Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 789–793
Bioorganic & Medicinal Chemistry

QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index

Abhilash Thakur,[a] Mamta Thakur,[a] Padmakar V. Khadikar,[b,*] Claudiu T. Supuran[c] and Purushottam Sudele[a]

[a]*Department of Chemistry, Bhopal Institute of Technology and Science, Bhopal, India*
[b]*Research Division, Laxmi Fumigation and Pest control, (P) Ltd. 3, Khatipura, Indore 452007, India*
[c]*University of Florence, Dipartmento di Chimica, Laboratoria di Chimica Bioinorganica, Via della Lastruccia 3, Rm. 188, 50019-Sesto Fiorentino, Firenze, Italy*

Received 13 August 2003; accepted 27 October 2003

Abstract—QSAR study on benzenesulphonamide carbonic anhydrase inhibitors has been made using the most discriminatory Balaban index (J). The regression analysis has shown that even in monoparametric regression this index gave excellent results. Furthermore, using the combination of the Balaban Index (J) with the first-order Randic connectivity index ($^1\chi$) and indicator parameters, tremendous improvement in the statistics has been observed. The results are critically discussed on the basis of regression data and cross-validation parameters.

1. Introduction

The physiological and physio-pathological processes in which carbonic anhydrases (CAs, EC 4.2.1.1) are involved were thoroughly investigated due to the pharmacological applications of their inhibitors, the unsubstituted sulfonamides.[1–3] It has also been shown[4] that such inhibitors are important clinical agents, mainly used in the treatment of glaucoma, gastro-duodenal ulcers, certain neurological disorders, motion and altitude sickness and some forms of tumors among others.[4b–e]

Consequently, a large number of aromatic and heterocyclic sulfonamides were synthesized and tested for their CA inhibitory potential.[5] Among them benzenesulfonamides have attracted much attention.[5,6]

Earlier QSAR studies on benzene-sulfonamide CAIs were mostly based on Hansch's[7] approach and some other such related molecular descriptors other than topological indices. However, we have shown that topological indices can be successfully used for this purpose, with interesting results being obtained and reported earlier.[8–14]

Literature data showed that even for the benzene-sulfonamide inhibitors no QSAR study using topological indices has been reported. In view of this and in continuation to our earlier work,[8–14] the present study deals with topological investigation on this class of CAIs, in which we have used the most discriminatory Balaban Index (J).[15]

At this stage, it is interesting to record that topological indices are the graph theoretical descriptors obtained by transforming molecular structure into molecular graph. In obtaining molecular graph all the carbon–hydrogen as well as heteroatom–hydrogen bonds are suppressed. A detailed information of these topological indices is reported in the literature.[16–20] Out of these topological indices Wiener (W)[21] and first-order connectivity ($^1\chi$) indices[22,23] have been used extensively. Comparatively less examples of quantitative structure–activity relationship (QSAR) studies based on Balaban index (J) are available in the literature. The question arises as to why more examples of QSAR based on the most discriminating Balaban index (J) have not appeared. We believe that the primary reason is that theoretical chemists have been very slow to appreciate the overriding importance

* Corresponding author. Tel.: +91-731-253-1906; fax: +91-731-276-3618; e-mail: pvkhadikar@rediffmail.com

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.10.058

of the Balaban index (J) in modifying their biological process. Nevertheless, earlier we have used this index successfully in developing QSPR/QSAR models.[24,25]

2. Result and discussion

The sulphonamides used (Fig. 1), their adopted CA II inhibitory activity[26] and structural parameters are given in Table 1.

The values of the Balaban index (J) used in the present study are given in Table 2. The use of the Balaban index (J) in modeling CA inhibitory activity of the sulphonamides is not enough, we need to prove its relative correlation power with respect to other widely used topological indices. We have, therefore, calculated the Wiener (W),[21] the Szeged (Sz)[27,28] and the first-order Randic connectivity ($^1\chi$) indices[22,23] for comparison. These are shown in Table 2. The details of calculation of these indices are given in the experimental section of this paper. In addition, we have used three indicator parameters, namely, I, I_1 and I_2 accounting for presence or absence of substitution at the 4-, 3-, and 2-positions in the aromatic nucleus.

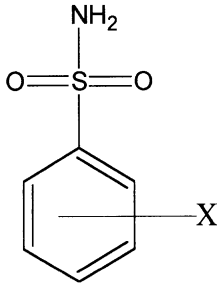

Figure 1. General structure of the sulphonamide CAIs used in present study (for detail see Table 1).

Table 1. Sulphonamides, their carbonic anhydrase II inhibitory activity (logKc) and indicator parameters[a]

Compd	X	I	I_1	I_2	logKc (exp)	logKc (calcd)	Residual
1	H	1	0	0	6.69	6.73	−0.04
2	4-Me	1	0	0	7.09	6.95	0.14
3	4-Et	1	0	0	7.53	7.45	0.08
4	4-*n*Pr	1	0	0	7.77	8.00	−0.23
5	4-*n*Bu	1	0	0	8.30	8.55	−0.25
6	4-*n*Am	1	0	0	8.86	9.06	−0.20
7	4-COOMe	1	0	0	7.98	7.75	0.23
8	4-COOEt	1	0	0	8.50	8.10	0.40
9	4-COOPr	1	0	0	8.77	8.48	0.29
10	4-COOBu	1	0	0	9.11	8.85	0.26
11	4-COOAm	1	0	0	9.39	9.18	0.21
12	4-COOHe	1	0	0	9.39	9.46	−0.07
13	4-CONHMe	1	0	0	7.08	7.65	−0.57
14	4-CONHEt	1	0	0	7.53	7.82	−0.29
15	4-CONHPr	1	0	0	8.08	8.04	0.04
16	4-CONHBu	1	0	0	8.49	8.30	0.19
17	4-CONHAm	1	0	0	8.75	8.54	0.21
18	4-CONHHe	1	0	0	8.88	8.76	0.12
19	4-CONHHp	1	0	0	8.93	8.93	0.00
20	3-COOMe	0	1	0	5.87	5.59	0.28
21	3-COOEt	0	1	0	6.21	5.94	0.27
22	3-COOPr	0	1	0	6.44	6.35	0.09
23	3-COOBu	0	1	0	6.95	6.78	0.17
24	3-COOAm	0	1	0	6.86	7.18	−0.32
25	2-COOMe	0	0	1	4.41	4.48	−0.07
26	2-COOEt	0	0	1	4.80	4.83	−0.03
27	2-COOPr	0	0	1	5.28	5.29	−0.01
28	2-COOBu	0	0	1	5.76	5.80	−0.04
29	2-COOAm	0	0	1	6.18	6.30	−0.12

[a] I, I_1 and I_2 are the indicator parameters accounting for the presence/absence of substituents respectively at the 4-, 3- and 2-positions on the aromatic nucleus, in presence of which, indicator parameters assumes the value of 1, otherwise they are zero.

In proposing QSAR models for modeling the activity, we have used the maximum R^2 method and adopted stepwise regression.[29] Initially, we have used Pogliani's quality factor Q[30,31] for investigating relative predictive power and finally the cross-validation techniques.[29]

The correlatedness of the topological indices used and their correlation with CA inhibitory activity is demonstrated in Table 3. A perusal of Table 3 shows that the Balaban index (J) alone correlates excellently with the inhibitory activity (logKc). This means that in monoparameteric regressions, this index alone is the most appropriate index giving statistically significant results.

Among the three indicator parameters mentioned above; the indicator parameter I correlates well with logKc (Table 3). This indicates that substitution at the 4-position is favourable for augmenting the inhibitory activity against CA II.

The stepwise regression resulted into the following monoparametric, statistically significant model using the Balaban index (J):

Table 2. Balaban index J and other related topological indices used in the present study

Compd	W	$^1\chi$	J	Sz
1	188	5.54855	2.53244	269
2	240	5.94239	2.52426	348
3	305	6.48040	2.48451	440
4	384	6.98040	2.42860	546
5	478	7.48040	2.36599	667
6	588	7.98040	2.30237	804
7	454	7.39108	2.48693	643
8	552	7.89108	2.45330	768
9	667	8.39108	2.40870	910
10	800	8.89108	2.35875	1070
11	952	9.39108	2.30717	1249
12	1124	9.89108	2.25632	1448
13	538	7.76376	2.51194	754
14	639	8.30177	2.50898	882
15	758	8.80177	2.48720	1028
16	896	9.30177	2.45339	1193
17	1054	9.80177	2.41248	1378
18	1233	10.3017	2.36798	1584
19	1434	10.8017	2.32232	1812
20	430	7.39108	2.62976	595
21	522	7.89108	2.59881	708
22	631	8.39108	2.55090	838
23	758	8.89108	2.49427	986
24	904	9.39108	2.43434	1153
25	406	7.40792	2.81074	547
26	492	7.90792	2.78232	648
27	595	8.40792	2.72865	766
28	716	8.90792	2.66177	902
29	856	9.40792	2.58961	1057

Table 3. Correlation matrix demonstrating correlation of the topological indices used and their correlation with the inhibitory activity (logKc)

	logKc	W	$^1\chi$	J	Sz	I
LogKc	1.00000					
W	0.49750	1.00000				
$^1\chi$	0.37907	0.97047	1.00000			
J	0.95413	−0.49995	−0.37093	1.00000		
Sz	0.53667	0.99791	0.96694	−0.53053	1.00000	
I	0.82201	0.11144	−0.04627	−0.72300	0.15446	1.00000

Table 4. Cross-validation parameters for the proposed models

Eq	Parameters	PRESS	SSY	PRESS/SSY	r^2_{cv}	SPRESS	PSE	R^2_A
1	J	4.9630	50.3900	0.0985	0.9015	—	—	—
2	J, I	2.9358	52.4161	0.0560	0.9440	0.3360	0.3182	0.9429
3	J, $^1\chi$, I	1.8755	53.4805	0.0350	0.9650	0.2736	0.2540	0.9621
4	J, $^1\chi$, W, I	1.3963	53.9557	0.0259	0.9741	0.2412	0.2194	0.9706

$$\text{logKc} = 31.4072 - 9.6189\ (\pm 0.5809)\ \text{J} \quad (1)$$

$$n = 29,\ \text{Se} = 0.4287,\ r = -0.9541,\ F = 274.192,\ Q = 2.2256.$$

Here, and thereafter, n is number of data point, Se is the standard error of estimation, r is the correlation coefficient, F is the F ratio and Q is the quality factor.

The aforementioned (eq 1) shows that the Balaban index (J) is negatively correlated with logKc. The inhibitory activity increases with decrease in the magnitude of the Balaban index (J). One should remember that this index is a highly discriminating descriptor whose value do not substantially increase with the molecular size and the number of rings present in the molecule of the investigated CAIs. Furthermore, Balaban index (J) is a variant of connectivity index, represents extended connectivity and is a good descriptor for the shape of the molecules. Decrease in all these parameters, therefore, favors the exhibition of logKc. Compared to this model all other monoparametric models were of comparatively poor statistics.

Among the bi-parametric regressions attempted, the one containing J and I gave better results than the one discussed above (eq 1):

$$\text{logKc} = 25.8505 - 7.6003\ (\pm 0.6591)\ \text{J} + 0.805\ (\pm 0.190)\ \text{I} \quad (2)$$

$$n = 29,\ \text{Se} = 0.336,\ r = 0.9731,\ F = 232.104,\ Q = 2.8961.$$

The positive coefficient of indicator parameter I supports favourable contribution of substituents at the 4-position for the value of logKc, as mentioned earlier. The negative coefficient of J has the same significance as mentioned above. All other bi-parameter regressions, though statistically significant, were of lower quality than the bi-parameter model (eq 2) discussed above.

Successive regression resulted into several tri-parametric models, out of which one containing Balaban index (J), first-order Randic connectivity index ($^1\chi$) and I was found to be the best:

$$\text{logKc} = 20.3759 - 6.1371\ (\pm 0.6623)\ \text{J} + 0.194\ (\pm 0.0514)\ ^1\chi + 1.1330\ (\pm 0.1775)\ \text{I} \quad (3)$$

$$n = 29,\ \text{Se} = 0.2736,\ r = 0.9829,\ F = 238.138,\ Q = 3.5925.$$

In this model (eq 3), the parameter $^1\chi$ is involved. This parameter coveys more information about the number of atoms in a molecule, which in turn is favorable in the present case. Significance of J and I is same as mentioned for the model expressed by (eq 2).

Finally, looking to the data point and following Thumb's rule for the use of number of descriptors compared to sample size, we have attempted tetra-parametric regressions. We observed that the tetra-parameteric model containing J, $^1\chi$, W and I gave the best results.

$$\text{logKc} = 18.3126 - 6.538\ (\pm 0.6004)\ J + 0.7504\ (\pm 0.1999)\ ^1\chi - 0.0025\ \left(\pm 8.6322 \times 10^{-4}\right)\ \text{W} + 1.2837\ (\pm 0.1651)\ \text{I} \quad (4)$$

$$n = 29,\ \text{Se} = 0.2412,\ r = 0.9873,\ F = 231.859,\ Q = 4.0933.$$

Here, Wiener (W) index accounts for the size, shape and branching of the molecule. Hence decrease in these factors favors logKc. Significance of all other parametersis the same as mentioned for the model expressed by (eq 4).

These results showed that our choice for indicator parameters I_1 and I_2 accounting for substitution at the 3- and 2-positions is not that justified. Although all the models involving these indicators were of sound statistics they were inferior to those containing the indicator I. These favor the dominating role of substitution at the 4-position for increasing activity logKc.

We also observed that as we pass from mono- to tetra-parameter models, the Q values goes on increases giving highest value for tetra-parameter model. Therefore, quality and predictivity of models run parallel to each other.

Todeschini[32] has criticized the use of the Q factor. Although in the present case the use of this factor gives quite good results, even then we have undergone a cross-validation methodology for deciding predictive power of the proposed models. This is needed because a model with good statistics may not have good predictivity.

The various cross-validation parameters, calculated for the proposed models, are presented in Table 4 and are discussed below.

Predicted residual sum of squares (PRESS) appears to be the most important cross-validation parameter accounting for a good estimate of the real predictive error of the models. Its value less than sum of the squares of response value (SSY) indicates that the model predicts better than chance and can be considered statistically significant. In our case (Table 4), PRESS $<<$ SSY indicating that all the models obtained are statistically significant and are better than chance.

To be a reasonable QSAR model, PRESS/SSY should be smaller than 0.4 and the value of the ratio smaller than 0.1 indicates an excellent model. The data presented in Table 4 indicates that all the models proposed by us are excellent.

Other important cross-validated parameters are S_{PRESS} (uncertainty of prediction) and r^2_{cv} (overall predictive ability). In our case the former is found to be the same as that of standard error of estimation (Se). Therefore, S_{PRESS} is not useful in our case. The later parameter r^2_{cv} is of important as the predictive power goes on increasing with increase in r^2_{cv}. The tetra-parametric model has the highest value for r^2_{cv} and is thus has highest predictive power.

Under the condition that S_{PRESS} equals Se, another cross-validation parameter namely PSE (predictive square of error) is some times used. Lowest value of PSE is in favors of highest predictive power. The PSE values recorded in Table 4 once again show the tetra-parametric model has the highest predictive power.

3. Conclusions

From the aforementioned results and discussion, we conclude that the Balaban index (J) can be used successfully for modeling CA inhibitory activity of the sulfonamides studied here. The J index was found better than the widely used Wiener index (W) and first-order Randic connectivity index ($^1\chi$). The predictive power of J index increased by the combination with first-order Randic ($^1\chi$) index and an indicator parameter I.

4. Experimental

4.1. Carbonic anhydrous activity (logKc)

The CA II inhibitory activity expressed as logKc, was taken from the literature.[26]

4.2. Topological indices

All the topological indices used are calculated from the hydrogen suppressed molecular graphs. Though their calculations are exclusively discussed in the literature, we give below the expressions used for their calculations.

4.2.1. Wiener index (W). Wiener index $W = W(G)$ of G is defined as the half sum of the elements of the distance matrix.

$$W - W(G) = 1/2\sum_{i=1}, \sum_{j-1}\left(D_{ij}\right) \tag{5}$$

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

4.2.2. The first-order connectivity index ($^1\chi$). The first-order connectivity index $^1\chi = {}^1\chi(G)$ of G is defined by Randic as

$${}^1\chi {}^1\chi(G) = \sum_{i,j}[d(i).d(j)]^{-0.5} \tag{6}$$

4.2.3. Balaban index (J). The Balaban index $J = J(G)$ of G is defined as

$$J = M/\mu + 1\sum_{\text{bonds}}\left(d_i.d_j\right)^{-0.5} \tag{7}$$

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

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

$$\grave{\imath} = M - N + 1 \quad (8)$$

4.2.4. Szeged index (Sz). The Szeged index, $Sz = Sz(G)$, is calculated according to the following expression:

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

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

4.2.5. Regression analysis. All the regression are carried out using maximum R^2 method.

4.2.6. Software. The calculation of topological indices and regression analysis were performed using softwares developed by Professor Istvan Lukovits, Hungarian academy of sciences, Budapest, Hungary.

Acknowledgements

Authors thanks are due to Istvan Lukovits for providing softwares. One of the authors (P.V.K.) is thankful to Professor Ivan Gutman for introducing him (P.V.K.) to these fascinating fields: Molecular Topology and Graph Theory.

References and notes

1. (a) Maren, T. H.; Wynns, G. C.; Wistrand, P. J. *Mol. Pharmacol.* **1993**, *44*, 901. (b) Maren, T. H.; Wynns, G. C.; Wistrand, P. J. *Mol. Pharmacol.* **1993**, *44*, 901.
2. Supuran, C. T.; Scozzafava, A.; Casini, A. *Med. Res. Rev.* **2003**, *23*, 146.
3. Supuran, C. T.; Scozzafava, A. *Exp. Opin. Ther. Pat.* **2002**, *12*, 217.
4. (a) Supuran, C. T. *Roumanian Chem. Quart. Rev.* **1993**, *1*, 77. (b) Scozzafava, A.; Owa, T.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Med. Chem.* **2003**, *10*, 925. (c) Casini, A.; Antel, J.; Abbate, F.; Scozzafava, A.; David, S.; Waldeck, H.; Schäfer, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 841. (d) Casini, A.; Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Cancer Drug Targets* **2002**, *2*, 55. (e) Supuran, C. T. *Exp. Opin. Invest. Drugs* **2003**, *12*, 283.
5. Owa, T.; Yoshino, H.; Okauchi, T.; Yoshimatsu, K.; Ozawa, Y.; Sugi, N. H.; Nagasu, T.; Koyanagi, N.; Kitoh, K. *J. Med. Chem.* **1999**, *42*, 3789.
6. Kakeya, N.; Yata, N.; Kamada, A.; Akadi, M. *Chem. Pharm. Bull.* **1970**, *18*, 191.
7. Hansch, C.; McLarin, J.; Klein, J.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493.
8. Agrawal, V. K.; Shrivastava, S.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *11*, 5353.
9. Saxena, A.; Agrawal, V. K.; Khadikar, P. V. *Oxid. Commun.* **2003**, *26*, 9.
10. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 447.
11. Agrawal, V. K.; Sharma, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 2993.
12. Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287.
13. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 171.
14. Saxena, A. Quantitative Structure Activity Relationship Study for a Group of Sulphonamide Schiff's Base Inhibition of Carbonic Anhydrase. PhD Thesis, D.A. University, Indore, India, 2000.
15. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
16. Balaban, A. T. *Chemical Application of Graph Theory*; Academic: London, 1976.
17. Devillers, J., Balaban, A. T., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon & Breach: Amsterdam, 1999.
18. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
19. Diudea, M. V., Ed. *QSPR/QSAR Studies for Molecular Descriptors*; Nova Science: Huntington, New York, 2001.
20. Diudea, M. V.; Khadikar, P. V. *Molecular Topology and its Application*; New Age Int. Nat.: New Delhi, India. In press.
21. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
22. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
23. Randic, M. *J. Mol. Graph. Model.* **2001**, *20*, 19.
24. Sapre, N. S.; Sikarwar, A.; Khadikar, P. V. *Oxid. Commun.* **2001**, *24*, 28.
25. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chem. Belg.* **1997**, *106*, 772.
26. Maren, T. H.; Clare, B. W.; Supuran, C. T. *Roumanian Chem. Quart. Rev.* **1994**, *2*, 259.
27. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
28. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Compt. Sci.* **1995**, *35*, 547.
29. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley VCH: New York, 2000.
30. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
31. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
32. Todeschini, R. Some Observations about Pogliani Q Quality Index; In *Chemometrics Web News*; Milano Chemometrics & QSAR Research Group (file: ///C/Windows/Desktop/Web news on chemometrics.html), 2001.