Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 40 (2005) 1002–1012
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

QSAR study on carbonic anhydrase inhibitors: water-soluble sulfonamides incorporating β-alanyl moieties, possessing long lasting-intra ocular pressure lowering properties—a molecular connectivity approach

Vijay K. Agrawal [a,*], Meghna Banerji [a], Madhu Gupta [b], Jyoti Singh [a], Padmakar V. Khadikar [c,*], Claudiu T. Supuran [d]

[a] *QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa 486 003, India*
[b] *Department of Chemistry, MMH P.G. College, Ghaziabad, India*
[c] *Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd., 3 Khatipura, Indore 452 007, India*
[d] *Laboratorio di Chimica Bioinorganica, Dipartmento di Chimica, University of Florence, Via della Lastruccia, 3, RM.188, Polo Scientifico, 50019 Sesto Fiorentino, Firenze, Italy*

Received 20 November 2004; received in revised form 8 April 2005; accepted 13 April 2005

Available online 14 June 2005

Dedicated to Professor P.V. Khadikar on his 70th birthday

Abstract

A QSAR study on a series of carbonic anhydrase (CA, EC 4.2.1.1) inhibitors, and more precisely on water-soluble sulfonamides incorporating β-alanyl moieties, possessing long lasting intra-ocular pressure lowering properties, was carried out using a series of distance-based topological indices. The regression analysis has shown that out of the pool of topological indices used, the $^{1}\chi$ (first-order Randic connectivity index) is the best one for modeling CA inhibitory properties against all three investigated isozymes, the cytosolic CA I, CA II and the membrane-bound CA IV, and that excellent results are obtained in multiparametric regressions. The results are critically discussed on the basis of statistical parameters.

Keywords: QSAR; Carbonic anhydrase inhibitor; Topological index; Molecular connectivity index; Regression analysis

1. Introduction

The studies on quantitative structure–activity relationship (QSAR) for relating the chemical or biological activity of drugs to their structure has led to a proliferation of methods and indices [1–5]. Such methods include various forms of univariate or multivariate regressions as well as essentially non-linear methods such as back-propagation neural networks. At a fundamental level, indices include pure graph invariants such as those proposed by Wiener [6], Randic [7], Kier and Hall [8,9]. There have been a large number of other indices, which attempted to capture the essence of molecular shape, reactivity and polarity. Some other type of such indices are: Szeged [10,11], Padmakar–Ivan (PI) [12] and Sadhana (Sd) [13]-indices. Recently, Mandloi et al. [14], Basak and Mills [15], Randic and Trinajstic [16] have thoroughly reviewed many of these indices. On the other hand, Kirby [17] has noted that many topological indices have a value that depends on the size of a molecule and because it is on this simple property that important physiochemical properties depend, these indices are widely used in QSAR studies.

In our laboratories, exhaustive work has been carried out for investigating the role of topological indices in developing QSAR as well as for drug design. The indices used being chiefly distance-based topological indices. The methodology used by us was to use initially a large pool of such indices

* Corresponding author. Tel.: +91 94 2519 4556.
E-mail addresses: vijay-agrawal@lycos.com (V.K. Agrawal), pvkhadikar@rediffmail.com (P.V. Khadikar).

0223-5234/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2005.04.006

and then to discover the most useful index. In many cases we observed that a use of single index was not fruitful and we have to use combination of topological indices. In one such attempts on the topological investigation of carbonic anhydrase (CA, EC 4.2.1.1) inhibitory properties of sulfonamide type of drugs, we observed that a single topological index (first-order valence connectivity index, $^{1}\chi^{V}$) in combination with indicator parameters yielded model(s) with excellent statistics. This has prompted us to undertake the present investigations in which we used the Randic molecular connectivity approach. This molecular connectivity approach, which is a method of quantifying the molecular structure of a compound, was introduced by Randic [7] and extensively developed by Kier and Hall [8,9]. The method describes the structure of a molecule by a set of molecular (Randic) connectivity indices, $^{m}\chi$, which encode information about the size, branching, cyclization, unsaturation, etc. In the present study we have initially used a large pool of distance-based topological indices (Table 2) and then arrived at the conclusion that $^{1}\chi$ in combination with indicator parameter and/or the Balaban [18] index yielded excellent models for modeling CA inhibitory properties of a series of compounds designed as topically acting antiglaucoma agents (Table 1).

At this stage it is interesting to mention that several thousand different aromatic/heterocyclic sulfonamide CA inhibitors have been synthesized in the last 50 years in the search for diverse pharmacological agents [19,20], but the number of amino acyl/oligopeptidyl such derivatives is unexpectedly small. One of our groups has recently reported [21] such a series of water-soluble sulfonamides incorporating β-alanyl moieties, possessing long lasting-intraocular pressure lowering properties via the topical route in an animal model of glaucoma (Table 1). Such an interesting set of CA inhibitors was synthesized by the reaction of 26 aromatic/heterocyclic sulfonamides containing amino, imino, hydrazino or hydroxyl groups with *N*-tetrabutyloxycarbonyl-β-alanyl (Boc-β-ala; Boc = t-butoxycarbonyl) in the presence of carbodiimide derivatives. In this paper [21], these derivatives were assayed as inhibitors of three CA isozymes, i.e. CA I, II (cytotsolic forms) and IV (membrane-bound form), involved in important physiological process. Good inhibition was observed against all three isozymes, but especially against CA II and CA IV (in the nano-molar range), the two isozymes known to play a critical role in aqueous humor secretion within the ciliary processes of the eye. However, up to now, no attempt was made to investigate the QSAR of these new sets of CA inhibitors. This is, therefore, the primary aim of our present investigation.

2. Results and discussion

The set of 52 CA inhibitors, first synthesized by Supuran et al. [21] and used in the present investigation is presented in Table 1. The CA inhibition data of the sulfonamides against isozymes I, II and IV and the indicator parameters used are given in Table 2. In attempting QSAR, these inhibition data were converted into their log units and in that form they are also presented in Table 2. The calculated values of the topological indices used: Wiener [6], Szeged [10,11], Balaban [18], Randic [7], and Branching [22] indices, respectively, are given in Table 3. In addition, we have used four indicator parameters: IP_1, IP_2, IP_3, IP_4 for the presence (1) or absence (0) of halogen, five-member ring, methyl group and β-alanyl respectively (Table 2).

The intercorrelations of the topological indices and the indicator parameters as well as their correlations with the inhibitory properties: log K_1(hCA I), log K_1(hCA II), log K_1(hCA IV) are presented in Table 4. The perusal of Table 4 shows that out of the five distance-based topological indices used $^{1}\chi$ (first-order Randic connectivity index) is the best for modeling all the three inhibitory properties. Also, among the three indicator parameters used, IP_2 correlates the best with all the three inhibitory properties. This indicates that the combinations of $^{1}\chi$ and IP_2 with other topological indices and/or indicator parameters may yield models with excellent statistics.

The preliminary regression analysis [22] have indicated that out of the 52 compounds, three compounds: **17**, **26** and **52** are outliers. At this stage we can not explain the reason for this behavior but it may be due to their structural properties. They are, therefore, deleted from the further regression procedure. Our QSAR study, therefore, pertain to 49 compounds only. Also, stepwise regression attempted by us indicated that the penta-parametric regression gave excellent statistics in modeling all the three inhibitory properties mentioned above. For the relative comparison of the results, we have attempted mono- to penta-parametric regressions using the following combinations: (1) $^{1}\chi$; (2) $^{1}\chi$, IP_2; (3) $^{1}\chi$, IP_1, IP_2; (4) $^{1}\chi$, J, IP_1, IP_2; and (5) $^{1}\chi$, J, IP_1, IP_2, IP_3. The reason for using these sets being that they are the best among several regression attempted under a particular category i.e. mono-, di-, tri-, tetra- and penta-parametric regressions.

2.1. Modeling of K_1 (hCA-I) inhibitory property

The use of simple regression analysis yielded the following mono-parametric regression expression containing $^{1}\chi$.

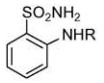
$$\log K_1(\text{hCA I}) = 6.4759 - 0.4563(\pm 0.0558)\ {}^{1}\chi \qquad (1)$$

$n = 49$, S.E. = 0.7814, $r = -0.7664$, $F = 66.885$.

The negative sign of the coefficient of $^{1}\chi$ indicates that the activity (inhibitory property) is inversely proportional to $^{1}\chi$. Since $^{1}\chi$ accounts for number of atoms and first-order branching, the negative sign of $^{1}\chi$ indicates that log K_1(hCA-I) increases with decrease in number of atoms and first-order branching.

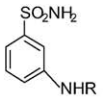
In the above Eq. (1), *n* – is the number of compounds, S.E. – the standard error of estimation, *r* – the simple correlation coefficient (*R* is used in multiple correlation), and *F* – is the *F*-statistics. In the following regression expressions also they follow the same meaning.

Table 1
Structural details of carbonic anhydrase used in the present study



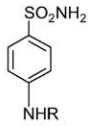
1.R=H

27. R= β-alanyl



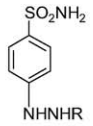
2. R=H

28. R= β-alanyl



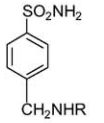
3. R=H

29. R= β-alanyl



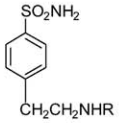
4. R=H

30. R= β-alanyl



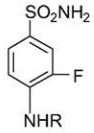
5. R=H

31. R= β-alanyl



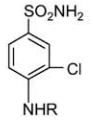
6. R=H

32. R= β-alanyl



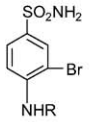
7. R=H

33. R= β-alanyl



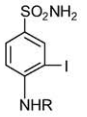
8. R=H

34. R= β-alanyl



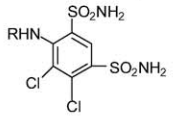
9. R=H

35. R= β-alanyl



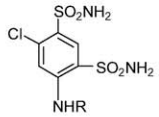
10. R=H

36. R= β-alanyl



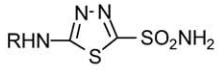
11. R=H

37. R= β-alanyl



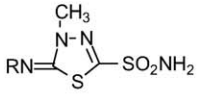
12. R=H

38. R= β-alanyl



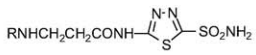
13. R=H

39. R= β-alanyl



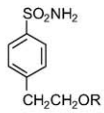
14. R=H

40. R= β-alanyl



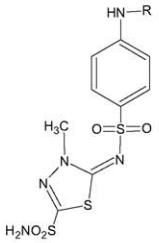
15. R=H

41. R= β-alanyl



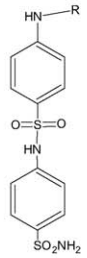
16. R=H

42. R= β-alanyl



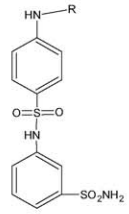
17. R=H

43. R= β-alanyl



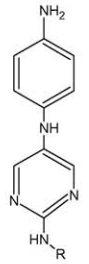
18. R=H

44. R= β-alanyl



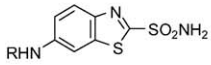
19. R=H

45. R= β-alanyl



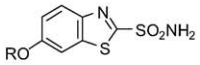
20. R=H

46. R= β-alanyl



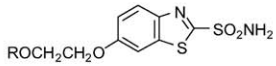
21. R=H

47. R= β-alanyl



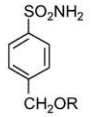
22. R=H

48. R= β-alanyl



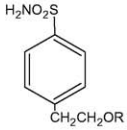
23. R=H

49. R= β-alanyl



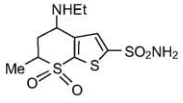
24. R=H

50. R= β-alanyl



25. R=H

51.R= β-alanyl



26. R=H (Dorzolamide)

52. R= β-alanyl

Table 2
The inhibition activity and indicator parameters for the carbonic anhydrase inhibitors used in the present study (Ref. Table 1)

Compound numbers	R	log K_1 (hCA-I)	log K_1 (hCA-II)	log K_1 (hCA-IV)	IP_1	IP_2	IP_3	IP_4
1.	H	4.6571	2.4698	3.1173	0	0	0	0
2.	H	4.3979	2.3802	3.3424	0	0	0	0
3.	H	4.4472	2.4771	3.4771	0	0	0	0
4.	H	4.8949	2.5052	3.5072	0	0	0	0
5.	H	4.3979	2.2304	3.4472	0	0	0	0
6.	H	4.3222	2.2041	3.3892	0	0	0	0
7.	H	3.9191	1.7782	2.2553	1	0	0	0
8.	H	3.9912	2.0414	2.5052	1	0	0	0
9.	H	3.8129	1.6021	1.8195	1	0	0	0
10.	H	3.7853	1.8451	2.0969	1	0	0	0
11.	H	3.7853	1.4472	2.2430	1	0	0	0
12.	H	3.9243	1.8751	2.2041	1	0	0	0
13.	H	3.9345	1.7782	2.7324	0	1	0	0
14.	H	3.9685	1.2787	2.5502	0	1	1	0
15.	H	2.6580	0.4771	2.0969	0	1	0	0
16.	H	0.7782	0.3010	0.6990	0	1	0	0
17.	H	0.0000	-0.2219	-0.0969	0	1	1	0
18.	H	1.6232	0.7782	1.6990	0	0	0	0
19.	H	1.6435	0.9542	1.7243	0	0	0	0
20.	H	2.8388	1.0792	2.1875	0	0	0	0
21.	H	1.8451	0.9542	1.2798	0	1	0	0
22.	H	1.7404	0.9031	1.2304	0	1	0	0
23.	H	1.6990	0.8451	1.1761	0	1	0	0
24.	H	4.3802	2.0969	2.7482	0	0	0	0
25.	H	4.2553	2.0414	2.6532	0	0	0	0
26.	H	4.6990	0.9542	1.6532	0	1	1	0
27.	β-Alanyl	4.3617	2.3010	2.4065	0	0	0	1
28.	β-Alanyl	4.2455	2.2798	2.3674	0	0	0	1
29.	β-Alanyl	4.1761	2.0607	2.2304	0	0	0	1
30.	β-Alanyl	4.3979	2.3729	2.4983	0	0	0	1
31.	β-Alanyl	3.2041	1.5563	1.8865	0	0	0	1
32.	β-Alanyl	2.9320	1.5315	1.8513	0	0	0	1
33.	β-Alanyl	2.7404	1	1.6232	1	0	0	1
34.	β-Alanyl	2.7781	1.5798	1.8573	1	0	0	1
35.	β-Alanyl	2.7959	1.5563	1.8451	1	0	0	1
36.	β-Alanyl	2.7404	1.6021	1.8388	1	0	0	1
37.	β-Alanyl	2.6981	1.0414	1.7324	1	0	0	1
38.	β-Alanyl	2.5441	1	1.6990	1	0	0	1
39.	β-Alanyl	2.4814	0.9542	1.6232	0	1	0	1
40.	β-Alanyl	2.5185	1.0414	1.6628	0	1	1	1
41.	β-Alanyl	1.4472	0.6990	1.2553	0	1	0	1
42.	β-Alanyl	1.1139	0.4771	1.0792	0	1	0	1
43.	β-Alanyl	1.1761	0.6021	1.2041	0	1	1	1
44.	β-Alanyl	1.8388	1.1761	1.9031	0	0	0	1
45.	β-Alanyl	1.8325	1.1461	1.8751	0	0	0	1
46.	β-Alanyl	1.7559	1.2788	1.9138	0	0	0	1
47.	β-Alanyl	1.1761	0.6990	1.0792	0	1	0	1
48.	β-Alanyl	1.2041	0.6021	1.1461	0	1	0	1
49.	β-Alanyl	0.9542	0.4771	1.1394	0	1	0	1
50.	β-Alanyl	3.2304	1.8451	2.2670	0	0	0	1
51.	β-Alanyl	3.2304	1.7924	2.1761	0	0	0	1
52.	β-Alanyl	1.9731	0.6990	1.4771	0	1	0	1

Table 3
Distance based before topological indices calculated for the carbonic anhydrase inhibitors used in the present study (Ref. Table 1)

Compound numbers	*W*	$^{1}\chi$	*J*	Sz	log RB
1.	144	5.0159	2.5451	220	46.5485
2.	148	4.9990	2.4607	228	47.5056
3.	152	4.9990	2.3936	236	48.2757
4.	201	5.5370	2.3588	306	63.1776
5.	201	5.5370	2.3588	306	63.1776
6.	262	6.0370	2.3049	388	80.9368
7.	189	5.4097	2.5123	291	60.5587
8.	189	5.4097	2.5123	291	60.5587
9.	189	5.4097	2.5123	291	60.5587
10.	189	5.4097	2.5123	291	60.5587
11.	458	7.4592	2.9912	668	147.7629
12.	399	7.0317	2.8525	582	128.1666
13.	113	4.4990	2.4489	132	35.9157
14.	146	4.9097	2.5376	171	46.8124
15.	403	6.9309	2.3042	452	121.6654
16.	853	9.1825	1.8612	1124	248.1982
17.	948	9.5932	1.9370	1237	277.0616
18.	1004	9.6825	1.8165	1502	28.9915
19.	360	9.6825	1.8996	1414	279.3234
20.	687	8.5378	1.3774	964	196.3445
21.	287	6.4654	1.9874	430	90.6985
22.	287	6.4654	1.9874	430	90.6985
23.	543	8.0034	1.8560	776	162.4138
24.	192	5.3123	2.4689	297	60.9234
25.	262	6.0370	2.3049	388	80.9368
26.	634	8.6749	2.2223	918	199.007
27.	459	7.6851	2.4597	600	141.3952
28.	470	7.4309	2.3860	635	142.4736
29.	494	7.4309	2.2676	683	145.9187
30.	610	7.9309	2.1912	820	175.0800
31.	610	7.9309	2.1912	820	175.0800
32.	743	8.4309	2.125	974	207.1639
33.	558	7.8416	2.4255	770	166.5446
34.	558	7.8416	2.4255	770	166.5446
35.	558	7.8416	2.4255	770	166.5446
36.	558	7.8416	2.4255	770	166.5446
37.	807	9.0529	2.7568	1088	243.2091
38.	896	9.4636	2.8807	1214	271.0257
39.	403	6.9309	2.3047	452	121.6654
40.	463	7.3416	2.4521	518	140.9050
41.	1003	9.3247	2.1554	1082	272.7856
42.	1723	11.6143	1.7761	2264	456.1391
43.	1875	12.025	1.8336	2449	497.1511
44.	1996	12.0967	2.3924	2455	515.7920
45.	1940	12.0967	2.4731	2399	509.3550
46.	1129	10.3752	1.5221	1715	319.0601
47.	755	8.8972	1.8351	1048	219.8339
48.	755	8.8972	1.8351	1048	219.8339
49.	1272	10.3972	1.6521	1655	339.3797
50.	610	7.9309	2.1912	820	175.0804
51.	743	8.4309	2.1251	974	207.1539
52.	1212	11.0343	2.2933	1651	365.7268

During the successive regression analysis we obtained several bi-parametric regression expressions out of which the one containing $^{1}\chi$ and IP_2 gave the best result. This is found as below:

$$\log K_1(\text{hCA I}) = 6.5421 - 0.4148(\pm 0.0377)\ {}^{1}\chi - 1.2582(\pm 0.1636)IP_2 \quad (2)$$

$n = 49$, S.E. = 0.5225, $R = 0.9052$, $F = 104.363$.

The comparison of Eqs. (1) and (2) indicates that there is a tremendous improvement in the quality of regression such that (*r*) *R*-value changes from –0.7664 to 0.9052 and standard error of estimation decreases from 0.7814 to 0.5225. Also, that here also the sign of the coefficient of $^{1}\chi$ is negative in both the cases. In addition, the coefficient of IP_2 in Eq. (2) is also negative. We have used IP_2 for demonstrating presence/absence of heterocyclic ring in the sulfonamide moiety. The negative sign of IP_2 indicates that the presence of a heterocyclic ring is not favorable for the exhibition of log K_1(hCA I) activity. The meaning of the negative coefficient of $^{1}\chi$ is the same as discussed for Eq. (1).

The stepwise regression resulted into several tri-parametric regression expressions out of which the one containing $^{1\prime}\chi$, IP_1 and IP_2 was found the best:

$$\log K_1(\text{hCA I}) = 6.7621 - 0.4239(\pm 0.0361)\ {}^{1}\chi - 0.4307(\pm 0.1795)IP_1 - 1.4046(\pm 0.1673)IP_2 \quad (3)$$

$n = 49$, S.E. = 0.4974, $R = 0.9165$, $F = 78.695$.

We observe that by the addition of IP_1 to regression expression expressed by Eq. (2), the quality of correlation is improved. However, the improvement is not that large when we arrived at Eq. (2). Once again the coefficients of $^{1}\chi$ and IP_2 are negative, therefore, their significance is the same as discussed above. In addition, the coefficient of IP_1 term is also negative. This indicator parameter is used to account for the presence/absence of halogen in the sulfonamide moiety. The negative coefficient of IP_1 indicates that the absence of halogen is not favorable for the exhibition of the activity.

Further stepwise regression indicated that tetra-parametric regression expression containing $^{1}\chi$, *J*, IP_1 and IP_2 yielded the better statistics than what we have observed before:

$$\log K_1(\text{hCA I}) = 3.7166 - 0.3551(\pm 0.0330)\ {}^{1}\chi + 1.1389(\pm 0.2402)J - 0.8425(\pm 0.1713)IP_1 - 1.2586(\pm 0.1410)IP_2 \quad (4)$$

$n = 49$, S.E. = 0.4092, $R = 0.9455$, $F = 92.817$.

Once again we observed that there is quite a good improvement in the quality of statistics when *J* index is added to Eq. (3). Here also, the coefficients of $^{1}\chi$, IP_1 and IP_2 are negative and thus they carry the same significance as above. However, the added Balaban index (*J*) has positive coefficient.

Thus, the increase in the magnitude of *J* increases the exhibition of log K_1(hCA I) activity.

Table 4
Correlation matrix showing correlation among various topological indices, activity and indicator parameters

	log K_1(hCA-I)	log K_1(hCA-II)	log K_1(hCA-IV)	*W*	$^1\chi$	*J*	Sz	log-RB	IP_1	IP_2	IP_3	IP_4
log K_1(hCA-I)	1.0000											
log K_1(hCA II)	0.8921	1.0000										
log K_1 (hCA IV)	0.8713	0.8918	1.0000									
W	–0.6533	–0.5453	–0.5254	1.0000								
$^1\chi$	–0.7311	–0.6368	–0.6434	0.9372	1.0000							
J	0.6009	0.4945	0.4580	–0.3726	–0.4574	1.0000						
Sz	–0.6797	–0.3378	–0.3342	0.9707	0.9659	–0.4365	1.0000					
log RB	–0.6444	–0.5415	–0.5406	0.9377	0.9272	–0.3422	0.9418	1.0000				
IP_1	0.1548	0.1034	–0.0106	–0.1838	–0.1632	0.5779	–0.1828	–0.1544	1.0000			
IP_2	–0.5468	–0.6939	–0.5893	0.1779	0.1758	–0.4062	0.1266	0.1903	–0.3922	1.0000		
IP_3	–0.1223	–0.3351	–0.2695	0.1398	0.1260	–0.0539	0.1181	0.1535	–0.1829	0.4663	1.0000	
IP_4	–0.3089	–0.1474	–0.3077	0.5558	0.5796	–0.0910	0.5048	0.5714	0.0109	–0.0277	–0.0595	1.0000

Finally, adding IP_3 to the Eq. (4) yielded a penta-parametric model with slight improvement in the quality of regression:

$$\log K_1(\text{hCA I}) = 4.0181 - 0.3617(\pm 0.0323)\ ^1\chi + 1.0254(\pm 0.2407)J - 0.8013(\pm 0.1678)IP_1 - 1.3744(\pm 0.1498)IP_2 + 0.5054(\pm 0.2647)IP_3 \quad (5)$$

$n = 49$, S.E. = 0.3974, $R = 0.9499$, $F = 79.447$.

Thus, by the addition of IP_3 term to Eq. (4), *R* increases from 0.9455 to 0.9499 and S.E. falls down from 0.4092 to 0.3979. Thus, such a slight improvement in the statistics is at the cost of IP_3 which has a positive coefficient in Eq. (5). This indicator parameter accounts for the presence/absence of β-alanyl, the positive coefficient of which indicates that the presence of β-alanyl is favorable for the exhibition of log K_1(hCA I) activity.

No other higher parameteric regression gave better results than that of Eq. (5). The regression parameters and quality of correlation for modeling log K_1 (hCA-I) activity are given in Table 5.

2.2. Modeling log K_1(hCA II) activity

As stated earlier, and for making exact comparison of the results, we have used the similar regression procedures as above for modeling log K_1(hCA II) activity. The results are given in Table 6. Since here also the penta-parametric regression containing $^1\chi$, *J*, IP_1, IP_2 and IP_3 yielded the best regression, the regression expression for the sake of exact comparison is given below:

$$\log K_1(\text{hCA II}) = 2.2447 - 0.1593(\pm 0.0219)\ ^1\chi + 0.3728(\pm 0.1656)J - 0.5423(\pm 0.1140)IP_1 - 0.9305(\pm 0.1018)IP_2 + 0.1013(\pm 0.1799)IP_3 \quad (6)$$

$n = 49$, S.E. = 0.2700, $R = 0.9166$, $R^2A = 0.8215$, $F = 45.172$.

This regression Eq. (6) suffers from the defect that the coefficient of IP_3 term is less than its standard deviation. Such expressions (models) are not allowed statistically. Furthermore, the data presented in Table 6 shows that as we pass from the tetra-parametric regression containing $^1\chi$, *J*, IP_1 and

Table 5
Regression parameters and quality of correlation for modeling log K_1(hCA I) activity

Model (Eqs.)	Parameter involved	S.E.	*R*(*r*)	R^2_A	*F*	*Q*
1.	$^1\chi$	0.7814	–0.7664	–	66.885	–0.9808
2.	$^1\chi$, IP_1	0.5225	0.9052	0.8116	104.363	1.7324
3.	$^1\chi$, IP_1, IP_2	0.4974	0.9165	0.8292	78.695	1.8426
4.	$^1\chi$, *J*, IP_1, IP_2	0.4092	0.9455	0.8844	92.817	2.3106
5.	$^1\chi$, *J*, IP_1, IP_2, IP_3	0.3974	0.9499	0.8910	79.447	2.3903

Table 6
Regression parameters and quality of correlation for modeling log K_1(hCA II) activity

Model (Eqs.)	Parameter involved	S.E.	*R*(*r*)	R^2_A	*F*	*Q*
6.	$^1\chi$	0.4997	–0.6336	–	31.520	–1.2680
7.	$^1\chi$, IP_1	0.3260	0.8664	0.7398	69.251	2.6577
8.	$^1\chi$, IP_1, IP_2	0.2834	0.9032	0.8034	66.401	3.1870
9.	$^1\chi$, *J*, IP_1, IP_2	0.2679	0.9159	0.8242	57.275	3.4188
10.	$^1\chi$, *J*, IP_1, IP_2, IP_3	0.2700	0.9166	0.8215	45.172	3.3948

Table 7
Regression parameters and quality of correlation for modeling log K_1(hCA IV) activity

Model (Eqs.)	Parameter involved	S.E.	$R(r)$	R^2_A	F	Q
1.	$^1\chi$	0.5342	–0.6486	–	34.123	–1.2142
2.	$^1\chi$, IP_1	0.4210	0.8049	0.6325	42.305	1.9119
3.	$^1\chi$, IP_1, IP_2	0.3404	0.8802	0.7598	51.604	2.5858
4.	$^1\chi$, J, IP_1, IP_2	0.3110	0.9035	0.7995	48.862	2.9051
5.	$^1\chi$, J, IP_1, IP_2, IP_3	0.3080	0.9077	0.8034	40.223	2.9471

IP_2 to penta-parametric regression having, $^1\chi$, J, IP_1, IP_2 and IP_3 as the correlating parameters, the adjusted R^2, R^2_A (= 0.8242) lowers down to 0.8215. It means that the added parameter IP_3 does not have its contribution in exhibiting log K_1(hCA II) activity. This means that the presence of methyl group is not required for the exhibition of the activity. Also, that instead of penta-parametric regression, the regression is the tetra-parametric regression:

$$\log K_1(\text{hCA II}) = 2.1843 - 0.1580(\pm 0.0216)\ ^1\chi + 0.3955(\pm 0.1573)J - 0.5805(\pm 0.1122)IP_1 - 0.9073(\pm 0.0923)IP_2 \quad (7)$$

$n = 49$, S.E. = 0.2679, $R = 0.9159$, $R^2_A = 0.8242$, $F = 57.275$.

In both the above Eqs. (6) and (7) the coefficient of J is positive, while that of $^1\chi$, IP_1 and IP_2 are negative. Thus, the physical significances of these regression expressions are the same as discussed for modeling log K_1(hCA I) activity.

2.3. Modeling of log K_1 (hCA-IV) activity

The regression parameters and quality of correlations for the different regression expressions, starting from simple to penta-parametric regressions, for modeling log K_1 (hCA-IV) activity are given in Table 7. Once again, we observed that the penta-parametric model containing $^1\chi$, J, IP_1, IP_2 and IP_3 yielded the best results:

$$\log K_1(\text{hCA IV}) = 2.7452 - 0.1811(\pm 0.0250)\ ^1\chi + 0.5123(\pm 0.1866)J - 0.8039(\pm 0.1300)IP_1 - 0.9189(\pm 0.1161)IP_2 + 0.2796(\pm 0.2052)IP_3 \quad (8)$$

$n = 49$, S.E. = 0.3080, $R = 0.9077$, $R^2_A = 0.8034$, $F = 40.223$.

At this stage it is interesting to comment upon the statistical acceptability of the proposed equations. It is correct that the equation with coefficient of the involved term less than its standard deviation is not allowed statistically. This is correct, but does not go far enough. It is traditional in sciences that a term is not acceptable to its t-probability exceeds 0.05 (i.e. 95% confidence). This is so only if the coefficient of the term exceeds t-times, not 1 time its standard deviation. The t term is Student's t, and is a function of probability and the degree of freedom. For probability of 0.05 and degrees of freedom large, t is approximately 2. On this basis, therefore, along with Eq. (5), Eq. (6) and Eq. (8) are also permitted statistically, because of the lack of significance of the same term.

In this regression expression also, J and IP_3 terms have positive coefficients, while the coefficients of the other terms viz. $^1\chi$, IP_1 and IP_2 are negative. Hence, the physical significance of this model is similar to the earlier models used for the exhibition of log K_1(hCA I) and log K_1(hCA IV) activities.

It is interesting to record that in all the cases, in lower as well as higher parametric models, the coefficients of $^1\chi$ and J terms more or less remain similar. That is, both these topological indices are quite stable. This supports that the choice of these topological indices i.e. $^1\chi$ and J is quite justified. That is, out of the larger pool of the topological indices used, $^1\chi$ and J are more useful for modeling all the three inhibitory activities (properties).

It is interesting to mention that our earlier QSAR studies for modeling CA inhibitory property was based on non-topological molecular descriptors [23–26], while in the present case we have used topological indices [27–33]. Non-topological QSAR studies made by us on several series of CA inhibitors emerged that the enhancement of CA inhibitory activity is correlated with increased positive charge of the heterocyclic/aromatic ring, incorporated in such molecules, as well as with 'long' inhibitor molecules (i.e. molecule extending in the direction passing through the Zn (II) ion of the enzyme, one sulfonamide nitrogen atom and the long axis of the inhibitor [23–26]. The topological modeling [27–30] and the one used herein gives additional information regarding the effect of size, shape, branching, presence of heteroatom on the exhibition of CA inhibitory property/activity of the sulfonamides used.

2.4. Comparison of the results

We can now compare the usefulness of the topological methodology used by us in modeling log K_1(hCA I), log K_1(hCA II) and log K_1(hCA IV) property/activity. We can use the data presented in Tables 5–7 for this purpose. The perusal of these Tables show that our method is most suited for the modeling of log K_1(hCA I) activity. The relative correlating power of our methodology goes on decreasing as we start modeling log K_1(hCA I) to log K_1(hCA IV) activity. Furthermore, interesting results are obtained for modeling log K_1(hCA II) in that tetra-parametric model gave better statistics than the penta-parametric models used in other two cases.

In order to support our findings we have estimated log K_1(hCA I), log K_1(hCA II) and log K_1(hCA IV) activities using the best models and compared their values. Such a comparison is shown in Table 8. Furthermore, we have estimated

Table 8
Comparison of observed and estimated log K_1(hCA I), log K_1(hCA II) and log K_1(hCA IV) activity using models 5, 10, 15 (Refs. Tables 5–7)

Compd No	Model-5 log (hCA I)			Model-10 log (hCA II)			Model-15 log (hCA IV)		
	Obs.	Est.	Residue	Obs.	Est.	Residue	Obs.	Est.	Residue
1.	4.657	4.858	–0.2011	2.470	2.417	0.053	3.117	3.180	–0.063
2.	4.398	4.765	–0.367	2.380	2.383	–0.003	3.342	3.131	0.211
3.	4.447	4.687	–0.24	2.477	2.354	0.123	3.477	3.090	0.387
4.	4.895	4.452	0.443	2.505	2.253	0.252	3.507	2.970	0.537
5.	4.398	4.452	–0.054	2.230	2.253	–0.023	3.447	2.970	0.477
6.	4.322	4.209	0.113	2.204	2.149	0.055	3.389	2.846	0.543
7.	3.919	3.820	0.0993	1.778	1.774	0.004	2.255	2.243	0.012
8.	3.991	3.820	0.171	2.041	1.774	0.267	2.505	2.243	0.262
9.	3.813	3.820	–0.007	1.602	1.774	–0.172	1.820	2.243	–0.423
10.	3.785	3.820	–0.035	1.845	1.774	0.071	2.097	2.243	–0.146
11.	3.785	3.643	0.142	1.447	1.652	–0.205	2.243	2.162	0.081
12.	3.924	3.635	0.289	1.875	1.660	0.215	2.204	2.155	0.049
13.	3.935	3.587	0.348	1.778	1.540	0.238	2.732	2.319	0.413
14.	3.969	4.168	–0.199	1.279	1.455	–0.176	2.550	2.263	0.287
15.	2.658	2.542	0.116	0.477	1.088	–0.611	2.097	1.785	0.312
16.	0.778	1.212	–0.434	0.301	0.536	–0.235	0.699	1.101	–0.402
17.	0	–	0	–0.222	–	–0.222	–0.097	–	–0.097
18.	1.624	2.324	–0.7	0.778	1.354	–0.576	1.699	1.879	–0.178
19.	1.644	2.422	–0.778	0.954	1.390	–0.436	1.724	1.930	–0.206
20.	2.839	2.221	0.618	1.079	1.348	–0.269	2.188	1.819	0.369
21.	1.845	2.338	–0.493	0.954	1.026	–0.072	1.280	1.676	–0.396
22.	1.740	2.338	–0.598	0.903	1.026	–0.123	1.230	1.676	–0.446
23.	1.699	1.631	0.068	0.845	0.723	0.122	1.176	1.313	–0.137
24.	4.380	4.662	–0.282	2.097	2.336	–0.239	2.748	3.079	–0.331
25.	4.255	4.209	0.046	2.041	2.149	–0.108	2.653	2.846	–0.193
26.	4.699	–	4.699	0.945	–	0.945	1.653	–	1.653
27.	4.362	3.797	0.565	2.301	1.952	0.349	2.407	2.639	–0.232
28.	4.246	3.802	0.444	2.280	1.961	0.319	2.367	2.640	–0.273
29.	4.176	3.663	0.513	2.061	1.910	0.151	2.230	2.568	–0.338
30.	4.398	3.394	1.004	2.373	1.797	0.576	2.418	2.429	0.011
31.	3.204	3.394	–0.19	1.556	1.797	–0.241	1.887	2.429	–0.542
32.	2.932	3.136	–0.204	1.532	1.688	–0.156	1.851	2.297	–0.446
33.	2.740	2.843	–0.103	1	1.346	–0.346	1.624	1.745	–0.121
34.	2.778	2.843	–0.065	1.580	1.346	0.234	1.857	1.745	0.112
35.	2.796	2.843	–0.047	1.556	1.346	0.21	1.845	1.745	0.1
36.	2.740	2.843	–0.103	1.602	1.346	0.256	1.839	1.745	0.094
37.	2.698	2.795	–0.097	1.041	1.295	–0.254	1.732	1.726	0.006
38.	2.544	2.792	–0.248	1.000	1.283	–0.283	1.699	1.727	–0.028
39.	2.481	2.543	–0.062	0.954	1.088	–0.134	1.623	1.785	–0.162
40.	2.519	3.192	–0.673	1.041	1.028	0.013	1.663	1.765	–0.102
41.	1.447	1.506	–0.059	0.699	0.640	0.059	1.255	1.255	0.000
42.	1.114	0.237	0.877	0.477	0.109	0.368	1.079	0.603	0.476
43.	1.176	0.781	0.395	0.602	0.011	0.591	1.204	0.528	0.676
44.	1.839	2.130	–0.291	1.176	1.216	–0.04	1.903	1.790	0.113
45.	1.833	2.225	–0.392	1.146	1.251	–0.105	1.875	1.839	0.036
46.	1.756	1.730	0.026	1.279	1.116	0.163	1.914	1.571	0.343
47.	1.176	1.284	–0.108	0.699	0.570	0.129	1.079	1.137	–0.058
48.	1.204	1.284	–0.080	0.602	0.570	0.032	1.146	1.137	0.009
49.	0.954	0.530	0.424	0.477	0.251	0.226	1.139	0.750	0.389
50.	3.230	3.394	–0.164	1.845	1.797	0.048	2.267	2.429	–0.162
51.	3.230	3.136	0.094	1.792	1.688	0.104	2.176	2.429	–0.253
52.	1.973		1.973	0.699		0.699	1.477		1.477

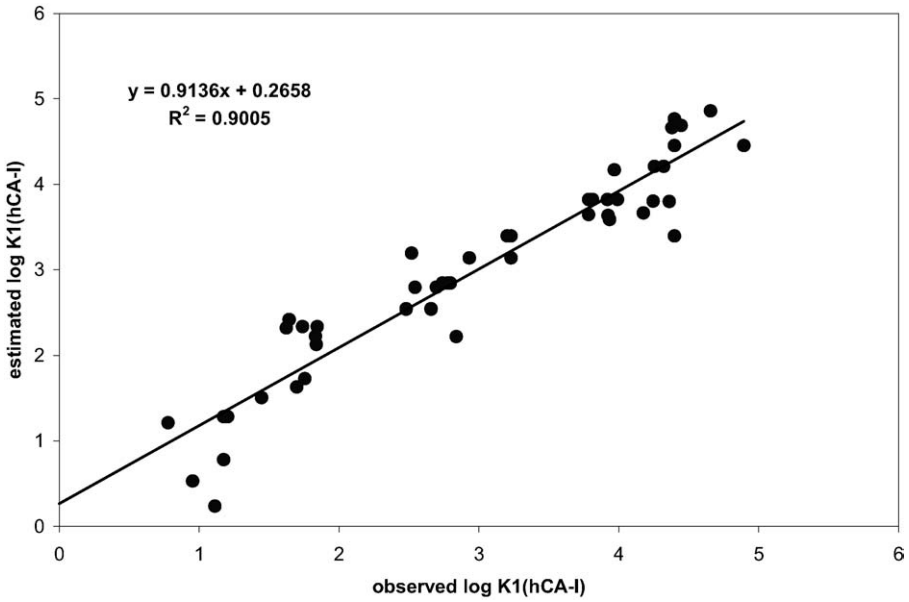


Fig. 1. Correlation of observed and estimated log K_1(hCA I).

predictive correlation coefficients (R^2_{pred}) from the correlation between observed and estimated activities in each case. Such correlations, as demonstrated in Figs. 1–3, yielded R^2_{pred} as 0.9005, 0.8372 and 0.8116, respectively, thus, supporting our results.

In Table 9, we have presented the data characterizing the statistical significance of each of the parameters present in the best regression expressions for modeling log K_1(hCA I), log K_1(hCA II) and log K_1(hCA IV) activities. The statistical significance is described by *t*-value and *t*-probability associated with each parameter. The perusal of Table 9 shows that the most statistically significant parameters for modeling all the activities are $^1\chi$, IP_1 and IP_2.

Further comparison can be made using quality factor, *Q*. This factor is defined [34,35] as the ratio of correlation coefficient (*R*) to the standard error of estimation (S.E.), i.e. $Q = R/\text{S.E.}$ and accounts for the predictive power of the model (regression expression). The calculated value of this parameter is recorded in Tables 5–7. Though the use of quality factor (*Q*) is criticized [36] it is found useful in the present study indicating that the predictive power of the methodology used by us follow the sequence:

$$\log K_1(\text{hCA I}) > \log K_1(\text{hCA II}) > \log K_1(\text{hCA IV})$$

log K_1(hCA I) > log K_1(hCA II) > log K_1(hCA IV).

This confirms our findings discussed above.

Now, it is interesting to comment on R^2_A values obtained for various regression expressions (Tables 5–7). In all the three cases, R^2_A goes on increasing as we pass from bi- to penta-

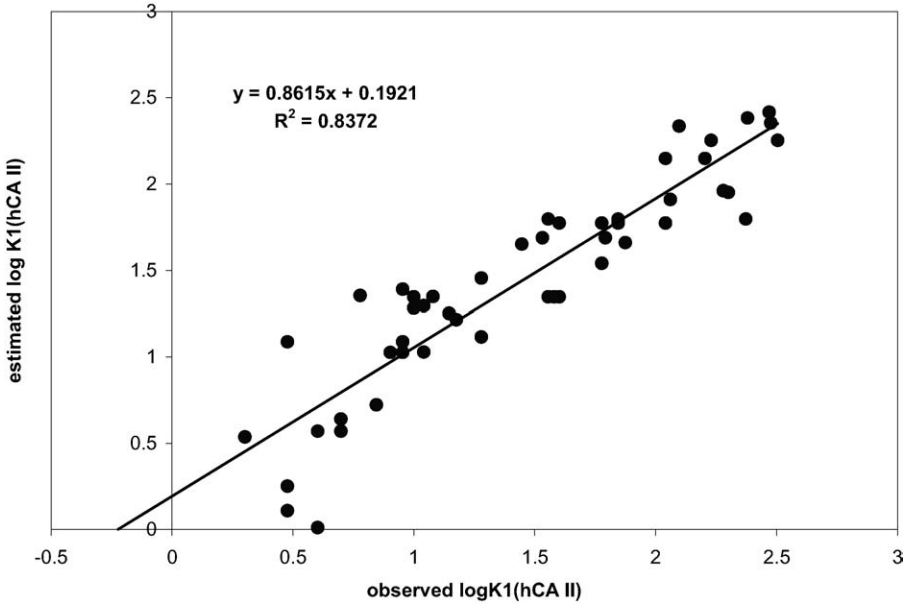


Fig. 2. Correlation of observed and estimated log K_1(hCA II).

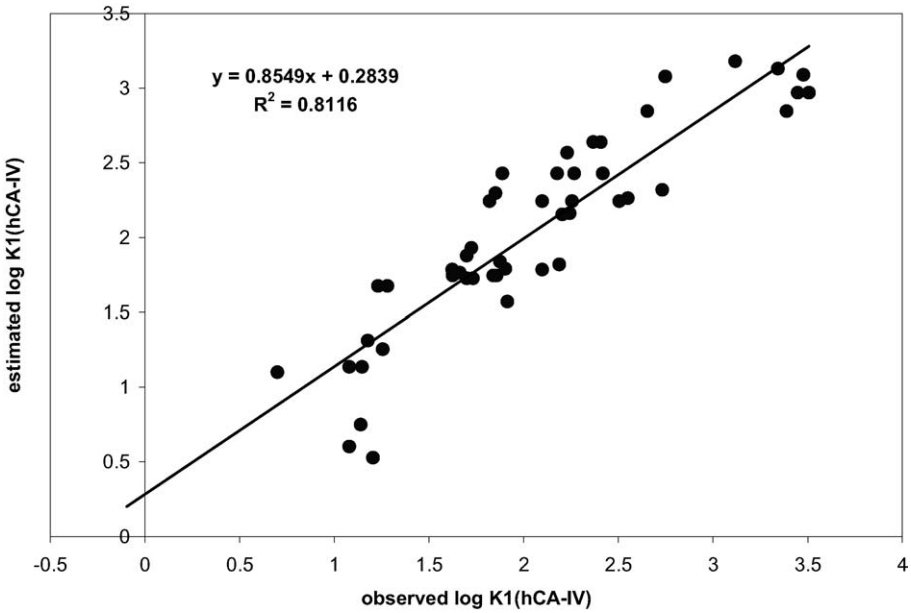


Fig. 3. Correlation of observed and estimated log K_1(hCA IV).

parametric models. Except for modeling of log K_1(hCA II) the value of R^2_A is found the highest for the penta-parametric regression expression. However, for modeling log K_1(hCA II) it is found highest for tetra-parametric model. R^2_A is a measure of % explained variation in the dependent variable, log K_1(hCA I), log K_1(hCA II), log K_1(hCA IV), in the present case, that takes into account the relationship between the number of cases (compounds used) and the number of independent variables, topological indices and indicator parameters used in the regression expression (models); R^2 will always increase when an independent variable is added; R^2_A on the other hand will decrease if the added variable does not reduce the unexplained variation enough to offset the loss of degrees of freedom. Furthermore, decrease in R^2_A with the addition of independent variable indicates that the added variable does not have enough contribution to the proposed model. The data presented in Tables 5–7 indicate that in case of modeling log K_1(hCA I) and log K_1(hCA IV) activities goes on increasing as we pass from mono- to penta-parametric regression expressions. This means that successive addition of the independent variables (topological indices and indicator parameters) have enough potential for increasing the quality of the models (regression expressions). However, the case is different for modeling log K_1(hCA II) activity. Here, the R^2_A value for penta-parametric regression expression is lower than that of tetra-parametric regression expression. This means that addition of IP_3 to the tetra-parametric regression expression does not have any contribution for modeling log K_1(hCA II). However, as stated above R^2 is increased from 0.9159 to 0.9166 by such addition of IP_3.

3. Conclusion

The topological designing of CA inhibitory activity is quite successful for modeling, monitoring and estimating log K_1(hCA I), log K_1(hCA II), log K_1(hCA IV) activities for the set of sulfonamides used in the present study. Out of the larger pools of topological indices used, $^1\chi$ is found to be the best for the purpose which yielded excellent models upon addition of indicator parameters. Our methodology found more useful for modeling log K_1(hCA I) activity as compared to other two activities viz. log K_1(hCA II) and log K_1(hCA IV).

4. Experimental

4.1. CA inhibitory activity

The CA inhibitory activity reported earlier [21] was used after converting it into log units.

Table 9
Statistical significance of the parameter of the best model used for modeling log K_1(hCA I), log K_1(hCA II) and log K_1(hCA IV) activities

Parameter	*t*-Value for			*t*-Probability for		
	log K_1(hCA I)	log K_1(hCA II)	log K_1(hCA IV)	log K_1(hCA I)	log K_1(hCA II)	log K_1(hCA IV)
$^1\chi$	–11.207	–7.265	–7.241	0.00000	0.00000	0.00000
J	4.259	2.279	2.746	0.00011	0.02769	0.00877
IP_1	–4.776	–4.757	–6.183	0.00002	0.00002	0.00000
IP_2	–9.175	–9.142	–7.916	0.00000	0.00000	0.00000
IP_3	1.909	0.563	1.363	0.06292	0.57642	0.18003

4.2. Topological indices

All the topological indices were calculated using hydrogen suppressed molecular graphs in that all the carbon–hydrogen as well as heteroatom–hydrogen bonds were deleted. The calculations of the topological indices were made using the software made available by Professor Istvan Lukovits, Hungarian Academy of Science, Budapest, Hungary. The details for the calculations of these indices are available in the literature and thus they are not given here.

4.2.1. Regression analysis

The regression analysis was carried out by stepwise regression following maximum R^2-method. Here also we have used Regress-1 program of Lukovits.

Acknowledgements

Authors are thankful to Professor Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary for providing software to carry out regression analysis. Authors are also thankful to CSIR New Delhi, India for providing financial support through project No. 01(1785)/02/EMR-II. Authors are also thankful to the referees for their valuable suggestions making the paper in publishable form.

References

[1] M. Karelson, Molecular Descriptors in QSAR/QSPR, Wiley, New York, 2000.
[2] M.V. Diudea, Ed. QSPR/QSAR Studies by Molecular Descriptors, in: Babes-Bolyai University, Cluj, Romania, 2000.
[3] N. Trinajstic, Chemical Graph Theory, CRC Press, Boca Raton, FL, 1992.
[4] R. Todeschini, V. Cosonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[5] J. Devillers, A.T. Balaban, Topological Indices and Related Descriptors in QSAR and QSPR, Gorden & Breach, Williston, VT, 2000.
[6] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17–20.
[7] M. Randic, J. Am. Chem. Soc. 97 (1975) 6609–6615.
[8] L.B. Kier, L.H. Hall, Molecular Connectivity in Structure–activity Relationship, Wiley, New York, 1986.
[9] L.B. Kier, L.H. Hall, Molecular Structure Description, Academic Press, New York, 1999.
[10] I. Gutman, Graph Theor. Notes New York 27 (1994) 9.
[11] P.V. Khadikar, N.V. Deshpande, P.P. Kale, A. Dobrynin, I. Gutman, G. Domotor, J. Chem. Inf. Comput. Sci. 35 (1995) 547.
[12] P.V. Khadikar, S. Karmarkar, V.K. Agrawal, J. Chem. Inf. Comput. Sci. 41 (2001) 934–949.
[13] P.V. Khadikar, V.K. Agrawal, S. Karmarkar, Bioorg. Med. Chem. 10 (2002) 3499.
[14] M. Mandloi, A. Sikarwar, N.S. Sapre, S. Karmarkar, P.V. Khadikar, J. Chem. Inf. Comput. Sci. 40 (2000) 57.
[15] S.C. Basak, D. Mills, Commun. Math. Chem. 44 (MATCH) 4415 (2001).
[16] M. Randic, N. Trinajstic, J. Molec, Stratum 300 (1993) 551.
[17] E.C. Kirby, J. Chem. Inf. Comput. Sci. 34 (1994) 1030.
[18] A.T. Balaban, Chem. Phys. Lett. 89 (1982) 399.
[19] a) C.T. Supuran, A. Scozzafava, A. Casini, Med. Res. Rev. 23 (2003) 146; b) C.T. Supuran, A. Scozzafava, A. Casini, Development of sulfonamide carbonic anhydrase inhibitors (CAIs), in: C.T. Supuran, A. Scozzafava, J. Conway (Eds.), Carbonic Anhydrase, Its Inhibitors and Activators, CRC Press, Boca Raton, FL, 2004, pp. 67–147.
[20] F. Mincione, L. Menabuoni, C.T. Supuran, Clinical applications of the carbonic anhydrase inhibitors in ophthalmology, in: C.T. Supuran, A. Scozzafava, J. Conway (Eds.), Carbonic Anhydrase, Its Inhibitors and Activators, CRC Press, Boca Raton, FL, 2004, pp. 243–254.
[21] C.T. Supuran, F. Briganti, L. Menabuoni, G. Mincione, F. Mincione, A. Scozzafava, Eur. J. Med. Chem. 35 (2000) 309–315.
[22] S. Chaterjee, A.S. Hadi, B. Price, Regression Analysis by Examples, third ed, Wiley, New York, 2000.
[23] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 30 (1995) 687–696.
[24] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 34 (1999) 41–50.
[25] B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 32 (1995) 311–319.
[26] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 33 (1998) 489–497.
[27] V.K. Agrawal, S. Sinha, S. Bano, P.V. Khadikar, Acta Microbiol. 48 (2001) 17–21.
[28] V.K. Agrawal, R. Shrivastava, P.V. Khadikar, Bioorg. Med. Chem. 9 (2001) 3287–3293.
[29] A. Saxena, V.K. Agrawal, P.V. Khadikar, Oxid. Commun. 26 (2003) 9–13.
[30] V.K. Agrawal, R. Sharma, P.V. Khadikar, Bioorg. Med. Chem. 10 (2002) 2993–2999.
[31] V.K. Agrawal, P.V. Khadikar, Bioorg. Med. Chem. 13 (2002) 447–453.
[32] V.K. Agrawal, S. Shrivastava, P.V. Khadikar, C.T. Supuran, Bioorg. Med. Chem. 11 (2003) 5353–5362.
[33] L. Pogliani, Amino Acids 6 (1994) 141–144.
[34] L. Pogliani, J. Phys. Chem. 100 (1996) 18065–18077.
[35] R. Benigni, Quantitaitive Structure–activity Relationshiop (QSAR) Models of Mutagenes and Carcinogens, CRC Press, Boca Raton, FL, 2003.
[36] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors Wiley-VCH, Molecular Descriptors in SAR/QSPR, Wiley, New York, 2000 (Weinheim, GER. 2000, Karelson M.).