Pergamon

SCIENCE DIRECT

Bioorganic & Medicinal Chemistry 11 (2003) 5353–5362

BIOORGANIC & MEDICINAL CHEMISTRY

Quantitative Structure–Activity Relationship Study on Sulfanilamide Schiff's Bases: Carbonic Anhydrase (CA) Inhibitors

Vijay K. Agrawal,[a] Shachi Srivastava,[a]
Padmakar V. Khadikar[b,*] and Claudiu T. Supuran[c]

[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*
[c]*University of Florence, Dipartmento di Chimica, Laboratorio di Chimica Bioinorganica, Via della Lastruccia, 3, RM.188, Polo Scientifico, 50019-Sesto Fiorentino (Firenze), Italy*

Received 11 March 2003; accepted 24 September 2003

Abstract—The paper deals with quantitative structure–activity studies on a group of sulfanilamide Schiff's base inhibitors of carbonic anhydrase (CA) using distance-based topological indices. The regression analysis of the data has shown that the activities of the compounds used in inhibiting Carbonic AnhydraseII (CAII) activity can be modeled excellently in multi-parametric model in that some indicator parameters are also involved. The results are discussed critically.

Introduction

Quantitative structure–activity relationship (QSAR) studies have often been carried out by using regression analysis. The biological activities are being modeled using a set of molecular descriptors that often include topological indices, experimental physico-chemical data, geometrical structural parameters, theoretical elctronic indices obtained from quantum mechanical calculations, atomic charges, Koopman's energies and electrostatic potentials and so on. The sources of such studies depends on whether the molecular descriptors chosen are appropriate to explain biological activities. Among the topological indices the distance-based topological indices are most prominent ones in QSAR studies. Likewise, among the electronic indices the most commonly used in QSAR studies are the atomic charges.[1,2] In an earlier report by Supuran and Clare[3] a QSAR study on a group of sulfanilamide inhibitors of carbonic anhydrase, using the semi-empirical Austin Model-1 (commonly known and available as AM_1 method) was presented. This method is an improved version of modified neglect diatomic overlap (MNDO) method. By using such studies these authors[4–7] have shown that increased carbonic anhydraseII (CAII) and carbonic anhydraseIV (CAIV) inhibition properties of aromatic/heterocyclic sulfonamides are connected with the presence of elongated inhibitor molecules.

In another report we have investigated QSAR for the same group of sufanilamide Schiff's base inhibitors of carbonic anhydrase,[8] using Wiener index.[9] The result presented therein indicted that we can use such distance-based topological indices for the purpose. However, the simple regression attempts were not of that importance.

Prompted by the fact that one can use topological indices for modeling inhibitors of carbonic anhydrase we have recently modeled carbonic anhydrase inhibition activity of sulfonamides using molecular negentropy.[10] Earlier, we have reported QSAR studies on carbonic anhydrase inhibition of ureido and thioureido derivatives of aromatic/ heterocyclic sulfonamides using topological indices.[11] Also, using topological indices we have reported QSAR studies on some antimalarial sulfonamides.[12]

In view of the above findings, in this paper we investigated QSAR for a group of sulfanilamide Schiff's base

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

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.09.037

inhibitors[3] of the physiologically relevant isozyme CAII using a large pool of topological indices. The results, as discussed below, show that the use of topological indices as QSAR descriptors in investigating such problems may be of relevance for the drug design. The set of sulfanilamide Schiff's base inhibitors of carbonic anhydrase used and inhibition of CAII is adopted from the previous work of Supuran and Clare.[3]

Results and Discussion

The investigated Schiff's bases, inhibition data of CAII,[3] and the indictor parameters are shown in Table 1. The indictor parameters used (Ip_1 and Ip_2) are related to styryl or substituted styryl and phenyl or substituted phenyl groups respectively, the details of which is given in the experimental section.

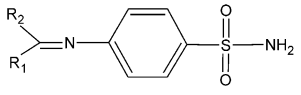
The values of the calculated topological indices for the aforementioned Schiff's bases are given in Table 2. The details of their calculation are also given in the experimental section.

Earlier we have stated that the success of QSAR studies depends on whether the molecular descriptors chosen are appropriate to explain biological activities. The same statement is applicable in the possible use of topological indices in the present study also. However, initially one is not sure which topological index/indices is/are useful for such studies. The common practice adopted is to use initially a large set of topological indices and then fix the appropriate topological indices adopting step-wise regression analysis. This is the reason for the use of large set of topological indices presented in Table 2. Record that all these topological indices are calculated using hydrogen suppressed molecular graphs of the compounds used in that all the carbon–hydrogen as well as hetero-atom-hydrogen bonds are depleted.

The inter-correlatedness among the topological indices as well as indictor parameters is demonstrated in Table 3. Such inter-correlatedness help in arriving at statistically most significant model.

The data presented in Table 3 show that W, Sz, logRB, $^1\chi^v$, $^1\chi$ and B are highly linearly correlated. Comparatively correlatedness of J and MRI with other topologi-

Table 1. Compounds used in present study and their activity [logK(CAII)]



Compd	R_1	R_2	LogK(CAII)	Ip_1	Ip_2
1	Phenyl	H	1.4314	0	0
2	2-Hydroxyphenyl	H	1.6128	0	1
3	2-Nitrophenyl	H	1.3222	0	1
4	4-Chlorophenyl	H	1.4472	0	1
5	4-Hydroxyphenyl	H	1.2788	0	1
6	4-Methoxyphenyl	H	1.2788	0	1
7	4-Dimethylaminophenyl	H	0.9031	0	1
8	4-Nitrophenyl	H	0.6990	0	1
9	4-Cyanophenyl	H	1.0414	0	1
10	3-Methoxy-4-hydroxyphenyl	H	0.9031	0	0
11	3,4-Dimethoxyphenyl	H	0.4771	0	0
12	3-Methoxy-4-acetoxyphenyl	H	1.0000	0	0
13	2,3-Dihydroxy-5-formylphenyl	H	0.3010	0	0
14	2-Hydroxy-3-methoxy-5-formylphenyl	H	0.4771	0	0
15	3,4,5-Trimethoxyphenyl	H	0.4771	0	0
16	3-Methoxy-4-hydroxy-5-bromophenyl	H	0.6021	0	0
17	2-Furyl	H	0.6990	0	0
18	5-Methyl-2-furyl	H	0.6021	0	0
19	Pyrol-2-yl	H	0.3010	0	0
20	Imidazol-4(5)-yl	H	1.0792	0	0
21	2-Pyridyl	H	0.9542	0	0
22	3-Pyridyl	H	0.9031	0	0
23	4-Pyridyl	H	0.6990	0	0
24	Styryl	Me	−0.4089	1	0
25	4-Methoxystyryl	Me	−0.9208	1	0
26	4-Dimethylamino styryl	Me	−1.0000	1	0
27	3,4,5-Trimethoxy styryl	Me	−0.6198	1	0
28	Styryl	Ph	−0.2518	1	0
29	4-Methoxy styryl	Ph	0.1761	1	0
30	4-Dimethylaminostyryl	Ph	0.2279	1	0
31	3,4,5-Trimethoxy styryl	Ph	0.3711	1	0
32	3,4,5-Trimethoxy styryl	4-$MeOC_6H_4$	0.1038	1	0
33	3-Nitrostyryl	4-$MeOC_6H_4$	−0.1871	1	0
34	3,4,5-Trimethoxy styryl	4-$NH_2C_6H_4$	−0.0706	1	0
35	3,4,5-Trimethoxy styryl	4-PhC_6H_4	0.3945	1	0

$Ip_1 = 1$, if R_1 = Styryl or substituted styryl group, otherwise $Ip_1 = 0$. $Ip_2 = 1$, if R_1 = Phenyl or substituted phenyl group, otherwise $Ip_2 = 0$.

Table 2. Molecular descriptors used in present study

Compd	W	$^1\chi$	J	Sz	LogRB	$^1\chi^v$	MRI
1	702	8.5547	1.6396	1068	199.6263	5.5010	0.183
2	802	8.9653	1.6804	1210	227.3501	2.3790	0.165
3	1042	9.876	1.7364	1534	291.6513	4.8440	0.176
4	826	8.9485	1.6327	1258	230.6517	5.3750	0.178
5	826	8.9485	1.6327	1258	230.6517	5.4050	0.165
6	969	9.4865	1.6194	1467	265.3184	4.1500	0.151
7	1114	9.8592	1.6240	1678	300.6783	6.2990	0.147
8	1114	9.8592	1.6240	1678	300.6783	6.6890	0.176
9	969	9.4865	1.6194	1467	265.3184	5.7550	0.386
10	1076	9.8972	1.6801	1618	295.9659	6.3670	0.339
11	1228	10.4352	1.6895	1838	333.6283	6.3020	0.470
12	1382	10.8079	1.7110	2060	371.9839	7.6780	0.339
13	1180	10.3079	1.7600	1761	327.1764	6.3879	0.540
14	1326	10.8459	1.7872	1966	365.2919	6.8330	0.206
15	1507	11.3839	1.7815	2237	408.3353	7.6198	0.395
16	1200	10.3079	1.7278	1801	329.5616	6.8130	0.228
17	593	8.0547	1.6474	832	171.1056	5.2340	0.246
18	701	8.4485	1.6468	975	199.2208	6.0380	0.209
19	593	8.0547	1.6474	832	171.1056	5.660	0.209
20	593	8.0547	1.6474	832	171.1056	5.4750	0.209
21	702	8.5547	1.6396	1068	199.6263	5.8970	0.285
22	702	8.5547	1.6396	1068	199.6263	7.760	0.285
23	702	8.5547	1.6396	1068	199.6263	5.9490	0.285
24	1012	9.914	1.7951	1494	287.1814	6.3990	0.602
25	1290	10.8459	1.8449	1896	360.5508	6.3580	0.426
26	1442	11.2186	1.8714	2110	400.1013	7.720	0.351
27	1890	12.7601	2.0300	2752	522.6562	7.4390	0.344
28	1674	12.4865	1.6387	2523	473.3709	8.0230	0.523
29	2025	13.4184	1.6824	3036	566.5435	8.6350	0.417
30	2216	13.791	1.7024	3308	616.3422	8.6930	0.435
31	2771	15.3326	1.8167	4114	768.0574	9.6040	0.399
32	3290	16.2644	1.8298	4895	894.9283	7.9110	0.399
33	2610	14.7397	1.7605	3868	720.2449	10.270	0.370
34	3014	15.7264	1.8286	4488	828.8765	9.0800	0.345
35	4544	18.2989	1.5254	6988	1182.912	10.701	0.400

cal indices is less significant. Same is the case with the indicator parameter Ip_1. This means that appropriate topological indices to be used in mathematical analysis are J, MRI, and Ip_1.

The aforementioned results (Table 3) also indicate that a model in which any combination of highly correlated indices occurs may suffer from the defect due to collinearity. However, such conditions are discussed critically by Randic[13] and we will use recommendations of Randic[21] to deal with such cases.

In arriving at a statistically most significant model we have used maximum R^2 method in undergoing multiple regression analysis.[14] The results obtained in simple as well as multiple regression analysis are presented in Table 4.

A perusal of Table 4 shows that two mono-parametric regressions involving J and MRI, but of quite lower quality are obtained. This result of regression analysis is in accordance with the data presented in Table 3. The details of these models are hardly worthwhile mentioning at all. The successive regression analysis resulted into five bi-parametric models all having significantly better statistics than the mono-parametric model discussed above. It is interesting to record that except for model 5 (Table 4) involving J and Ip_1, all the bi-parametric models are of similar quality. The model involving J and Ip_1 is the best bi-parametric model for modeling CAII inhibition. This model is found as:

$$\begin{aligned}&\log K(\mathrm{CAII}) = \\ &\quad -2.3666(\pm 0.7414)\mathrm{J} - 0.8239(\pm 0.1520)\mathrm{Ip1} + 4.8475 \\ &n = 35,\ \mathrm{Se} = 0.3664,\ R = 0.8418,\ F = 38.917, \\ &Q = 2.2975\end{aligned} \tag{1}$$

Here and thereafter n is the number of compounds used, Se is the standard error of estimation, R is the multiple correlation coefficient, F is the F-ratio and Q is the quality factor.[15,16] The quality factor Q is a useful parameter to be used in deciding predictive potential of the model. Q is defined as the ratio of correlation coefficient to the standard error of estimation. Its use is criticized[17] recently but we find that it is a good parameter to explain the predictive potential of the models proposed by us. The higher the value of Q the better is the predictive potential of the models.

The aforementioned model (eq 1) indicates that J and Ip_1 are negatively correlated with logK(CAII). That decreased in the magnitude of J and Ip_1 favours the exhibition of logK(CAII). This means that the presence of styryl or substituted styryl group at R_1 has a retarding effect.

With a hope of obtaining a still better model we attempted successive regressions resulting at nine tri-parametric models (Table 4). Except for models 9 and 10 all other models were of improved statistics than the

Table 3. Correlation matrix

	LogK(CAII)	W	$^1\chi$	J	Sz	LogRB	$^1\chi^v$	MRI	Ip_1	Ip_2
LogK(CAII)	1.0000									
W	−0.4236	1.0000								
$^1\chi$	−0.4925	0.9858	1.0000							
J	−0.6643	0.2449	0.3405	1.0000						
Sz	−0.4084	0.9995	0.9825	0.2224	1.0000					
logRB	−0.4389	0.9991	0.9907	0.2662	0.9978	1.0000				
$^1\chi^v$	−0.5695	0.8071	0.8276	0.2541	0.8032	0.8127	1.0000			
MRI	−0.6708	0.4191	0.4788	0.3433	0.4153	0.4314	0.4943	1.0000		
Ip_1	−0.7848	0.7343	0.7917	0.5133	0.7245	0.7523	0.6955	0.6356	1.0000	
Ip_2	0.5661	−0.2839	−0.3057	−0.3459	−0.2736	−0.2940	−0.5377	−0.5021	−0.3932	1.0000

Table 4. Regression parameters and quality of the correlations of the activity logK(CAII) with various parameters used in present study

Model no.	Parameters used	A_i (i = 1, 2, 3, 4, 5, 6)	B (intercept)	Se	Corr. coeff. (R)	R^2	F-ratio	Q = R/Se
1	J	4.4290(±0.8676)	8.0877	0.4996	−0.6643	0.4412	26.060	−1.3297
2	MRI	−3.4939(±0.6724)	1.5989	0.4957	−0.6708	0.4500	27.003	−1.3532
3	W	$2.4384\times10^{-4}(\pm1.1064\times10^{-4})$	0.6592	0.3920	0.8164	0.6665	31.975	2.0827
	Ip_1	−1.4058(±0.2056)						
4	J	−3.2801(±0.7280)	6.9303	0.3937	0.8146	0.6635	31.549	2.0691
	MRI	2.6143(±0.5686)						
5	J	−2.3666(±0.7414)	4.8475	0.3664	0.8418	0.7087	38.917	2.2975
	Ip_1	−0.8239(±0.1520)						
6	MRI	−1.5035(±0.6897)	1.2678	0.3925	0.8158	0.6655	31.838	2.0785
	Ip_1	−0.8221(±0.1811)						
7	LogRB	$9.6536\times10^{-4}(\pm4.2721\times10^{-4})$	0.6367	0.3907	0.8178	0.6687	32.299	2.0932
	Ip_1	−1.4320(±0.2112)						
8	W	$1.7814\times10^{-4}(\pm1.0302\times10^{-4})$	4.1941	0.3555	0.8569	0.7343	28.555	2.4104
	J	−2.0770(±0.7386)						
	Ip_1	−1.0975(±0.2163)						
9	W	$2.2374\times10^{-4}(\pm1.0565\times10^{-4})$	1.0236	0.3728	0.8413	0.7078	25.032	2.2567
	MRI	−1.370(±0.6577)						
	Ip_1	−1.1486(±0.2309)						
10	W	−0.0049(±0.0027)	0.8076	0.3762	0.8381	0.7024	24.395	2.2278
	Sz	0.0034(±0.0017)						
	Ip_1	−1.2092(±0.2220)						
11	W	$2.4151\times10^{-4}(\pm9.8491\times10^{-4})$	0.4987	0.3489	0.8625	0.7440	30.028	2.4721
	Ip_1	−1.2399(±0.1909)						
	Ip_2	0.4679(±0.1528)						
12	J	−2.3236(±0.6910)	5.1396	0.3414	0.8689	0.7549	31.830	2.5451
	MRI	−1.4516(±0.6000)						
	IP_1	−0.5862(±0.1724)						
13	J	−2.0408(±0.7432)	4.1373	0.3554	0.8570	0.7344	28.576	2.4114
	SZ	$1.1691\times10^{-4}(\pm6.7395\times10^{-4})$						
	IP_1	−1.0966(±0.2156)						
14	J	−2.0371(±0.6942)	4.1609	0.3373	0.8722	0.7608	32.862	2.5858
	IP_1	−0.7228(±0.1453)						
	Ip_2	0.3903(±0.1502)						
15	LogRB	$9.6192\times10^{-4}(\pm3.7947\times10^{-4})$	0.4742	0.3470	0.8642	0.7468	30.478	2.4905
	IP_1	−1.2673(±0.1950)						
	Ip_2	0.4697(±0.1519)						
16	Sz	$1.6023\times10^{-4}(\pm6.3777\times10^{-5})$	0.5025	0.3475	0.8637	0.7460	30.354	2.4855
	IP_1	−1.2384(±0.1878)						
	Ip_2	0.4641(±0.1522)						
17	W	$1.5845\times10^{-4}(\pm9.6726\times10^{-5})$	4.5415	0.3325	0.8804	0.7750	25.840	2.6478
	J	−2.0685(±0.6908)						
	MRI	−1.3677(±0.5866)						
	IP_1	−0.8433(±0.2298)						
18	W	$1.8731\times10^{-4}(\pm9.3360\times10^{-5})$	3.4561	0.3219	0.8883	0.7891	28.058	2.7596
	J	−1.7241(±0.6807)						
	IP_1	−1.0079(±0.1985)						
	Ip_2	0.4005(±0.1434)						
19	W	−0.0179(±0.0048)	0.4120	0.3320	0.8807	0.7757	25.935	2.6527
	LogRB	0.0294(±0.0094)						
	Sz	0.0069(±0.0019)						
	IP_1	−1.5938(±0.2313)						
20	J	−2.0710(±0.6845)	4.5236	0.3306	0.8818	0.7776	26.225	2.6673
	MRI	−1.3612(±0.5833)						
	logRB	$6.4984\times10^{-4}(\pm3.714\times10^{-4})$						
	IP_1	−0.8695(±0.2326)						
21	J	−2.0336(±0.6946)	4.4864	0.3321	0.8806	0.7755	25.914	2.6516
	MRI	−1.3728(±0.5856)						
	Sz	$1.0491\times10^{-4}(\pm6.3191\times10^{-5})$						
	IP_1	−0.8439(±0.2285)						
22	J	−2.0853(±0.6754)	4.5332	0.3279	0.8839	0.7812	26.785	2.6956
	MRI	−1.0335(±0.6168)						
	IP_1	−0.5778(±0.1656)						
	Ip_2	0.2969(±0.1563)						
23	J	−1.7274(±0.6732)	3.4379	0.3194	0.8902	0.7924	28.622	2.7871
	LogRB	$7.6428\times10^{-4}(\pm3.5770\times10^{-4})$						
	IP_1	−1.0357(±0.2009)						
	Ip_2	0.4017(±0.1423)						
24	W	−0.0044(±0.0019)	5.0457	0.3092	0.9010	0.8119	25.052	2.9140
	J	−2.5623(±0.6751)						
	MRI	−1.3288(±0.5458)						
	LogRB	0.0178(±0.0075)						

(*Continued on next page*)

Table 4 (*continued*)

Model no.	Parameters used	A_i (i = 1, 2, 3, 4, 5, 6)	B (intercept)	Se	Corr. coeff. (R)	R^2	F-ratio	Q = R/Se
	IP_1	−1.1322(±0.2458)						
25	W	−0.0199(±0.0041)	0.8889	0.2841	0.9172	0.8412	30.719	3.2284
	MRI	−1.7587(±0.5086)						
	LogRB	0.0315(±0.0081)						
	Sz	0.0078(±0.0016)						
	IP_1	−1.2527(±0.2211)						
26	W	−0.0156(±0.0043)	0.2622	0.2982	0.9083	0.8251	27.360	3.0459
	LogRB	0.0279(±0.0084)						
	Sz	0.0057(±0.0018)						
	IP_1	−1.4977(±0.2104)						
	IP_2	0.3868(±0.1352)						
27	W	−0.0165(±0.0048)	2.6986	0.2799	0.9226	0.8512	26.685	3.2962
	J	−1.1113(±0.8115)						
	MRI	−1.6463(±0.5078)						
	LogRB	0.0297(±0.0080)						
	Sz	0.0059(±0.0022)						
	Ip_1	−1.1855(±0.2233)						
28	W	−0.0180(±0.0041)	0.6823	0.2741	0.9259	0.8573	28.027	3.3780
	MRI	−1.3572(±0.5403)						
	LogRB	0.0301(±0.0078)						
	Sz	0.0068(±0.0017)						
	Ip_1	−1.2702(±0.2136)						
	Ip_2	0.2430(±0.1368)						
29	W	−0.0197(±0.0038)	1.3627	0.2604	0.9334	0.8712	31.564	3.5845
	$^1\chi^v$	−0.1201(±0.0470)						
	MRI	−1.5152(±0.4758)						
	Sz	0.0076(±0.0015)						
	LogRB	0.0328(±0.0074)						
	Ip_1	−1.2420(±0.2027)						
30	W	−0.0218(±0.0042)	2.8137	0.2769	0.9243	0.8544	27.379	3.3380
	$^1\chi$ MRI	−0.3133(±0.1967)						
	Sz	−1.5444(±0.5136)						
	LogRB	0.0078(±0.0016)						
	Ip_1	0.0421(±0.0103)						
31	W	−0.0142(±0.0047)	2.6323	0.2675	0.9322	0.8689	25.570	3.4849
	J	−1.2043(±0.7770)						
	MRI	−1.2139(±0.5353)						
	LogRB	0.0281(±0.0077)						
	Sz	0.0047(±0.0022)						
	Ip_1	−1.1983(±0.2135)						
	Ip_2	0.2560(±0.1338)						
32	W	−0.0213(±0.0039)	2.8538	0.2567	0.9377	0.8793	28.107	3.6529
	$^1\chi^v$	−0.1107(±0.0469)						
	B	−0.2488(±0.1844)						
	MRI	−1.3641(±0.4821)						
	Sz	0.0076(±0.0015)						
	LogRB	0.0411(±0.0096)						
	Ip_1	−1.2435(±0.1998)						

models discussed above. Furthermore, the model containing J, Ip_1 and Ip_2 as the correlating parameters is found to be the best tri-parametric model. This model is found as below:

$$\begin{aligned}\log K(CAII) &= -2.0371(\pm 0.6942)J \\ &- 0.7228(\pm 0.11453)Ip_1 + 0.3903(\pm 0.11502)Ip_2 + 4.1609 \\ n &= 35,\ Se = 0.3373,\ R = 0.8722,\ F = 32.802, \\ Q &= 2.5838\end{aligned} \quad (2)$$

The negative coefficient of J and Ip_1 have the same influence on logK(CAII) as discussed above. In addition to J and Ip_1, this model also contains the Ip_2 term having positive coefficient. This indicates, therefore, that the presence of a phenyl or substituted phenyl group at R_1 is favourable for exhibition of logK(CAII).

Further, step-wise regression resulted into seven tetra-parametric models (Table 3). All of them have better statistics than the tri-parametric model discussed above. Out of these seven tetra-parametric model the model consisting of J, logRB, Ip_1 and Ip_2 is the best. This model is as below:

$$\begin{aligned}\log K(CAII) &= -1.7274(\pm 0.6732)J + 7.6428 \times 10^{-4} \\ &(\pm 3.5770 \times 10^{-4})\log RB - 1.0357(\pm 0.2009)Ip_1 \\ &+ 0.4017(\pm 0.1423)Ip_2 + 3.4374 \\ n &= 35,\ Se = 0.3194,\ R = 0.8902,\ F = 30.719, \\ Q &= 3.2284\end{aligned} \quad (3)$$

The influence of parameters J, Ip_1 and Ip_2 in the exhibition of logK(CAII) is the same as discussed above. Additionally, the increase in the magnitude of logRB is in favour of an increase in logK(CAII).

Among the three penta-parametric models, the model containing W, MRI, logRB, Sz and Ip_1 gave better results as:

$$\begin{aligned}\log K(CAII) = {} & -0.0199(\pm 0.0041)W - \\ & 1.7587(\pm 0.5086)MRI + 0.0315(\pm 0.0081)\log RB \\ & + 1.2527(\pm 0.2211)Ip_1 + 0.8889\end{aligned}$$

$n = 35$, Se $= 0.2841$, $R = 0.9172$, $F = 30.719$, Q $= 3.719$ (4)

Similarly, out of the four hexa-parametric models the model containing W, $^1\chi$, MRI, Sz, logRB, and Ip_1 gave better results. This model is found as:

$$\begin{aligned}\log K(CAII) = {} & -0.0197(\pm 0.0038)W - \\ & 0.1201(\pm 0.0490)\ {}^1\chi^v - 1.5152(\pm 0.4758)MRI + \\ & 0.0076(\pm 0.0015)Sz + 0.0328(\pm 0.0074)\log RB - \\ & 1.2420(\pm 0.2027)Ip_1 + 1.3627\end{aligned} \quad (5)$$

$n = 35$, $Se = 0.2604$, $R = 0.9334$, $F = 31.564$, $Q = 3.5645$

Finally statistically significant, two hepta-parametric models were obtained and the one containing W, $^1\chi^v$, $^1\chi$, MRI, Sz, logRB and Ip_1 gave the best statistics. This model is presented as below:

$$\begin{aligned}\log K(CAII) = {} & -0.0213(\pm 0.0039)W \\ & - 0.1107(\pm 0.0469)\ {}^1\chi^v - 0.\ 2488(\pm 0.1844)^1\chi - \\ & 1.3641(\pm 0.4821)MRI + 0.0076(\pm 0.0015)Sz + \\ & 0.0411(\pm 0.0096)\log RB - 1.2435(\pm 0.1998)Ip_1 + 2.8538\end{aligned}$$

$n = 35$, Se $= 0.2567$, $R = 0.9377$, $F = 28.107$, $Q = 3.6529$ (6)

Table 5. Various correlation models and their qualities of correlations

Model no.	Regression expression
1	LogK(CAII) = 4.4290(±0.8676)J + 8.0877
2	LogK(CAII) = −3.4939(±0.6724) + 1.5989
3	LogK(CAII) = 2.4384×10^{-4}(±1.1064×10^{-4})W−1.4058(±0.2056)Ip_1 + 0.6592
4	LogK(CAII) = −3.2801(±0.7280)J + 2.6143(±0.5686)MRI + 6.9303
5	LogK(CAII) = −2.3666(±0.7414)J−0.8239(±0.1520)Ip_1 + 4.8475
6	LogK(CAII) = −1.5035(±0.6897)MRI−0.8221(±0.1811) Ip_1 + 1.2678
7	LogK(CAII) = 9.6536×10^{-4} (±4.2721×10^{-4})logRB−1.4320(±0.2112) Ip_1 + 0.6367
8	LogK(CAII) = 1.7814×10^{-4} (±1.0302×10^{-4})W−2.0770(±0.7386)J−1.0975(±0.2163) Ip_1 + 4.1941
9	LogK(CAII) = 2.2374×10^{-4} (±1.0565×10^{-4})W−1.370(±0.6577)MRI−1.1486(±0.2309) Ip_1 + 1.0236
10	LogK(CAII) = −0.0049(±0.0027)W + 0.0034(±0.0017)Sz−1.2092(±0.2220) Ip_1 + 0.8076
11	LogK(CAII) = 2.4151×10^{-4} (±9.8491×10^{-4})W−1.2399(±0.1909) Ip_1 + 0.4679(±0.1528) Ip_2 + 0.4987
12	LogK(CAII) = −2.3236(±0.6910)J−1.4516(±0.6000)MRI−0.5862(±0.1724) Ip_1 + 5.1396
13	LogK(CAII) = −2.0408(±0.7432)J + 1.1691×10^{-4} (±6.7395×10^{-4})Sz−1.0966(±0.2156) Ip_1 + 4.1373
14	LogK(CAII) = −2.0371(±0.6942)J−0.7228(±0.1453) Ip_1 + 0.3903(±0.1502) Ip_2 + 4.1609
15	LogK(CAII) = 9.6192×10^{-4} (±3.7947×10^{-4})logRB−1.2673(±0.1950) Ip_1 + 0.4697(±0.1519) Ip_2 + 0.4742
16	LogK(CAII) = 1.6023×10^{-4} (±6.3777×10^{-5})Sz−1.2384(±0.1878) Ip_1 + 0.4641(±0.1522) Ip_2 + 0.5025
17	LogK(CAII) = 1.5845×10^{-4} (±9.6726×10^{-5})W−2.0685(±0.6908)J−1.3677(±0.5866)MRI−0.8433(±0.2298) Ip_1 + 4.5415
18	LogK(CAII) = 1.8731×10^{-4} (±9.3360×10^{-5})W−1.7241(±0.6807)J−1.0079(±0.1985) Ip_1 + 0.4005(±0.1434) Ip_2 + 3.4561
19	LogK(CAII) = −0.0179(±0.0048)W + 0.0294(±0.0094)logRB + 0.0069(±0.0019)Sz−1.5938(±0.2313) Ip_1 + 0.4210
20	LogK(CAII) = −2.0710(±0.6845)J−1.3612(±0.5833)MRI + 6.4984×10^{-4} (±3.714×10^{-4})logRB−0.8695(±0.2326) Ip_1 + 4.5236
21	LogK(CAII) = −2.0336(±0.6946)J−1.3728(±0.5856)MRI + 1.0491×10^{-4} (±6.319110^{-5})Sz −0.8439(±0.2285) Ip_1 + 4.4864
22	LogK(CAII) = −2.0853(±0.6754)J−1.0335(±0.6168)MRI−0.5778(±0.1656) Ip_1 + 0.2969(±0.1563) Ip_2 + 4.5332
23	LogK(CAII) = −1.7274(±0.6732)J + 7.6428×10^{-4} (±3.5770×10^{-4})logRB−1.0357(±0.2009) Ip_1 + 0.4017(±0.1423) Ip_2 + 3.4379
24	LogK(CAII) = −0.0044(±0.0019)W−2.5623(±0.6751)J−1.3288(±0.5458)MRI + 0.0178(±0.0075)logRB
25	LogK(CAII) = −0.0199(±0.0041)W−1.7587(±0.5086)MRI + 0.0315(±0.0081)logRB + 0.0078(±0.0016)Sz −1.2527(±0.2211) Ip_1 + 0.8889
26	LogK(CAII) = −0.0156(±0.0043)W + 0.0279(±0.0084)logRB + 0.0057(±0.0018)Sz−1.4977(±0.2104) Ip_1 0.3868(±0.1352) Ip_2 + 0.2622
27	LogK(CAII) = −0.0165(±0.0048)W−1.1113(±0.8115)J−1.6463(±0.5078)MRI + 0.0297(±0.0080)logRB + 0.0059(±0.0022)Sz −1.1855(±0.2233) Ip_1 + 2.6986
28	LogK(CAII) = −0.0180(±0.0041)W−1.3572(±0.5403)MRI + 0.0301(±0.0078)logRB + 0.0068(±0.0017)Sz −1.2702(±0.2136) Ip_1 + 0.2430(±0.1368) Ip_2 + 0.6823
29	LogK(CAII) = −0.0197(±0.0038)W−0.1201(±0.0470)$^1\chi^v$−1.5152(±0.4758)MRI + 0.0076(±0.0015)Sz 0.0328(±0.0074)logRB−1.2420(±0.2027) Ip_1 + 1.3627
30	LogK(CAII) = −0.0218(±0.0042)W−0.3133(±0.1967)B−1.5444(±0.5136)MRI + 0.0078(±0.0016)Sz + 0.0421(±0.0103)logRB−1.2535(±0.2155) Ip_1 + 0.8137
31	LogK(CAII) = −0.0142(±0.0047)W−1.2043(±0.7770)J−1.2139(±0.5353)MRI + 0.0281(±0.0077)logRB + 0.0047(±0.0022)Sz−1.1983(±0.2135) Ip_1 + 0.2560(±0.1338) Ip_2 + 2.6323
32	LogK(CAII) = −0.0213(±0.0039)W−0.1107(±0.0469)$^1\chi^v$−0.2488(±0.1844)$^1\chi$−1.3641(±0.4821)MRI + 0.0076(±0.0015)Sz + 0.0411(±0.0096)logRB−1.2435(±0.1998) Ip_1 + 2.8538

At this point it is interesting to comment on the models expressed by eqs 4–6. All these models contain highly linearly correlated topological indices (Table 3). The change in the sign of the coefficients of W and Sz is probably due to this effect. However, in all the cases the coefficients of highly linearly correlated topological indices are considerably higher than their respective standard deviations. Such models, even though they contain highly correlated parameters, are considered statistically valid. Further evidence in this favour is obtained from the recommendation of Randic.[21] He stated that one should be particularly aware of a common fit in regression analysis in describing descriptors that are highly inter-correlated. Randic[21] further stated that, by discarding one of the descriptors which commonly duplicates another, we may be discarding a descriptor that nevertheless may carry useful structural information in the part in which it does not parallel with another descriptors. Thus, following Randic,[21] we may safely say that the occurrence of highly inter-correlated topological indices in the aforementioned models may perhaps be due to their different structural information content.

For a comparative study we have summarized all the models in Table 5. The aforementioned study indicates that models 29 and 32 (Tables 4 and 5), expressed by eqs 5 and 6, respectively, are the most appropriate for modeling logK(CAII). Which out of these two models have better predictive potential is decided by estimating logK(CAII) using both the models and comparing them with the observed values of logK(CAII). Such a comparison is shown in Table 6. The residue, that is the difference between observed and estimated logK(CAII) is in favour of the model 32 expressed by eq 6.

The predictive potential of the aforementioned models is obtained from Figures 1 and 2, in which the observed logK(CAII) are plotted against their estimated values. The values of R^2, are found to be 0.8712 and 0.8794, respectively, for models 29 and 32 (Tables 4 and 5) expressed by eqs 5 and 6, respectively. Hence, the latter model is statistically most significant for modeling logK(CAII).

It is interesting to compare the results obtained in the present study with those obtained in the earlier communication.[3] In the earlier study the semi-empirical AM_1 electronic parameters were used while the present study is concerned with the use of topological indices. Like in the earlier study higher parametric regressions for modeling logK(CAII) activity were also used in that hepta-parametric models gave R^2 ranging between 0.821 and 0.854. In the present study R^2 value is found between 0.8689 and 0.8793. This shows that the present methodology based on topological indices approach gives comparable results, which is important from the point of view of the drug design of such therapeutical agents.

Conclusion

Our results leads to the conclusion that inhibition of CA can be successfully modeled using distance-based topological indices. Also, that the topological indices may be used in conjuction with semi-empirical electronic parameters as calculated by AM_1 method.

Table 6. Comparison of observed and estimated logK(CAII) using models **29** and **32**

Compd	Obs. logK(CAII)	Estimated logK(CAII)			
		Model-29	Residue	Model-32	Residue
1	1.4314	1.2120	0.2194	1.2540	0.1774
2	1.6128	1.6260	−0.0132	1.6140	−0.0012
3	1.3222	1.1410	0.1812	1.0980	0.2242
4	1.4472	1.2450	0.2022	1.2580	0.1892
5	1.2788	1.2610	0.0178	1.2720	0.0068
6	1.2788	1.3320	−0.0532	1.2670	0.0118
7	0.9031	0.9770	−0.0739	0.9140	−0.0109
8	0.6990	0.8860	−0.187	0.8310	−0.132
9	1.0414	0.7830	0.2584	0.7690	0.2724
10	0.9031	0.8180	0.0851	0.7940	0.1091
11	0.4771	0.5300	−0.0529	0.4750	0.0021
12	1.0000	0.7760	0.224	0.6760	0.324
13	0.3010	0.5660	−0.265	0.5730	−0.272
14	0.4771	0.9410	−0.4639	0.8640	−0.3869
15	0.4771	0.4540	0.0231	0.3630	0.1141
16	0.6021	0.9750	−0.3729	0.9280	−0.3259
17	0.6990	0.5750	0.124	0.6760	0.023
18	0.6021	0.4090	0.1931	0.4840	0.1181
19	0.3010	0.5800	−0.279	0.6790	−0.378
20	1.0792	0.6020	0.4772	0.6990	0.3802
21	0.9542	1.0100	−0.0558	1.0710	−0.1168
22	0.9031	0.7860	0.1171	0.8650	0.0381
23	0.6990	1.0040	−0.305	1.0650	−0.366
24	−0.4089	−0.7910	0.3821	−0.7570	0.3481
25	−0.9208	−0.5530	−0.3678	−0.5880	−0.3328
26	−1.0000	−0.6840	−0.316	−0.7110	−0.289
27	−0.6198	−0.5960	−0.0238	−0.6680	0.0482
28	−0.2518	−0.0260	−0.2258	−0.0800	−0.1718
29	0.1761	0.07800	0.0981	0.0260	0.1501
30	0.2279	−0.0310	0.2589	−0.0470	0.2749
31	0.3711	0.0450	0.3261	0.0720	0.2991
32	0.1038	0.0860	0.0178	0.1350	−0.0312
33	−0.1871	−0.2460	0.0589	−0.2250	0.0379
34	−0.0706	0.2230	−0.2936	0.2770	−0.3476
35	0.3945	0.3090	0.0855	0.3800	0.0145

Experimental

Inhibition of CAII. The inhibition value logK(CAII) was adopted from the work of Supuran and Clare.[3]

Molecular graphs. The hydrogen suppressed molecular graphs[1,2] were used for the calculation of topological indices W, Sz, MRI, $^{p1}\chi^{v}$, $^{1}\chi$, J, logRB (Table 2).

Topological indices

Wiener index (W). The Wiener index (W) is a widely used topological index.[9] It is based on the vertex-distances of the respective molecular graph.

The molecular graph can be denoted by G and having $v_1, v_2, v_3, \ldots, v_n$ 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:

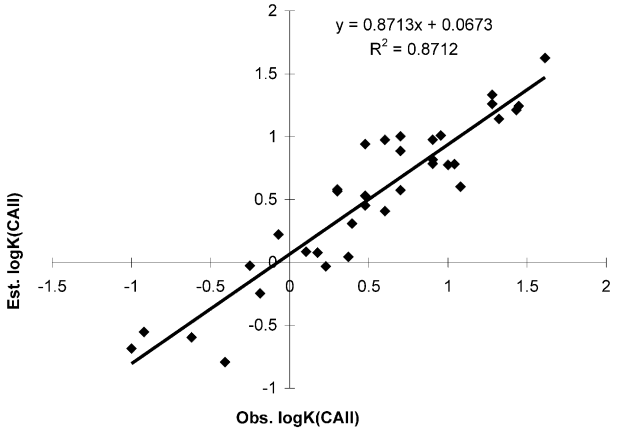


Figure 1. Comparison between observed and estimated logK(CAII) activity using model 29.

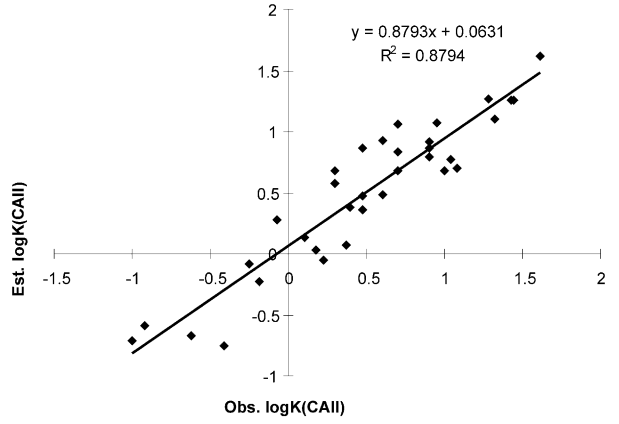


Figure 2. Comparison between observed and estimated logK(CAII) activity using model 32.

$$W = W(G) = 1/2\sum_{i=j}^{n}\sum_{j=1}^{n} d(v_i, v_j \mid G) \tag{7}$$

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[18,19] (Sz) is defined as:

$$Sz(G) = Sz = \sum_{e} n_1(e \mid G) n_2(e \mid G) \tag{8}$$

with the summation going over all the 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.

First-order connectivity index ($^1\chi$). The connectivity index $\chi = \chi(G)$ of a graph G is defined by Randic[21] as under:

$$\chi = \chi(G) = \sum_{ij} [\delta_i \delta_j]^{-0.5} \tag{9}$$

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[22,23] as follows:

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

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

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_j - 1} \tag{11}$$

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.

LogRB. The logRB has been calculated by the method as described in the literature.[25]

Balaban index (J). The Balaban index, J (the average distance sum connectivity index) is defined[20] by:

$$J = J(G)\frac{M}{\mu + 1}\sum_{\text{bonds}} (d_i d_j)^{-1/2} \tag{12}$$

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

$$d_i = \sum_{j=1}^{N} (D)_{ij} \tag{13}$$

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:

$$\mu = M - N + 1 \tag{14}$$

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

(i) The diagonal elements:

$$(D)_{ij} = 1 - (Z_c/Z_i) \tag{15}$$

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

(ii) The off-diagonal elements:

$$(D)_{ij} d_i = \sum_r k_r \tag{16}$$

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.

Molecular redundancy index (MRI). A message or information in the form of electron probability fields distributed around in space in a frame work of atomic radii has lead to introduction of molecular descriptor referred to as molecular redundancy index (MRI),[24] derived from information theory and molecular graph theory.[3] MRI is a molecular symmetry descriptor and indicates the capacity and symmetry of a molecule and can be computed as:

$$\mathrm{MRI} = \frac{\Sigma n_i \log n_i}{N \log N} \tag{17}$$

where n is the number of atoms of the same kind in the ith atom set, i is the number of different atoms sets and $N = \Sigma n_i$ is the total number of atoms in the molecule.

Eq 16 shows that calculation of MRI leads to quantification of the information content. It encodes the salient steric properties of the molecules in cases where biological activity is non specific. It ranks them correctly according to non-specific biological potency and thus, provides mechanistic interpretation of drugs at molecular level based on probability consideration.

Regression analysis. Maximum R^2 improvement method to identify prediction models.[14] 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.

Multiple regression analyses for correlating CAII inhibition activity 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 modeling the CAII inhibition activity of the compounds.

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 are also thankful to CSIR, New Delhi, India for sanctioning a research scheme.

References and Notes

1. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000.
2. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gorden & Breach: Williston, VT, 2000.
3. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1998**, *33*, 489.
4. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1995**, *30*, 687.
5. Clare, B. W.; Supuran, C. T. *Eur. J. Med. Chem.* **1997**, *32*, 311.
6. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1999**, *34*, 41.
7. Clare, B. W.; Supuran, C. T. *Eur. J. Med. Chem.* **1999**, *34*, 4643.
8. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 1179.
9. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
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.; Shrivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287.
13. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
14. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.

15. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
16. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
17. Todeschini, R., *Chemometrics Web News*; Milano Chemometric & QSAR Research Group, file://C\Windows/Desktop/Web News on Chemometrics.html, Feb. 2001.
18. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
19. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
20. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
21. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
22. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Relationship*; Wiley: New York, 1986.
23. Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic: New York, 1999.
24. Kier, L. B. *J. Pharm. Sci.* **1980**, *69*, 807.
25. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.