Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 14 (2004) 3283–3290
Bioorganic & Medicinal Chemistry Letters

Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides

Mona Jaiswal,[a] Padmakar V. Khadikar,[a,*] Andrea Scozzafava and Claudiu T. Supuran[b]

[a]*Research Division, Laxmi Fumigation and Pest Control (P), 3, Khatipura, Indore 452007, India*
[b]*Universita degli Studi di Firenze, Laboratorio di Chiica Bioinorganica, Rm. 118, Via della Lastruccia 3, I-500 la Sesto Fiorentino (Firenze), Italy*

Received 6 February 2004; accepted 27 March 2004

Abstract—QSAR study on the tumor-associated transmembrane carbonic anhydrase IX (CA IX) isoenzyme has been made using a large pool of distance-based topological indices : W, Sz, PI, ${}^{0}\chi$, ${}^{1}\chi$, ${}^{2}\chi$, ${}^{0}\chi^{v}$, ${}^{1}\chi^{v}$, and ${}^{2}\chi^{v}$. A combined set of 32 aromatic and heterocyclic compounds, including the six clinically used derivatives: acetazolamide, methazolamide, ethoxyzolamide, dichlorophenamide, dorzolamide, and brinzolamide are used for this purpose. The results have shown that the inhibition of the tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides can be modeled excellently in multiparametric regression after introduction of indicator parameters. The predictive power of the models is discussed using probable error of correlation (PE), variance-inflation factor (VIF), and cross-validation parameters: PRESS, SSY, r^{2}_{cv}, S_{PRESS}, and PSE. This is the first report on QSAR study on inhibition of tumor-associated isoenzyme IX.

1. Introduction

Several QSAR studies of carbonic anhydrase inhibitors using physiochemical parameters as well as theoretical descriptors have been reported in the last years.[1–18] Encouraging results were obtained using distance-based topological indices. The favorable topological indices used being: Wiener (W)-,[19] Szeged (Sz)-,[20,21] PI (Padmakar–Ivan)-,[22–25] Randic connectivity (${}^{0}\chi$, ${}^{1}\chi$, ${}^{2}\chi$)-,[26] and Kier and Hall's valence connectivity (${}^{0}\chi^{v}$, ${}^{1}\chi^{v}$, ${}^{2}\chi^{v}$)-[27,28] indices. In addition, information theoretic indices, chiefly molecular redundancy (MRI)-, and negentropy (N) were also found useful. In some cases[29] the applicability of W and ${}^{m}\chi$ and/or ${}^{m}\chi^{v}$ was well established, however, that of Sz and PI is yet to be established thoroughly.

Quantitative structure–property/activity/toxicity relationships (QSPR/QSAR/QSTR) studies are tools of prediction endpoints of interest on organic compounds acting as drugs, which have not been experimentally determined.[30–34] Many physiological activities of compounds can be related to their composition and structure. Since topological indices are the numerical representation of molecular structure, they are the best candidates for QSPR/QSAR/QSTR studies.[30–34]

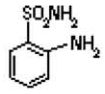
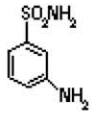
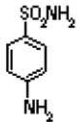
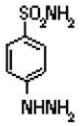
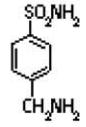
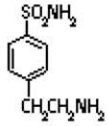
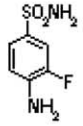
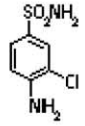
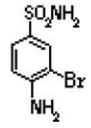
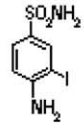
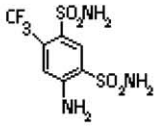
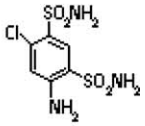
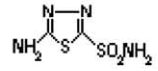
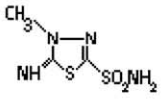
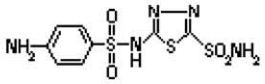
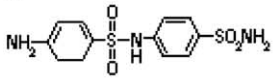
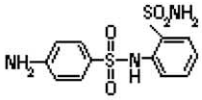
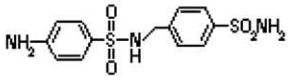
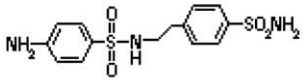
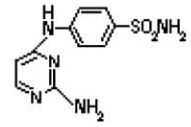
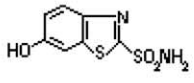
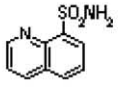
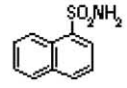
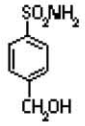
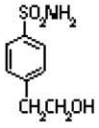
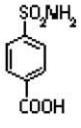
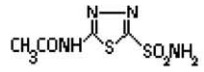
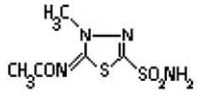
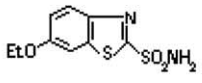
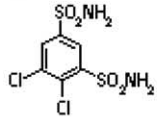
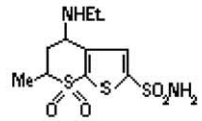
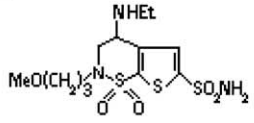
The carbonic anhydrase (CA, EC 4.2.1.1) inhibitors are extensively studied in the last period, due to their potential applications for the prevention and treatment of a large number of diseases.[35–38]

Recently, Supuran and co-workers[39] have investigated the inhibition of the tumor-associated transmembrane carbonic anhydrase IX (CAIX) isoenzyme with a series of aromatic and heterocyclic sulfonamides, including the six clinically used derivatives: acetozalamide, methazolamide, ethoxyzolamide, dichlorophenamide, dorzolamide, and brinzolamide (Table 1). A very interesting and unusual inhibition profile against CA IX with these sulfonamides has been observed. Several nanomolar

Keywords: QSAR; Carbonic anhydrase inhibitors; Aromatic sulfonamides; Heterocyclic sulfonamides; Regression analysis; Topological modeling; Distance-based indices.
* Corresponding author. Tel.: +91-731-2531906; fax: +91-731-27011-70; e-mail addresses: pvkhadikar@rediffmail.com, tomona@rediffmail.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.03.099

Table 1. Tumor-associated isoenzymes IX with aromatic and heterocyclic sulfonamides used in the present involving

 **1**	 **2**	 **3**	 **4**
 **5**	 **6**	 **7**	 **8**
 **9**	 **10**	 **11**	 **12**
 **13**	 **14**	 **15**	
 **16**		 **17**	
 **18**		 **19**	
 **20**	 **21**	 **22**	
 **23**	 **24**	 **25**	 **26**
 **27**	 **28**	 **29**	
 **30**	 **31**	 **32**	

(K_{I-s} in the range of 14–50 nM) CA IX inhibitors have been detected, both among the aromatic as well as the heterocyclic sulfonamides examined (Table 1). The referred report[39] is the first CA IX inhibition study with a series of aromatic and heterocyclic sulfonamides, as well as with the six clinically used CAIs mentioned above.

Since CA IX is a highly active isoenzyme predominantly expressed in tumor tissues with poor prognosis of disease progression and, QSAR study on such an interesting class of isoenzyme is yet to be performed, we have undertaken the present study, which will be very promising for the potential design of CA IX specific inhibitors with application as anti-tumor agents.

2. Results and discussion

The structure of aromatic and heterocyclic sulfonamides together with the six clinically used inhibitors are shown in Table 1. Inhibition data against CA IX isoenzymes[39] and the values of indicator parameters (I_1, I_2, I_3) are presented in Table 2. The indicator parameters: I_1, I_2, I_3 assumes the value 1 when heterocyclic, halogens, and $-NH_2$ group *para* to $-SO_2NH_2$ are, respectively, present in the compounds used. In absence of which the values of indicator parameters are zero.

Table 3 records the calculated values of the distance-based topological indices (W, Sz, PI, $^0\chi$, $^1\chi$, $^2\chi$, $^0\chi^v$, $^1\chi^v$, and $^2\chi^v$) used.

A close look of Table 2 shows that a very little degeneracy is present in the inhibitory activity: $\log K_I$ (hCAIX) activity of the compounds used. However, high degeneracy is observed in the topological indices, W, Sz, and PI are first-generation topological indices while the remaining topological indices ($^0\chi$, $^1\chi$, $^2\chi$, $^0\chi^v$, $^1\chi^v$, and $^2\chi^v$) are the second-generation topological indices.[40] Balaban[40] has shown that such indices, inspite of their degeneracy, can be successfully in QSAR studies. This is found to be the case in present study also.

The simple linear regression analysis has indicated that none of the topological indices used gave any statistically significant regression expression for modeling the activity. This means that no single variable model is possible for modeling the activity [$\log K_I$(hCA IX), nM]. Many correlations, particularly when involving molecules of different size, need not be linear.

In view of the above failure we have undertaken stepwise multiple regression using maximum R^2 method.[41] Before a multiple (multivariate) analysis is undertaken, it is convenient to tailor the data in certain ways to make the calculation easier. Normally, it is sufficient to propose the data by means of auto-scaling and mean-centering the variables. Auto-scaling gives each variable unit variance and the same chance to contribute to a calculated model; whereas mean-centering facilitates interpretation. Multicollinearity (auto-correlation) occurs when two independent variables are correlated with each other and therefore contribute redundant information to the model when highly correlated independent variables are included in the regression model, they can effect the regression results. However, the multicollinearity problem is recently discussed by Randic[42] and we will use his recommendations in discussing our results.

At this stage it is worthy to mention that the problems caused by multicollinearity; and how to deal with them, continue to be of prime concern to theoretical statistician. From a decision makers viewpoint, one should be aware of the fact that multicollinearity can (and usually does) exist and recognize the basic problems it can cause. Some of the most obvious problems and indications of severe multicollinearity are:

(i) incorrect signs of the co-efficients;
(ii) a change in the values of the previous co-efficient when a new variable is added to the model;
(iii) change in insignificant of a previously significant variable when a new variable is added to the model; and
(iv) an increase in the standard errors of the estimate when a new variable is added to the model.

The simplest method of investigating occurrence of multicollinearity is to obtain correlation matrix.[41] In addition, correlation matrix is very useful for determining which independent variables are likely to help explain variation in the dependant variable. In our case such a correlation matrix obtained (not shown here) indicated that all the distance-based topological indices used in the present investigation are highly linearly correlated. This means that any combination of these indices occurring in the regression expression (model) it may suffer from the defect due to collinearity. In practice every term in the correlation matrix >0.4 may be taken as being suspicious due to collinearity. In the present case this term was found as high as 0.99 for collinearity between W and Sz, $^0\chi$ and $^1\chi$, and $^0\chi^v$ and $^2\chi^v$. The high linear collinearity between these pairs is well known.

Another test for multicollinearity is the variance inflation factor (VIF).[41,43] The VIF is defined as $1/(1-R_i^2)$, where R_j is the multiple correlation co-efficient for the ith variable regressed on the $p-1$ others, p being the number of others. A VIF less than 10 indicates that the model contains no multicollinearity. However, requiring VIFs less than 10 is not so traditional, but it is recommended by regression texts. Based on this requirement only the regression equation 1, as discussed below, will be entirely satisfactory. However, the best regression Eq. 5, which is discussed below, with 32 compounds and the independent variables of Eq. 5 (see below), the VIFs come out to be as under:

Variable	VIF
I_2	1.3041
I_3	1.1646
PI	115.1883
$^1\chi$	153.6831
$^2\chi$	23.6657
$^2\chi^v$	22.0293

Thus, looking to the aforementioned requirement of VIFs value, four variables viz PI, $^1\chi$, $^2\chi$, and $^2\chi^v$ do not reach statistical significance, leaving us with a dependence on the two indicator variables that is, I_2 and I_3. That is, VIFs of Eq. 5 indicates a massive collinearity problem. However, for resolving such a problem we have to take help of Randic[42] recommendations.

Before resolving collinearity issue using Randic recommendation[42] it is interesting to consider the results of Shapiro and Guggenheim[43] who reported VIFs for the inhibition of oral bacteria by phenolic compounds using molecular connectivity. In two of the proposed models they obtained VIFs value as high as 313/362. Also, for

Table 2. $K_{I(hCA\ IX)}$ activity values of indicator parameters (I_1, I_2, I_3) and topological indices (W, Sz, PI, $^0\chi$, $^1\chi$, $^2\chi$, $^0\chi^v$, $^1\chi^v$, and $^2\chi^v$)

Compound no	$K_{I(hCA\ IX)}$	I_1	I_2	I_3	W	Sz	PI	$^0\chi$	$^1\chi$	$^2\chi$	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$
1	33	0	0	0	144	220	104	8.4831	5.0159	5.2343	5.6888	2.8892	2.0828
2	238	0	0	0	148	228	104	8.4831	4.9990	5.3347	5.6888	2.8832	2.1175
3	294	0	0	1	152	236	104	8.4831	4.9990	5.3228	5.6888	2.8832	2.1140
4	305	0	0	1	201	306	126	9.1902	5.5370	5.4919	6.1889	3.1332	2.2137
5	103	0	0	1	201	306	126	9.1902	5.5370	5.4919	6.3960	3.3563	2.3931
6	33	0	0	1	262	388	150	9.8973	6.0370	5.8723	7.1031	3.8563	2.7276
7	245	0	1	1	189	291	126	9.3534	5.4097	5.8306	5.9895	2.9989	2.2283
8	264	0	1	1	189	291	126	9.3534	5.4097	5.8306	5.9895	2.9989	2.2283
9	269	0	1	1	189	291	126	9.3534	5.4097	5.8306	5.9895	2.9989	2.2283
10	285	0	1	1	189	291	126	9.3534	5.4097	5.8306	5.9895	2.9989	2.2283
11	24	0	0	1	624	888	336	15.2236	8.2430	10.3662	8.9701	4.3066	3.4740
12	39	0	1	0	399	582	234	12.7236	7.0317	8.3710	7.7142	3.6787	2.9105
13	41	1	0	0	113	132	85	7.7760	4.4990	4.9812	4.6821	2.1173	1.4512
14	30	1	0	1	146	171	104	8.6463	4.9097	5.4889	5.6048	2.5258	1.8417
15	38	1	0	1	853	1124	398	14.9663	9.1825	9.8915	9.7163	4.9302	3.5927
16	34	0	0	1	1004	1502	450	15.6734	9.6825	10.2332	10.723	5.6962	4.2439
17	20	0	0	1	916	1326	450	15.6734	9.6993	10.1666	10.723	5.7021	4.2157
18	31	0	0	1	1195	1726	494	16.3805	10.1825	10.5677	11.4301	6.1533	4.5825
19	24	0	0	0	1408	1972	540	17.0876	10.6825	10.9213	12.1372	6.6533	4.9058
20	16	1	0	0	669	1057	330	13.1734	8.4485	8.4572	9.2380	4.9342	3.4945
21	14	1	0	0	287	430	191	10.3449	6.4653	6.9727	6.8368	3.5903	2.6768
22	32	1	0	0	272	510	196	10.1818	6.5883	6.6118	7.2134	3.9543	2.8763
23	30	0	0	0	272	510	196	10.1818	6.5883	6.6118	7.3435	4.0945	3.0223
24	21	0	0	0	201	306	126	9.19020	5.5370	5.4919	6.2658	3.2643	2.3471
25	22	0	0	0	262	388	150	9.8973	6.0370	5.8723	6.9729	3.7643	2.6626
26	26	0	0	0	252	378	150	10.0605	5.9097	6.2217	6.4670	3.2723	2.3745
27	25	1	0	0	257	294	148	10.0605	5.8929	6.3716	6.5130	3.0327	2.0993
28	27	1	0	0	304	347	173	10.9307	6.3035	6.9013	7.4602	3.4272	2.4500
29	34	1	0	0	442	645	251	11.7591	7.5033	7.5223	8.5050	4.5666	3.0873
30	50	0	1	0	398	580	234	12.7236	7.0317	8.3433	7.5148	3.5790	2.8072
31	52	1	0	0	634	918	356	14.4223	8.6749	9.3116	10.5814	5.8138	4.5548
32	37	1	0	0	1142	1578	524	17.2507	10.7129	10.6032	12.9808	7.1044	5.1729

Table 3. Results of stepwise regression for modeling $\log K_{I(hCA\ IX)}$

Eqn	Model	Parameter used	Se	R	F	Q	R^2_A
1	1	$^2\chi$, I_3	0.3085	0.7150	15.163	2.3177	0.4775
2	2	$^1\chi$, $^2\chi^v$, I_3	0.2955	0.7529	12.213	2.5478	0.5204
3	3	$^1\chi$, $^2\chi^v$, I_2, I_3	0.2796	0.7933	11.295	2.8372	0.5705
4	4	$^1\chi$, $^2\chi^v$, I_1, I_2, I_3	0.2770	0.8040	9.508	2.9025	0.5785
5	7	$^1\chi$, $^2\chi^v$, $^2\chi$, PI, I_2, I_3	0.2748	0.8158	8.290	2.9687	0.5852

Se—standard error of estimation, R—multiple correlation co-efficient, F—F-statistics, Q—quality factor, and R^2_A—adjustable R^2.

one of the best model, quadratic in $^1\chi^v$, resulted with VIF value of 25.4 for both the terms that is, $^1\chi^v$ and $(^1\chi^v)$.[2] Looking to these results Eqs. 2 and 3 yielding VIFs values around 19 and 21, respectively, for the parameter involved there in may be more acceptable than Eq. 5. Furthermore, we can argue that the statistical requirement of VIF less than 10 is tentative. In this connection we have to seriously consider Randic recommendations.[42] Randic stated that selection of the descriptors to be used in structure–property–activity studies should not be delegated solely to the computers although the statistical criteria will continue to be useful for preliminary screening of the descriptors taken from a large pool. Often in an automated selection of descriptors a descriptor will be discarded because it is highly correlated with another descriptor already selected. But what is important is not descriptor parallel one another, that is, duplicate much of the same structural information but whether they differ in those parts that are important for structural–property–activity correlations. If they differ in the domain, which is important for the property/activity considered both descriptors should be retained. If they differ in parts that are not relevant for the correlation of considered property/activity then one of them can be discarded. Consequently, in the following proposed Eqs. (1)–(5), therefore, all the linearly highly correlated descriptors are to be retained as they have different information content. Also, in all the five equations the co-efficients of the parameters involved are much higher than their respective standard deviations. In addition, none of the four problems due to collinearity, as mentioned earlier, are present. Such equations are statistically allowed. Further support in favor of our results is obtained from the consistent increase in R^2_A value when we pass from Eqs. 1–5. This is elaborately discussed in the following section:

The stepwise regression resulted into statistically significant biparametric regression equation as below:

$$\begin{aligned}\log K_I(\text{hCA IX}) &= 2.3511 - 0.1223(\pm 0.0287)^2\chi \\ &\quad + 0.4905(\pm 0.1122)I_3 \\ n = 32,\quad \text{Se} = 0.3085,\quad R &= 0.7150,\quad R^2_A = 0.47746, \\ F = 15.1629,\quad Q &= 2.3176\end{aligned} \tag{1}$$

where n is the number of the compounds, R is the correlation co-efficient, Se is the standard error of estimation, F is the Fischer F-ratio and Q is the quality factor. This quality factor Q is defined in the literature[44,45] as the ratio of correlation co-efficient (R) to the standard error of estimation (Se) that is, $Q = R/\text{Se}$ and is used to describe the predictive power of the model. The values in the paranthesis are the respective standard deviation of the co-efficients. All other biparametric regression expressions were of poor statistical quality than the one given above.

Subsequent step-wise regression resulted into slightly better triparametric regression containing $^1\chi$, $^2\chi^v$, and I_3 as the correlating parameters. All other triparametric correlations were of low quality. The regression equation under consideration is found as below:

$$\begin{aligned}\log K_I(\text{hCA IX}) &= 2.6141 - 0.3804(\pm 0.1258)^1\chi \\ &\quad + 0.5025(\pm 0.2425)^2\chi^v \\ &\quad + 0.4879(\pm 0.1070)\text{I}_3 \\ n = 32,\quad \text{Se} = 0.2955,\quad R &= 0.7529,\quad \text{R}^2_A = 0.5204, \\ \text{F} = 12.2126,\quad \text{Q} &= 2.5478\end{aligned} \tag{2}$$

Both the above expressions show that first and second order branching, presence of heteroatom, and the presence of $-NH_2$ group *para* to $-SO_2NH_2$ group have significant influence on the inhibition of the tumor associated isozyme IX with aromatic and heterocyclic sulfonamides.

Successive regression analysis gave a tetraparametric regression expression containing $^1\chi$, $^2\chi^v$, I_2, I_3 with still better statistics than both the models discussed above. Few more tetraparametric models having better quality than the model expressed by Eqs. 1 and 2 were also obtained but the one mentioned here was the best:

$$\begin{aligned}\log K_I(\text{hCA IX}) &= 2.4664 - 0.3643(\pm 0.1193)^1\chi \\ &\quad + 0.5092(\pm 0.2291)^2\chi^v \\ &\quad + 0.2962(\pm 0.1434)I_2 \\ &\quad + 0.3930(\pm 0.1112)I_3 \\ n = 32,\ \text{Se} = 0.2796,\ R &= 0.7933,\ R^2_A = 0.5705, \\ F = 11.2945,\ Q &= 2.8372\end{aligned} \tag{3}$$

This means that the quality of the model expressed by Eq. 2 is improved by the introduction of I_2 indicator parameter. That is, the introduction of halogen further

enhances inhibition of the tumor associated isozyme IX. The improvement in the R^2_A value indicates that introduction of I_2 is well justified. This has promoted us to have penta-parametric regression expression:

$$\begin{aligned}\log K_I(\text{hCA IX}) &= 3.7804 - 0.7091(\pm 0.3046)^1\chi \\ &\quad + 0.5123(\pm 0.2270)^2\chi^v + 0.00462 \\ &\quad \times (\pm 0.0038)I_1 + 0.3119(\pm 0.1427)I_2 \\ &\quad + 0.3520(\pm 0.1151)\mathrm{I}_3 \\ n = 32, \quad \mathrm{Se} = 0.2770, \quad R = 0.8040, \quad R^2_A &= 0.5785, \\ F = 9.5077, \quad Q &= 2.9025 \end{aligned} \tag{4}$$

The regression parameters associated with this Eq. 4 are in favor of the additional presence of I_1 indicator.

All the models (Eqs. 1–4) discuss above indicate that inhibition of the tumor associated isozyme IX is favored by the presence of heterocyclic ring, halogens, and $–NH_2$ group *para* to $–SO_2NH_2$ group. The third parameter has the dominating role in this respect. Also, that extent of branching and the presence of heteroatom are another favorable parameters for this purpose. Still at this stage we could not obtain any improved model which may account for the size as a whole. The appropriate distance-based topological indices for this purpose are Wiener (W)-, Szeged (Sz)-, and PI (Padmakar–Ivan)-indices. When these topological indices are used we observed that PI index gave better results than both W and Sz indices. Three such regressions with improved statistics than the models discussed were obtained (Models 5–7, Table 3), out of these the one containing $^1\chi$, $^2\chi$, $^2\chi^v$, PI, I_2, and I_3 gave better results:

$$\begin{aligned}\log K_I(\text{hCA IX}) &= 3.9171 - 0.5583(\pm 0.3274)^1\chi \\ &\quad - 0.1447(\pm 0.1212)^2\chi \\ &\quad + 0.4171(0.2403)^2\chi^v \\ &\quad + 0.0053(\pm 0.0038)\mathrm{PI} \\ &\quad + 0.3771(\pm 0.1517)I_2 \\ &\quad + 0.3480(\pm 0.1142)I_3 \\ n = 32, \quad \mathrm{Se} = 0.2748, \quad R = 0.8158, \quad R^2_A &= 0.5852, \\ F = 8.2903, \quad Q &= 2.9687 \end{aligned} \tag{5}$$

This Eq. 5 with positive co-efficient for PI term finally indicates that the size of the molecule is to some extent responsible for the inhibition of the tumor-associated isozyme. Further stepwise regressions indicated that no other higher parameteric regression gave better statistics than the model expressed by Eq. 5. The subsequent discussion is, therefore, centered at this model expressed by Eq. 5.

Now, it is interesting to mention that in the regression equation (2)–(5) the co-efficient of $^1\chi$ term is negative. This first-order Randic connectivity index conveys more information about the number of atom in a molecule, the negative co-efficient, which indicates such a situation is not favorable for the exhibition of the activity under present study. Also, the co-efficient of $^2\chi$ term in Eqs. 1 and 5 is negative. Note that this second order Randic connectivity index indicates more information of branching. In all the regression equations discussed above (except regression equation 1), the co-efficient of $^1\chi^v$ term is positive indicating thereby the degree of unsaturation and the presence of heteroatom are favorable for the exhibition of the activity under study.

The use of PI index and its positive co-efficient in Eq. 5 needs further explanation. As stated earlier[46–48] PI index is a Weiner-type index introduced to resolve nonapplicability of Sz index for acyclic molecules. Though Weiner index (W) is a well established index, we observed that inspite of this in many cases better results are obtained using Sz and PI indices in place of Weiner index. In the present study also this is found to be the case. It indicates that PI index has hitherto unknown information not present in W and Sz indices.

The models suggested in the present study invariably are based on the use of connectivity indices, both the molecular connectivity indices of Randic[26] and their modifications due to Kier and Hall.[27,28] Thus, the key objective becomes the molecular connectivity approach in favor of our earlier studies.[49,50]

We have used the regression parameters associated with Eq. 5 and calculated probable error of the correlation (PE). The following recommendations are used for this purpose:

(i) If $R <$ PE, R is not significant,
(ii) If $R >$ PE, several times, at least three times greater, then correlation is indicated, and;
(iii) If $R > 6$PE, the correlation is good.

In our case (Eq. 5), the PE value was found as 0.04 indicating that R is much greater than 6PE. This favors that the correlation (Eq. 5) is very good.

At this stage it is interesting to comment on R^2_A values, which accounts for the adjustment of R^2. If a variable is added that does not contribute its fair share, the R^2_A value actually decline. In passing from the Eqs. 1–5 we observed a consistent increase in R^2_A values. This means that the added descriptors in each case have their fare share in the proposed model. 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 number of independent variables 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 redeem the unexplained variation enough to offset the loss of degrees of freedom.

The predictive ability of the proposed models can be discussed on the basis of quality factor Q.[44,45] As mentioned above, the magnitude of Q goes on increasing as we pass from Eqs. 1–5 and is maximum for Eq. 5. That is, the model based on Eq. 5 has the highest predictive ability. This is in consistent with the definition $Q = R/\mathrm{Se}$

Table 4. Cross-validation parameters for most appropriate model for modeling $\log K_{I(hCA\ IX)}$

Model	Parameters used	PRESS	SSY	PRESS/SSY	r^2_{cv}	S_{PRESS}	PSE
1 [Eq. 1]	$^2\chi$, I_3	2.7540	2.8852	0.9563	0.0437	0.3084	0.2936
2 [Eq. 2]	$^1\chi$, $^2\chi^v$, I_3	2.4498	3.1992	0.7656	0.2343	0.2958	0.2767
3 [Eq. 3]	$^1\chi$, $^2\chi^v$, I_2, I_3	2.1113	3.5329	0.5976	0.4024	0.2796	0.2569
4 [Eq. 4]	$^1\chi$, $^2\chi^v$, I_1, I_2, I_3	1.9956	3.6487	0.5469	0.4531	0.2770	0.2497
5 [Eq. 5]	$^1\chi$, $^2\chi$, $^2\chi^v$, PI, I_2, I_3	1.8879	3.7563	0.5026	0.4974	0.2748	0.2429

PRESS—predicted residual sum of squares, SSY—sum of the squares of response values, r^2_{cv}—overall predictive ability (cross-validated correlation coefficient), S_{PRESS}—uncertainty of prediction and PSE-predictive square error.

given earlier. The higher the value of R (correlation coefficient), the lower the Se (standard error of estimation), the larger will be Q and the better will be the predictive ability. Further support in favor of our results is obtained from the cross-validation methodology.

To avoid the possibility of a chance correlation each model was validated using cross-validation method. As opposed to the traditional regression methods, cross-validation evaluates the validity of a model by how well it predicts data rather than how well it fits the data. The analysis uses a 'leave-one-out' scheme; a model is built with $N-1$ compounds and the Nth compound is predicted. Each compound is left out of the model derivative and predicted in turn. The calculated values of cross-validated parameters are presented in Table 4. The meaning of these parameters is given as a footnote to Table 4. The values are of the cross-validated parameters in favor of our results discussed above.

3. Conclusions

The results and discussion made above indicate that the distance-based topological indices can be used successfully for modeling, monitoring, and estimation inhibition of tumor-associated isozyme with aromatic and heterocyclic sulfonamides such as those used in the present study. The results also show that distance-based topological indices, which accounts for branching and heteroatom are more useful for this purpose. Also, that effect due to size can be more successfully dealt with using PI (Padmakar–Ivan) index. Finally, presence of $-NH_2$ group *para* to $-SO_2NH_2$ is the most favorable situation.

4. Experimental

4.1. Set of sulfonamides and CA IX activity

The set of sulfonamides and their hCA IX activities used are those reported earlier[39] by one of the author (CTS) on the present paper. These activity values are converted into their log units and then used for modeling.

4.2. Topological methodology

The topological indices: Wiener (W)-,[19] Szeged (Sz)-,[20,21] PI (Padmakar–Ivan)-,[22–25] Randic connectivity ($^0\chi$, $^1\chi$, $^2\chi$)-,[26] and Kier and Hall's valence connectivity ($^0\chi^v$, $^1\chi^v$, $^2\chi^v$)-[27,28] indices were calculated using all-hydrogen suppressed graph. Such molecular graphs are obtained by deleting all the carbon–hydrogen as well as heteroatom-hydrogen bonds present in the sulfonamides moiety. The calculations of these indices are reported in the literature (see references given above) and thus the same are not described in this publication.

4.3. Indicator parameter

These are dummy parameters[41] sometime used in QSAR analysis for accounting those structural feature not accounted for in the molecular/topological descriptors used. They assume only two values: 0 or 1. The presence of the structural feature under question is indicated by assuming the value of 1 for that indicators, in absence of which the referred indicator parameter is taken as zero. The indicator parameters: I_1, I_2, and I_3 accounting for presence or otherwise of heterocyclic ring, halogens, and of $-NH_2$ group *para* to $-SO_2NH_2$, respectively, are used in the present study.

4.4. Computation

The topological indices used are calculated using the program developed by Raj Singh Sisodia, while the regression analysis is carried out using Regress-1 program of Prof. Istvan Lukovits.

Acknowledgements

The authors are thankful to Professor B.W. Clare as well as the referees for suggesting modifications on the paper and providing calculations of VIFs, which has ultimately resulted into a better quality paper. The authors are also thankful to Professor Lukovits and Mr. Raj Singh Sisodia for providing their softwares. One of the authors (P.V.K.) is thankful to Professor Ivan Gutman for introducing him (P.V.K.) to the fascinating field: Chemical Topology and Graph Theory.

References and notes

1. Agrawal, V. K.; Shrivastava, S.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *11*, 5353.
2. Thakur, A.; Thakur, M.; Supuran, C. T.; Khadikar, P. V.; Sudele, P. *Bioorg. Med. Chem.* **2004**, *12*, 789.

3. Casini, A.; Antel, J.; Abbate, F.; Scozzafava, A.; David, S.; Waldeck, H.; Schafer, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 841.
4. Clare, B. W.; Supuran, C. T. *J. Pharm. Sci.* **1994**, *83*, 768.
5. Clare, B. W.; Scozzafava, A.; Supuran,C. T. *J. Med. Chem.*, submitted.
6. Scozzafava, A.; Supuran, C. T. *SAR. QSAR Environ. Res.* **2001**, *12*, 17.
7. Supuran, C. T.; Balaban, A. T. *Rev: Roum. Chem.* **1994**, *39*, 107.
8. Maren, T. H.; Clare, B. W.; Supuran, C. T. *Roum. Chem. Quart. Rev.* **1994**, *2*, 259.
9. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 447.
10. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1998**, *33*, 489.
11. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1995**, *30*, 687.
12. Clare, B. W.; Supuran, C. T. *Eur. J. Med. Chem.* **1997**, *32*, 311.
13. Supuran, C. T.; Clare, B. W. *Eur. J. Med. Chem.* **1999**, *34*, 41.
14. Clare, B. W.; Supuran, C. T. *Eur. J. Med. Chem.* **1999**, *34*, 4643.
15. Agrawal, V. K.; Sharma, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 2993.
16. Agrawal, V. K.; Shrivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287.
17. Agrawal, V. K.; Sinha, S.; Bano, S.; Khadikar, P. V. *Acta Microbiol. Immunol. Hung.* **2001**, *48*, 17.
18. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 171.
19. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
20. Gutman, I. Graph Theory Notes New York, **1994**, *27*, 9.
21. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
22. Khadikar, P. V. *Nat. Acad. Sci. Lett.* **2000**, *23*, 113.
23. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
24. Khadikar, P. V.; Karmarkar, S.; Varma, R. G. *Acta Chem. Slov.* **2002**, *49*, 755.
25. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 134.
26. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
27. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity–Relationship*; Wiley: NewYork, 1986.
28. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
29. Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmarkar, S.; Khadikar, P. V. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57.
30. Diudea, M. V.; Khadikar, P. V. *Molecular Topology and its Applications*, New Age Int: New Delhi, in press.
31. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
32. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; J. Wiley and Sons: New York, 2000.
33. Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer: Berlin, 1986.
34. Randic, M. *J. Mol. Graphis Modelling* **2001**, *20*, 19.
35. (a) Maren, T. H.; Wynns, G. C.; Wistrand, P. J. *Mol. Pharmacol.* **1993**, *44*, 901; (b) Maren, T. H.; Wynns, G. C.; Wistrand, P. J. *Mol. Pharmacol.* **1993**, *44*, 901.
36. Supuran, C. T.; Scozzafava, A.; Cassini, A. *Med. Res. Rev.* **2003**, *23*, 146.
37. Supuran, C. T.; Scozzafava, A. *Exp. Opin. Ther. Pat.* **2002**, *12*, 217.
38. (a) Supuran, C. T. *Roum. Chem. Quart. Rev.* **1993**, *1*, 77; (b) Scozzafava, A.; Owa, T.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Med. Chem.* **2003**, *10*, 925; (c) Casini, A.; Antel, J.; Abbate, F.; Scozzafava, A.; David, S.; Waldeck, H.; Schafer, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 841; (d) Casini, A.; Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Cancer Drug Targets* **2002**, *2*, 55; (e) Supuran, C. T. *Exp. Opin. Invest. Drugs* **2003**, *12*, 283.
39. Vullo, D.; Franchi, M.; Galloi, E.; Pastorek, J.; Scozzafava, A.; Pastorekova, S.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *13*, 1005.
40. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
41. Chatterjee, S.; Hadi, A.; Price, B. *Regression Analysis by Examples*. 3rd ed.; Wiley-VCH: New York, 2000.
42. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
43. Shapiro, S.; Guggenheim, B. *Quant. Struct-Act. Relat.* **1998**, *17*, 327.
44. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
45. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
46. Khadikar, P. V.; Mandloi, D.; Bajaj, A. V.; Joshi, S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 419.
47. Khadikar, P. V.; Singh, S.; Mandloi, D.; Joshi, S.; Bajaj, A. V. *Bioorg. Med. Chem.* **2003**, *11*, 5045.
48. Khadikar, P. V.; Joshi, S.; Bajaj, A. V.; Mandloi, D. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1187.
49. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3517.
50. Agrawal, V. K.; Sohgaura, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3295.