Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 923–930
Bioorganic & Medicinal Chemistry Letters

QSAR studies on benzene sulfonamide carbonic anhydrase inhibitors: need of hydrophobic parameter for topological modeling of binding constants of sulfonamides to human CA-II

Padmakar V. Khadikar,[a,*] Vimukta Sharma,[b] Sneha Karmarkar[a] and Claudiu T. Supuran[c]

[a]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd, 3, Khatipura, Indore 452 007, India*
[b]*Department of Pharmacy, Maharaja Ranjeet Singh College, Indore 452 007, India*
[c]*Laboratorio di Chimica Bioinorganica, Dipartimento di Chimica, University of Florence, Via della Lastruccia, 3, RM, 188, Polo Scientifico, 50019 Sesto Fiorentino (Florence), Italy*

Received 4 November 2004; accepted 17 December 2004
Available online 19 January 2005

Abstract—The binding constants ($\log K$) of benzene sulfonamides to human CA-II have been modeled using a large series of distance-based topological indices. The need or otherwise of the hydrophobic parameter ($\log P$) for such topological modeling of $\log K$ has been examined critically. In both the cases excellent results have been obtained. In multiparametric models involving indicator parameters we observed that introduction of hydrophobic parameter ($\log P$) yields much improved statistics. The results are discussed on the basis of statistical parameters and also by using cross-validation method.

1. Introduction

It is well established that now a days the quantitative structure–activity relationship (QSAR) study is a major field of research in medicinal chemistry and drug design. This methodology is very helpful in screening a large library of possible drug candidates for selectivity and potency.[1–17] In this methodology mathematical models are formed that correlate molecular structure to an activity, toxicity, or property of interest. Molecular structure is invaded through the generation of descriptors, which are numerical values corresponding to topological, geometric, or electronic structural features. The goal of QSAR methodology is to develop several QSAR models to predict activities of a compound. In the last two decades, QSAR modeling based on topological (graph theoretical) indices has shown an explosive growth. According to Trinajstic and co-workers,[18] two factors greatly influenced research in this area. The first one was a rapid progress of chemical graph theory and the second was an amazing advance of computer technology, especially in terms of the portability and performance of PCs (Personal Computers).

Looking to the potential of QSAR methodology, one of us has carried out extensive work on the inhibition power of carbonic anhydrases.[19] The parameters used therein are mostly non-topological parameters. Very recently we observed that distance-topological parameters. based topological indices are best suited for this purpose.[20–28]

Some cases of the use of information theoretic indices were also reported in the literature.[29] The present study is, therefore, the extension of our earlier work related to the use of topological indices in modeling binding constant ($\log K$) of benzene sulfonamides to human CA-II (Table 1).

At this stage it is worthy to mention that a topological index is a single number, derived following a certain rule, which can be used to characterize the molecule.[30–32] A plethora of topological indices have been proposed so far.[33] The calculations of these indices are well documented in the literature.[34] In the present study we have used Wiener (W)-,[35] Szeged (Sz)-,[36–38]

Keywords: QSAR; Benzene sulfonamide; Carbonic anhydrase; Binding constant; Regression analysis; Topological index; Cross-validation.
* Corresponding author. Tel.: +91 731 2531906; fax: +91 731 2763618; e-mail: pvkhadikar@rediffmail.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.12.056

Table 1. Sulfonamide, their binding constant to human CA-II (in log units; log K) lipophilicity (log P), and indicator parameters (IP_1, IP_2, IP_3)

Compd no	R	log K	log P	IP_1	IP_2	IP_3
1	H	6.69	0.21	1	0	0
2	4-Me	7.09	0.69	1	0	0
3	4-Et	7.53	1.31	1	1	0
4	4-*n*Pr	7.77	1.64	1	0	0
5	4-*n*Bu	8.30	2.45	1	0	0
6	4-*n*Am	8.86	2.97	1	0	0
7	4-COOMe	7.98	0.64	1	0	0
8	4-COOEt	8.50	1.17	1	0	0
9	4-COO*n*Pr	8.77	1.75	1	0	0
10	4-COO*n*Bu	9.11	2.34	1	0	0
11	4-COO*n*Am	9.39	2.71	1	0	0
12	4-COO*n*He	9.39	3.23	1	0	0
13	4-CONHMe	7.08	−0.31	1	0	0
14	4-CONHEt	7.53	0.03	1	0	0
15	4-CONH*n*Pr	8.08	0.51	1	0	0
16	4-CONH*n*Bu	8.49	1.05	1	0	0
17	4-CONH*n*Am	8.75	1.54	1	0	0
18	4-CONH*n*He	8.88	2.05	1	0	0
19	4-CONH*n*Hp	8.93	2.57	1	0	0
20	3-COOMe	5.87	0.62	0	1	0
21	3-COOEt	6.21	1.11	0	1	0
22	3-COO*n*Pr	6.44	1.72	0	1	0
23	3-COO*n*Bu	6.95	2.24	0	1	0
24	3-COO*n*Am	6.86	2.71	0	1	0
25	2-COOMe	4.41	0.45	0	0	1
26	2-COOEt	4.80	0.72	0	0	1
27	2-COO*n*Pr	5.28	1.49	0	0	1
28	2-COO*n*Bu	5.76	2.01	0	0	1
29	2-COO*n*Am	6.18	2.55	0	0	1

Me—methyl, Et—ethyl, Pr—propyl, Bu—butyl, Am—pentyl, He—hexyls, Hp—heptyl; log K—logarithm of binding constant (K) to human CA-II, log P—lipophilicity (hydrophobic parameter), IP_1, IP_2, IP_3—indicator parameter for the presence (=2) or absence (=0) of substitution at 4-, 3-, and 2-position, respectively.

Padmakar-Ivan (PI)-,[39–41] Randic connectivity ($^0\chi, ^1\chi, ^2\chi$)-[42] and Kier Hall valence connectivity ($^0\chi^2, ^1\chi^2, ^2\chi^2$)-[2,43] indices. In addition, we have used three indicator parameters (IP_1, IP_2, IP_3) related to substitution at 4-, 3-, and 2-position, respectively. These indicator parameters carry only two values, 1 in presence and zero in absence of that structural feature.

Recently Hansch et al.[44] have published a very interesting review related to chem-bioinformatics and QSAR in that they have critically reviewed QSAR lacking positive hydrophobic term. Hansch is considered as the father of QSAR and used non-topological parameters for his extensive work on QSAR. The approach used by Hansch school has been to use the electronic and steric parameters developed by physical organic chemists together with hydrophobic parameters based on octanol/water partition coefficients. In the review,[44] Hansch has stated that, *it seems timely to examine those instances where hydrophobic terms are not significant*. To the belief of Hansch at this point in time, further advances in understanding biological QSAR can best be attained by comparative study. Consequently, another objective of our study is to investigate need or otherwise of hydrophobic parameter (log P) in modeling binding constant (log K) of benzene sulfonamides (Table 1) to human CA-II. Needless to state that this will be a comparative approach for which we have used maximum R^2 method followed by cross-validation.[45] The results are discussed below.

2. Results and Discussion

The set of the benzene sulfonamides used in the present investigation are recorded in Table 1, and their structures are shown in Figure 1. This Table 1 also contains binding constant to human CA-II (log K) and lipophilicity (log P) taken from the literature.[46] The assumed indicator parameters (IP_1, IP_2, IP_3) are also recorded in Table 1. Based on the log K values the following order of binding constant to human CA-II can be proposed:

$$\begin{aligned} 1 = 12 > 10 > 19 > 18 > 6 > 9 > 17 > 8 > 16 \\ > 5 > 15 > 7 > 4 > 3 = 14 > 2 > 13 > 23 > 24 \\ > 1 > 22 > 21 > 29 > 20 > 28 > 27 > 26 > 25 \end{aligned} \quad (1)$$

This Eq. 1 shows that substitution effect due to 4-COO-*n*Am and 4-COO*n*He is similar and both these constituents exhibit highest value of log K. On the other hand substitution of 2-COOMe yields lowest log K value. Also, that 4-COOEt and 4-CONHEt have similar effects in the exhibition of log K. Eq. 1, however, does not establish any structure–activity relationship. Consequent to this we have used distance-based topological indices (Table 2) for this purpose. Out of the nine topological indices, three are first-generation topological indices, while the remaining six are second-generation topological indices.[47,48] Compared to first-generation topological indices, the degeneracy observed in the second-generation topological indices are comparatively small. Balaban[47] has shown that such topological indices in spite of their observed degeneracy can be successfully used in developing QSAR models. This is so in the present case also.

Because of the dual objective of the present study further discussion will be made under two headings: (i) modeling without hydrophobic parameter that is log P and (ii) modeling with the inclusion of hydrophobic parameter that is log P. From the comparison of such results we will arrive at the need or otherwise of hydrophobic parameter (log P) for modeling binding constant (log K) of benzene sulfonamides (Table 1) to human CA-II. In both the cases we have used maximum-R^2 method[45] for obtaining statistically significant models.

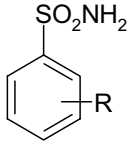

Figure 1. Structure of sulfonamides **1**–**29** (Me = methyl; Et = ethyl; Pr = propyl; Bu = butyl; Am = pentyl; He = Hexyl; Hp = heptyl).

Table 2. Distance-based topological indices for the benzene sulfonamides used in the present study (refer Table 1)

Compd no	W	Sz	PI	$^0\chi$	$^1\chi$	$^2\chi$	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$
1	114	177	84	7.6129	4.6052	4.701	5.1888	2.6838	1.8581
2	152	236	104	8.4831	4.999	5.3228	6.1115	3.0945	2.3581
3	201	306	126	9.1902	5.537	5.4919	6.8186	3.6552	2.5425
4	262	388	150	9.8973	6.037	5.8723	7.5257	4.1552	2.939
5	336	483	176	10.6044	6.537	6.2259	8.2328	4.6552	3.2925
6	424	592	204	11.3116	7.037	6.5795	8.9399	5.1552	3.6461
7	316	463	176	10.7676	6.4477	6.3633	7.428	3.661	2.561
8	393	562	204	11.4747	8.9471	6.7437	8.1351	4.2486	2.7899
9	487	676	234	12.1818	7.4477	7.0972	8.8422	4.7486	3.2053
10	596	806	266	12.8889	7.9477	7.4508	9.5493	5.2486	3.5589
11	722	953	300	13.596	8.4477	7.8043	10.2564	5.7486	3.9124
12	866	1118	336	14.3031	8.9477	8.1579	10.9635	6.2486	4.266
13	316	463	176	10.7676	6.4477	6.3633	7.5197	3.7987	2.6485
14	393	562	204	11.4747	6.9477	6.7437	8.2269	4.3593	2.9288
15	487	676	234	12.1818	7.4477	7.0972	8.934	4.8593	3.3253
16	596	806	266	12.8809	7.9477	7.4508	9.6411	5.3593	3.6788
17	722	953	300	13.596	8.4477	7.8043	10.3482	5.8593	4.0324
18	866	1118	336	14.3031	8.9477	8.1579	11.0553	6.3593	4.3859
19	1029	1302	374	15.0102	9.4477	8.5115	11.7624	6.8593	4.7395
20	300	431	176	10.7676	6.4477	6.3752	7.428	3.661	2.5644
21	374	522	204	11.4747	6.9477	6.7556	8.1351	4.2486	2.7933
22	463	628	234	12.1818	7.4477	7.1091	8.8422	4.7486	3.2088
23	568	750	266	12.8889	7.9477	7.4627	9.5493	5.2486	3.5623
24	690	889	300	13.596	8.4477	7.8163	10.2564	5.7486	3.9159
25	284	399	176	10.7676	6.4645	6.3064	7.428	3.667	2.5328
26	354	482	204	11.4747	6.9645	6.6868	8.1351	4.2545	2.7616
27	439	580	234	12.1818	7.4645	7.0403	8.8422	4.7545	3.1771
28	540	694	266	12.8889	7.9645	7.3939	9.5493	5.2545	3.5306
29	658	825	300	13.596	8.4645	7.7474	10.2564	5.7545	3.8842

2.1. Modeling of log K without log P

During the process of regression analysis using maximum-R^2 method, first we have obtained several mono-parametric models (Table 3) none of which gave any statistically significant model. The results, however, show the relative power of each of the topological indices for modeling log K. We observed the following sequence of such relative potential of the topological indices for modeling log K:

$$^2\chi^v > \mathrm{Sz} > W > {}^1\chi^2 > {}^0\chi^v > \mathrm{PI} > {}^1\chi > {}^2\chi > {}^0\chi \quad (2)$$

This Eq. 2 shows that second order branching and presence of heteroatom are favorable for the exhibition of log K.

Table 3. Regression parameters and quality of correlation for the models without hydrophobic parameter for modeling log K

Model	Parameters	Se	$r(R)$	R^2_A	F	Q
Mono-parametric regression						
1	W	1.2634	0.4734	—	7.9316	0.3747
2	Sz	1.2351	0.5106	—	8.9784	0.4134
3	PI	1.3330	0.3698	—	5.5686	0.2774
4	$^0\chi$	1.3523	0.3340	—	4.8868	0.2470
5	$^1\chi$	1.3417	0.3542	—	5.2648	0.2640
6	$^2\chi$	1.3485	0.3413	—	5.0213	0.2530
7	$^0\chi^v$	1.3020	0.4200	—	6.6294	0.3226
8	$^1\chi^v$	1.2659	0.4707	—	7.8601	0.3718
9	$^2\chi^v$	1.2336	0.5106	—	8.5263	0.4139
Bi-parametric models						
10	$^0\chi^v$, IP_1	0.5010	0.9394	—	97.666	1.8751
11	$^1\chi^v$, IP_1	0.5033	0.9389	0.8724	96.685	1.8655
Tri-parametric models						
12	W, Sz, IP_1	0.4745	0.9480	0.8866	73.942	1.9979
13	PI, $^0\chi$, IP_1	0.5020	0.9416	0.8730	65.175	1.8757
Tetra-parametric models						
14	W, Sz, $^1\chi^v$, IP_1	0.4564	0.9540	0.8951	60.701	2.0903
15	W, Sz, $^2\chi^v$ IP_1	0.4525	0.9548	0.8968	61.855	2.1100

Step-wise regression analysis yielded two bi-parametric models with excellent statistics. The one contains $^{0}\chi^{v}$ and IP_1 as the correlating parameters, while the other has $^{1}\chi^{v}$ and IP_1 as the correlating parameters. The former is slightly better than the latter and is found as below:

$$\begin{aligned}\log K &= 2.1443 + 0.4220(\pm 0.0623)\,{}^{0}\chi^{v} \\ &\quad + 2.4495(\pm 0.1959)IP_1 \\ n &= 29,\ Se = 0.5010,\ R = 0.9394,\ R^2_A = 0.8735, \\ F &= 97.666,\ Q = 1.8751\end{aligned} \tag{3}$$

Here and thereafter, n—number of compounds, Se—standard error of estimation, R—multiple correlation coefficient, R^2_A—adjustable-R^2, F—Fishers statistics, and Q—quality factor. This quality factor Q is defined[49,50] as the ratio of correlation coefficient to the standard error of estimation that is $Q = R/Se$ and is used to account for the predictive power of the model. Obviously, the larger the R, the smaller the Se, the higher will be Q, and the better will be the predictive power of the model.

In the above bi-parametric model (Eq. 3) both the correlating parameters have positive coefficients indicating their favorable contribution in the exhibition of $\log K$. The positive coefficient of $^{0}\chi^{v}$ indicates that increase in number of atoms and heteroatom increases $\log K$. Likewise, positive coefficient of indicator parameter IP_1 indicates that substitution at 4-position increases $\log K$. The results also indicate that a single parameter ($^{0}\chi^{v}$) is capable of giving excellent statistics.

The successive regression analysis yielded two tri-parametric models (Table 3), both having better statistics than the bi-parametric model discussed above. Out of the two tri-parametric models, the one containing W, Sz, IP_1 gave better results. Next to this model is the one containing PI, $^{0}\chi$ and IP_1. The former model is found as

$$\begin{aligned}\log K &= 3.5111 - 0.0215(\pm 0.0084)W \\ &\quad + 0.0200(\pm 0.0069)Sz + 1.9486(\pm 0.2289)IP_1 \\ n &= 29,\ Se = 0.4545,\ R = 9488,\ R^2_A = 0.8866, \\ F &= 73.942,\ Q = 1.9979\end{aligned} \tag{4}$$

The negative coefficient of W in the above model (Eq. 4) may be attributed to high colinearity between W and Sz. Both these indices basically account for size, shape, and branching and thus steric contribution to the exhibition of $\log K$. However, W takes account for tree-like (acyclic) side chain and Sz takes care of cyclic nature. Thus, the presence of cycle and acyclic side chain are favorable for the exhibition of $\log K$. The physical significance of IP_1 is the same as discussed for the model expressed by Eq. 3.

The step-wise regression once again yielded two tetra-parametric models (Table 3) and the one containing W, Sz, $^{2}\chi^{v}$ and IP_1 gave better results:

$$\begin{aligned}\log K &= 1.9828 - 0.0238(\pm 0.00813)W \\ &\quad + 0.0199(\pm 0.0066)Sz + 0.8407(\pm 0.4450)\,{}^{2}\chi^{v} \\ &\quad + 1.9132(\pm 0.2190)IP_1 \\ n &= 29,\ Se = 0.4525,\ R = 0.9548,\ R^2_A = 0.8968, \\ F &= 61.855,\ Q = 2.1100\end{aligned} \tag{5}$$

The improved statistics of this model (Eq. 5) is due to the addition of $^{2}\chi^{v}$ term to the model expressed by Eq. 4. This means that the number atoms, second order branching and heteroatom further increases the magnitude of $\log K$. The physical significance of all other terms involved in Eq. 5 is the same as before.

All the above models and the discussion made there in indicates that one can obtain excellent models without using hydrophobic parameter, that is, $\log P$. However, the results discussed below indicate that the quality of such models is increased by the introduction of hydrophobic term ($\log P$).

2.2. Modeling of $\log K$ using $\log P$ as one of the correlating parameter

We now discuss the cases wherein $\log P$ is also used as one of the correlating parameters. Investigating mono-parametric regression using $\log P$ will do this more successfully. This modeling gave statistically poor results (Table 4). Consequently, we have undergone bi-parametric regression analysis. All the two-variable models containing $\log P$ and one of the topological indices gave extremely poor statistics, even worse than the single-variable model containing $\log P$. In view of this we have carried out bi-parametric regression using $\log P$ and one of the indicator parameters. The results obtained are shown in Table 4, which indicates that combination of $\log P$ with IP_1 gave excellent results. Use of other two indicator parameters is not that useful. Comparison of this model with those without $\log P$ indicates that use of $\log P$ improved the statistics. The bi-parametric model containing $\log P$ and IP_1 is found as below:

$$\begin{aligned}\log K &= 4.8219 + 0.6748(\pm 0.0979)\log P \\ &\quad + 2.4335(\pm 0.1937)IP_1 \\ n &= 29,\ Se = 0.4955,\ R = 0.9408,\ R^2_A = 0.8763, \\ F &= 100.197,\ Q = 1.8987\end{aligned} \tag{6}$$

The positive coefficients of both $\log P$ and IP_1 indicate that $\log K$ increases with increase in the magnitude of $\log P$ as well as of IP_1. In other words $\log K$ is directly correlated with $\log P$ and IP_1.

The step-wise regression required the use of topological indices in addition to $\log P$ and IP_1. This yielded five tri-parametric models having better statistics than the above model (Table 4). The statistics is also better than the tri-parametric models without $\log P$ term (Table 3). Out of the five tri-parametric models, the model containing $\log P$, $^{2}\chi$, and IP_1 is found the best. This model is as below:

Table 4. Regression parameters and quality of correlation for the models with hydrophobic parameter for modeling $\log K$

Model	Parameters	Se	$r(R)$	R_A^2	F	Q
Mono-parametric regression						
16	$\log P$	1.2931	0.4332	—	14.9305	0.3350
Bi-parametric models						
17	$\log P$, IP_1	0.4955	0.9408	0.8763	100.197	1.8987
18	$\log P$, IP_2	1.2103	0.5609	0.2621	5.971	0.4634
19	$\log P$, IP_3	0.8382	0.8119	0.6460	26.547	0.9686
Tri-parametric models						
20	PI, $\log P$, IP_1	0.4211	0.9593	0.9107	96.157	2.2781
21	Sz, $\log P$, IP_1	0.4210	0.4594	0.9107	96.164	2.2789
22	$^1\chi$, $\log P$, IP_1	0.4113	0.9612	0.9148	101.183	2.3370
23	$^0\chi$, $\log P$, IP_1	0.4043	0.9625	0.9177	105.022	2.3507
24	$^2\chi$, $\log P$, IP_1	0.3952	0.9642	0.9213	110.276	2.4400
Tetra-parametric models						
25	Sz, $\log P$, IP_1, IP_2	0.2768	0.9845	0.9614	175.388	3.5567
26	$^0\chi^v$, $\log P$, IP_1, IP_2	0.2529	0.9861	0.9678	211.316	3.8200
27	PI, $\log P$, IP_1, IP_2	0.2506	0.9868	0.9684	215.305	3.9378
28	$^1\chi$, $\log P$, IP_1, IP_2	0.2288	0.9886	0.9736	259.505	4.3208
29	$^0\chi$, $\log P$, IP_1, IP_2	0.2199	0.9895	0.9756	281.325	4.4500
30	$^2\chi$, $\log P$, IP_1, IP_2	0.2180	0.9897	0.9761	286.499	4.5400

$$\log K = 2.2096 + 0.4347(\pm 0.0987)\log P + 0.4226(\pm 0.1060)^2\chi + 2.5167(\pm 0.1559)IP_1$$
$$n = 29,\ \mathrm{Se} = 0.3952,\ R = 0.9642,\ R_A^1 = 0.9213,\ F = 110.276,\ Q = 2.4400 \quad (7)$$

Thus, the improved statistics is due to the addition of $^2\chi$ term to the model expressed by Eq. 6. The positive coefficient of $^2\chi$ indicates that second order branching increases $\log K$ value.

It is interesting to mention that when we used indicator parameter IP_2 in place of IP_1, the resulting tri-parametric models were of very poor quality. On the other hand use of IP_3 in place of IP_1 though yielded statistically significant models all of them were of very inferior quality. This means that in the present case the use of indicator parameter IP_1 is most useful. This means, substitution at 4-position enhances $\log K$, while this is not the case with substitution at 2- and 3-positions.

Finally, successive regressions yielded six tetra-parametric models all having excellent statistics (Table 4). The statistics of these tetra-parametric models is also of high quality compared to the tetra-parametric models without $\log P$ term. This once again shows that though we obtained excellent models without $\log P$, the use of $\log P$ further enhances the quality of the models. Out of the six tetra-parametric models containing $\log P$ as one of the correlating parameter, the model containing $\log P$, $^2\chi$, IP_1 and IP_2 was found the best:

$$\log K = 1.6434 + 0.4033(\pm 0.0546)\log P + 0.4350(\pm 0.0585)^2\chi + 3.0451(\pm 0.1104)IP_1 + 1.0549(\pm 0.1383)IP_2$$
$$n = 29,\ \mathrm{Se} = 0.2180,\ R = 0.9897,\ R_A^2 = 0.9761,\ F = 286.500,\ Q = 4.5400 \quad (8)$$

Thus, the improved statistics is due to the addition IP_2 term to Eq. 7. That is, substitution at 3-position further increases the exhibition of $\log K$.

It is worthy to mention that when we used combinations of IP_1 and IP_3 as well as of IP_2 and IP_3 in place of the combination of IP_1 and IP_2 the resulting statistics remain the same.

3. Comments on the occurrence of highly linearly correlated topological indices in the proposed models

It is interesting to mention that all other bi-parametric models containing topological indices were of poor qualities than the models discussed above. Also, that tri- and tetra-parametric models without $\log P$ contained highly linearly correlated topological indices. Such models need further explanation. In our earlier reports[28,51,52] we have stated that the problems caused by multicollinearity, and how to deal with them continue to be of prime concern to theoretical statistician. From a decision maker's point of view point, one should be aware of the fact that multicollinearity can, and usually does, exists and recognize the basic problems it can cause. We have then mentioned four causes of multicollinearity and Randic recommendation[53] in favor of the occurrence of highly linearly correlated parameters in the proposed models. The same are applicable here. According to Randic[53] all the tri- and tetra-parametric models (Table 3) though serious for theoretical statistician are quite useful to medicinal chemistry.

4. Goodness of fit: probable error of the correlation (PE)

For supporting our results we have calculated an interesting statistical parameter called probable error of correlation (PE). This parameter, PE, is calculated by the following expression:

Table 5. Cross-validated and other supporting parameters

Model	Topological index	R^2_A	PRESS	SSY	PRESS/SSY	r^2_{cv}	S_{PRESS}	PSE	PE
10	${}^0\chi^v$, IP_1	0.8735	6.5285	49.0470	0.1331	0.8669	0.5010	0.4745	0.0596
12	W, Sz, IP_1	0.8866	5.6290	49.9465	0.1127	0.8873	0.4745	0.4406	0.0436
15	W, Sz, ${}^2\chi^v$, IP_1	0.8968	4.9142	50.6613	0.0970	0.9030	0.4525	0.4116	0.0344
17	$\log P$, IP_1	0.8763	6.3825	49.1930	0.1297	0.8703	0.4955	0.4691	0.0583
23	$\log P$, ${}^2\chi$, IP_1	0.9213	3.9047	51.6709	0.0756	0.9244	0.4043	0.3669	0.0302
30	$\log P$, ${}^1\chi$, IP_1, IP_2	0.9761	1.1400	54.4355	0.0209	0.9790	0.2180	0.1982	0.0079

$$PE = \frac{2}{3}\frac{1-r^2}{\sqrt{n}} \quad (9)$$

where r is either simple correlation coefficient or multiple correlation coefficient and n is the number of compounds used. It is recommended that,

If, $r < PE$, r is not significant.

$r > PE$, several times, at least three times greater, the correlation is indicated, and
$r > 6PE$, correlation is definitely good.

For all the proposed models we have calculated PE values, which are found significantly greater than 6PE. Thus, all the correlations attempted are definitely good.

5. Adjustable R^2 (R^2_A): a parameter in favor of step-wise regression

It is interesting to comment on the statistical parameter called adjustable-R^2 (R^2_A). This parameter tells us regarding the statistical significance of step-wise regression. This parameter takes into account of adjustment of R^2. Therefore, if a variable is added that does not contribute its fair share, the R^2_A will actually decline. It (R^2_A) is a measure of the % explained variation in the dependent variable that takes into account the relationship between the number of cases and the number of independent variable in the regression model. Whereas R^2 will always increase when an independent variable is added, R^2_A will decrease if the added variable does reduce the unexplained variation enough to offset the loss of degrees of freedom.

If R^2_A values (Table 5), in both the cases, that is, models with and without $\log P$, goes on increasing as we pass from bi- to higher parametric regressions. Hence, the added parameters (even though they are highly correlated) during the process of successive regression analysis have favorable contribution in the exhibition of $\log K$.

6. Predictive power of the models

It is necessary that the proposed model should have both the statistical quality as well as better predictive power. The aforementioned discussion and the regression parameters are good enough to establish the quality of model. However, the simplest parameter to decide the predictive power of the model is the quality factor Q, the highest value of which indicates best predictive power. In all the models discussed above Q goes on increasing as we pass from mono- to higher parametric regressions. It means that the quality of the model and its predictive power are parallel to each other's. That is, all the models discussed above are quality models having quality predictive power also. However, Todeschini[54] recently criticizes the use of quality factor Q. Consequently, we have undertaken cross-validation and obtained cross-validated parameters for the models discussed above and are recorded in Table 5.

7. Cross-validation

QSAR should be evaluated according to its ability to predict the activity of molecules, which were not used in the original QSAR table, which contains the data, the dependent activity and the independent variables. Such an evaluation can be done by cross-validation method, which is based on 'leave-n-out' concept. In each step 'n' molecules are randomly or on turn excluded from the QSAR table. The QSAR equation is then calculated and used to predict the activity of these n molecules. The methodology yields cross-validated parameters, PRESS (predictive residual sum of squares), SSY (sum of the square of the response value), r^2_{cv} (overall predictive ability), S_{PRESS} (uncertainty of predictive), and PSE (predictive square error). These parameters obtained for the model discussed above is calculated as given in Table 5.

A perusal of Table 5 shows that in each case PRESS $\ll$ SSY and also that PRESS/SSY < 0.4. This indicates that the proposed models are better than chance and indicate them to be excellent models. The PRESS/SSY value for the model-30, that is, 0.0209 indicates to the best model. The r^2_{cv} values also supports this findings. The cross-validated parameter S_{PRESS} is not useful as it is similar to the standard error of estimation (Se). The other cross-validated parameter viz., PSE is, therefore, used to estimate uncertainty of prediction, the lowest value of PSE for the model-30 establishes it to be the model with best statistics and the best predictive power.

8. Conclusions

The results and discussion made above indicate that topological modeling of binding constant ($\log K$) of benzene sulfonamides to human CA-II can be modeled excellently without the hydrophobic parameter. Under the

present situation there is no need of Hansch as well as Hammett parameters for modeling the binding constant ($\log K$). Perhaps the information contents of Hansch and Hammett parameters are equally contained in the topological indices used. However, the quality of models based on topological indices is increased significantly using the hydrophobic parameter ($\log P$).

9. Experimental

(i) *Binding constant of sulfonamides*. The binding constants ($\log K$) of benzene sulfonamides to human CA-II were taken from the literature.[46]

(ii) *Hydrophobic parameter*. The hydrophobic parameter ($\log P$) as reported by Hansch[52] was also used in the present study.

(iii) *Topological indices*. All the topological indices were calculated from the molecular graphs of benz- ene sulfonamides used in the present study. Such molecular graphs are obtained by deleting all the carbon–hydrogen as well as heteroatom–hydrogen bonds present in the molecule. Since all the topological indices used in the present study are well documented[30–34] in the literature, their calculations ate not described.

(iv) *Statistical analysis*. All the regression analyses were performed using maximum R^2 methods.[45]

(v) *Computation*. The topological indices as well as regression analysis were made using the software's provided by Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary.

Acknowledgements

Authors are thankful to Istvan Lukovits for providing the softwares for calculating topological indices as well as for making regression analysis.

References and notes

1. Simon, Z.; Chiriac, A.; Holban, S.; Ciubotariu, D.; Mihalas, G. I. *Minimum Steric Difference. The MTD Method for QSAR Studies*; Research Study Press: Letchworth, UK, 1984.
2. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Relationship*; Research Studies Press: Letchworth (UK), 1986.
3. Kaliszan, R. *Quantitative Structure–Chromatographic Retention Relationships*; Wiley & Sons: New York, 1987.
4. *Practical Applications of Quantitative Structure–Activity-Relationships (QSAR) in Environmental Chemistry and Toxicology in Chemical and Environmental Science*; Karcher, W., Devilliers, J., Eds.; Kluwer Academic Pub.: Dordrecht (The Netherlands), 1990.
5. *QSAR: Rational Approaches to the Design of Bioactive Compounds in Pharmacochemistry*; Silipo, C., Vittoria, A., Eds.; Elsevier: Amsterdam, The Netherlands, 1991.
6. Kubinyi, H. *3D QSAR in Drug Design: Theory, Methods and Applications*; ESCOM: Leiden, The Netherlands, 1993.
7. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: Weinheim, GER, 1994.
8. *QSQR in Molecular Modeling: Concepts, Computational Tools and Biological Applications*; Sanz, F., Giraldo, J., Manaut, F., Eds.; Porous Science: Barcelona, SP, 1995.
9. Hansch, C.; Leo, C. *Exploring QSAR Fundamental and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1996.
10. *Structure–Property Correlations in Drug Research*; Van de Waterbeemed, Ed.; Candes: Austin, TE, 1996.
11. Diudea, M. V.; Khadikar, P. V. *Molecular Topology and its Applications*, Gilotia Pub: New Delhi, India, in press.
12. *QSAR/QSPR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Nova Science, 2000.
13. *3D QSAR in Drug Design Ligand–Protein Interaction and Molecularity*; Kubinyi, H., Folkers, G., Martin, V. C., Eds.; Klawer/ESCOM: Amsterdam, The Netherlands, 1998.
14. *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Klawer/ESCOM: Amsterdam, The Netherlands, 1998.
15. *Comparative QSAR*; Devilliers, J., Ed.; Taylor & Francis: Washington, DC, 1998.
16. *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Pub.: Amsterdam, The Netherlands, 1999.
17. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; J. Wiley & Sons: New York, 2000.
18. Plavsic, D.; Xlikoli, S.; Trinajstic, N.; Klein, D. J. *Croatica Chem. Acta* **1993**, *66*, 345.
19. (a) Supuran, C. T.; Scozzafava, A.; Casini, A. Carbonic Anhydrase Inibitors. *Med. Res. Rev.* **2003**, *23*, 146–189; (b) Clare, B. W.; Supuran, C. T. QSAR Studies of Sulfonamide Carbonic Anhydrase Inhibitors. In *Carbonic Anhydrase, its Inhibitors and Activators*; Supuran, C. T., Scozzatava, A., Conway, J., Eds.; CRC: Boca Raton, FL, USA, 2004; pp 149–182.
20. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 171.
21. Agrawal, V. K.; Sinha, S.; Bano, S.; Khadikar, P. V. *Acta Microbiol. Immunol. Hung.* **2001**, *48*, 17.
22. Agrawal, V. K.; Shrivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287.
23. Agrawal, V. K.; Sharma, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 2993.
24. Agrawal, V. K.; Shrivastava, S.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *11*, 5353.
25. Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T.; Sudele, P. *Bioorg. Med. Chem.* **2004**, *12*, 789.
26. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2004**, *12*, 2477.
27. Jaiswal, M.; Khadikar, P. V.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3283.
28. Agrawal, V. K.; Bano, S.; Supuran, C. T.; Khadikar, P. V. *Eur. J. Med. Chem.* **2004**, *39*, 593.
29. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 447.
30. Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer: Berlin, 1986.
31. Trinajstic, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1993.
32. Diudea, M. V.; Ivanciuc, O. *Molecular Topology*; Complex: Cluj, 1995.
33. Trinajstic, N. In *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vol. II, Chapter 4.
34. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley–VCH: Weinheim, GER, 2000.
35. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
36. Gutman, I. *Graph Theory Notes, N.Y.* **1994**, *27*, 9.

37. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1975**, *35*, 547.
38. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *Commun. Math. Comput. Chem. (MATCH)* **2001**, *43*, 7.
39. Khadikar, P. V. *Natl. Acad. Sci. Lett.* **2000**, *23*, 113.
40. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
41. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 134.
42. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
43. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
44. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619.
45. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
46. King, R. W.; Burgen, A. S. V. *Proc. R. Soc. London B* **1976**, *193*, 107.
47. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
48. Balaban, A. T. *Math. Chem.* **1986**, *21*, 115.
49. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
50. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
51. Thakur, M.; Agrawal, A.; Thakur, A.; Khadikar, P. V. *Bioorg. Med. Chem.* **2004**, *12*, 2287.
52. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5661.
53. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
54. Todeschini, R., Some Observations about Pogliani *Q*, Quality Index; In *Chimometrics Web News*, Miland Chemo metrics and QSAR Research Group (File: 111C/Windows/Desktop/WebNewsonChemometrics.html), 2001.