QSAR study on murine recombinant isozyme mCAXIII: topological vs structural descriptors

Jyoti Singh,[a] Shalini Singh,[b] S. Thakur,[c] Meenakshi Lakhwani,[d] Padmakar V. Khadikar,[e*] Vijay K. Agrawal,[a] and Claudiu T. Supuran[f]

[a]QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa-486 003, India
E-mail: jyoti_singh07@rediffmail.com, vijay-agrawal@lycos.com
[b]Department of Chemistry, Bareilly College, Bareilly (U.P) 243001 India
E-mail: shalini_singh15@yahoo.com
[c]Department of Chemistry, PGV Post Graduate College, Gwalior, India
[d]Department of Chemistry, Holkar Model & Autonomous College, Indore, India
[e]Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd., 3, Khatipura, Indore 452 007 India
E-mail: pvkhadikar@rediffmail.com
[f]Laboratio di Chimica Bioinorganica, Departmento di Chimica, University of Florence, via della Lastruccia, 3, RM-188,Polo Scientifico, 50019 Sesto Fioventinol, Fireze Italy
E-mail: claudiu.supuran@unifi.it

Abstract

The paper describes a first novel QSAR study on murine recombinant isozyme mCAXIII. A comparative study on modeling of inhibition of this isozyme is made using a series of topological, structural descriptors as well as their combinations. The results have shown that distance-based topological indices yield significantly better models than the structural descriptors and the combination of topological and structural descriptors. In all the three cases the Balaban type indices played a dominating role.

Keywords: Carbonic anhydrase, murine recombinant isozyme, mCAXIII, aromatic sulfonamide, heterocyclic sulfonamide, topological index, Balaban-type indices, physicochemical parameter, regression analysis

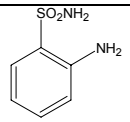
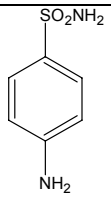
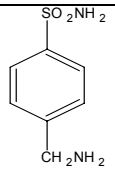
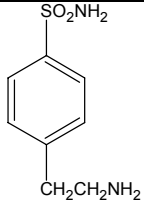
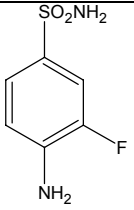
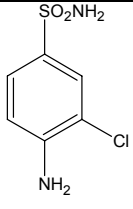
* Author for correspondence (PVK): Tel: + 91-731-2531906, E-mail: pvkhadikar@rediffmail.com

Introduction

Carbonic anhydrases (CAs, EC 4.2.1.1) occupy a special place among the In-metallo-enzymes and extensively studied by us using topological indices [1-19]. The reasons for such an extensive study being CAs are involved in crucial physiological processes. Consequent to the important role of CAs, their inhibition by carbonic anhydrase inhibitors may be explored for designing useful drugs in the management and prevention of many diseases [20-31]. The isozymes mCAXIII, which show a cytosolic sub cellular localization, is one among to 15 isozymes presently known in humans. It is also a sulfonamide inhibitable isozyme but is available in quite limited amount and this is the reason that a small number of sulfonamides has been tested. Consequently, less QSAR studies have been made / or are available in the literature related to mCAXIII. However, it is interesting to record here that mCAXIII plays an important role in pH regulation in the reproductive tract of both females and males. Such a potential of mCA is recently recognized [30,31].

Recently Supuran has reported [31] inhibition of mCAXIII with aromatic and heterocyclic sulfonamides. Out of the 32 inhibitors used by Supuran only 14 could be tested for mCA XIII. Furthermore, no QSAR study on this class of isoenzyme employing topological indices is reported in the literature. This is the main reason to carry out the first QSAR study on mCA XIII inhibitor. In the present study, therefore, we have used mCAXIII data with 14 sulfonamides [30,31], the structural details of these sulfonamides are given in Table 1. The topological indices, including Balaban and Balaban type indices were used and their calculation was made using DRAGON software [32]. These topological indices are presented in Tables 2 and 3. In addition, we have also calculated structural descriptors using ACD Labs software [33] and the same are recorded in Table 4. This set of descriptors is also used for modeling inhibitory activity of mCAXIII. Statistically significant modes were then obtained using stepwise regression analysis adopting maximum-R^2 method[34]. The results are discussed below.

Table 1. Structural details of carbonic anhysrases used for modeling $\log K_i$ (mCAXIII)

 1.	 2.	 3.
 4.	 5.	 6.

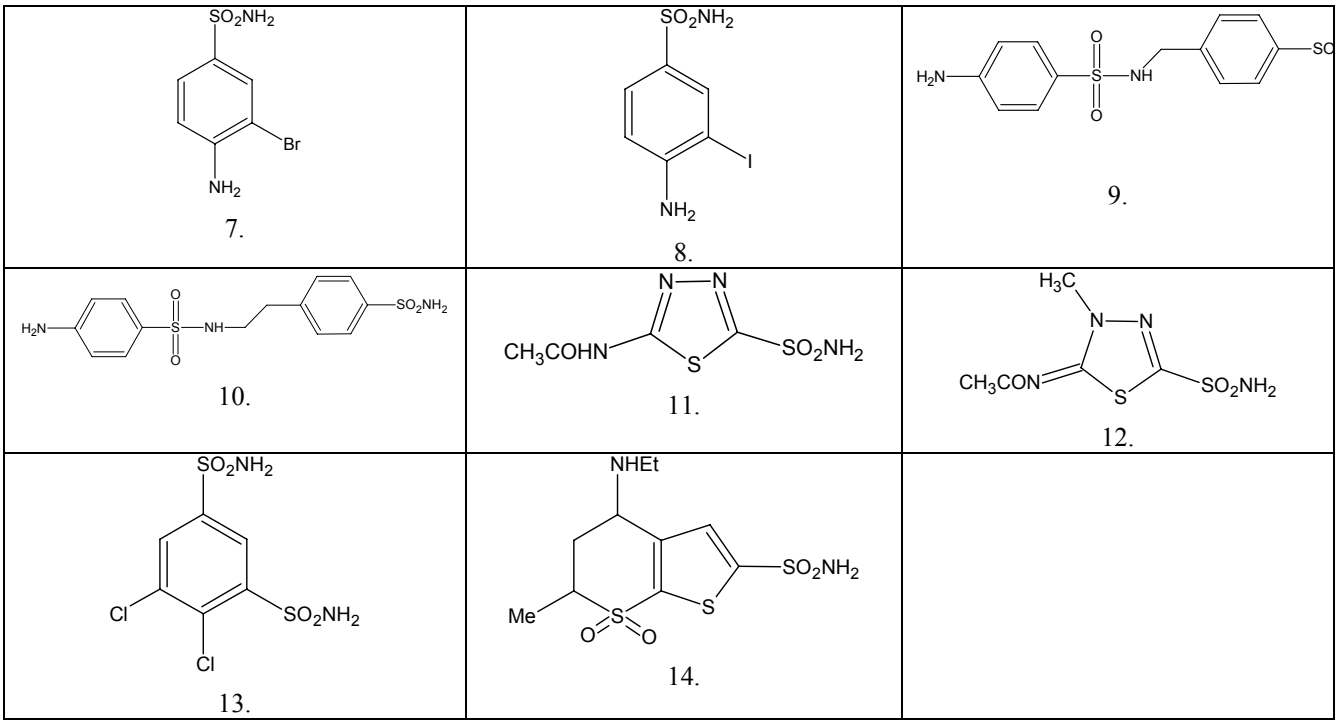


Table 2. Various topological descriptors and inhibition activity: log (mCAXIII) used in the present study and their values

Comp. No.	$\log K_i$ (mCAXIII)	W	$^{o}\chi$	$^{1}\chi$	$^{2}\chi$	$^{o}\chi^{v}$	$^{1}\chi^{v}$	$^{2}\chi^{v}$
1.	1.6334	144	8.4831	5.015	5.2343	5.6880	2.2889	2.0828
2.	1.5051	152	8.4831	4.999	5.3226	5.6880	2.8832	2.1140
3.	1.6127	201	9.1902	5.537	5.4919	6.3960	3.3563	2.3913
4.	1.6334	262	9.8973	6.037	5.8723	7.1031	3.8563	2.7276
5.	1.6989	189	9.3534	5.4097	5.8306	5.9895	2.9889	2.2283
6.	1.7481	189	9.3534	5.4097	5.8306	5.9895	2.9889	2.2283
7.	1.7323	189	9.3534	5.4097	5.8306	5.9895	2.9889	2.2283
8.	1.6989	189	9.3534	5.4097	5.8306	5.9895	2.9889	2.2283
9.	0.3010	1195	16.3805	10.1825	10.5877	11.4301	6.1533	4.5825
10.	0.2787	1408	17.0876	10.6825	10.9123	12.1372	6.6533	4.9508
11.	1.2304	257	10.605	5.8929	6.3716	6.5130	3.0317	2.0993
12.	1.2787	304	10.9307	6.3035	6.0013	7.4062	3.4272	2.4500
13.	1.3617	398	12.7238	7.0370	8.3433	7.5146	3.5790	2.8072
14.	1.2552	634	14.4223	8.6749	9.3126	10.5814	5.8318	4.5548

W- Wiener index [44]; $^{o}\chi$, $^{1}\chi$, $^{2}\chi$ –zero-, first-, and second-order Randic connectivity indices [45,46]; $^{0}\chi^{v}$, $^{1}\chi^{v}$ and $^{2}\chi^{v}$ –zero-, first- and second-order Kier and Hall Valence connectivity indices [47,48].

Table 3. Balaban and Balaban type indices used in the present study

Compd. No	J	Jhetz	Jhetv	Jhetm	Jhete	JhetP	BAC
1.	2.545	4.7880	4.789	3.092	3.614	3.531	18
2.	2.394	4.4250	4.426	2.953	3.415	3.438	18
3.	2.359	4.0008	4.009	2.895	3.257	3.216	21
4.	2.305	3.6320	3.633	2.791	3.672	3.046	22
5.	2.512	4.5260	4.54	2.878	3.592	3.083	27
6.	2.512	4.6450	4.652	3.101	3.568	3.509	27
7.	2.512	4.7200	4.73	3.105	3.553	3.562	27
8.	2.512	4.7450	4.754	3.178	3.523	3.623	27
9.	1.736	2.8280	2.828	1.882	2.65	2.068	38
10.	1.686	2.5490	2.549	1.788	2.108	1.943	38
11.	2.364	4.5530	4.555	2.161	3.007	2.372	31
12.	2.515	5.0150	5.0157	2.347	3.376	2.526	42
13.	2.859	5.8140	5.825	3.511	4.035	4.217	69
14.	2.222	3.8850	3.886	2.349	2.69	2.742	55

J- Balaban distance connectivity index [49]; Jhetz-Balaban-type index from Z-weighted distance matrix (Barysz matrix); Jhetm- Balaban-type index from mass weighted distance matrix; Jhetv- Balaban-type index from van der Waals weighted distance matrix; Jhete- Balaban-type index from electro negativity weighted distance [50] matrix; Jhetp- Balaban-type index from polarizability weighted distance matrix; BAC- Balaban centric index. [49,50]

Table 4. Values of physicochemical parameters calculated for compounds (Table 1) used in the present study

Compd.No	MW	MR	MV	α	η	γ	d	PC	MIM	NM	AM
1.	172.204	42.80	120.6	340.6	1.627	63.7	1.427	16.97	172.03	172	172.02
2.	172.204	42.80	120.6	340.6	1.627	63.7	1.427	16.97	172.03	172	172.02
3.	186.23	47.43	138.3	381.0	1.601	57.4	1.345	18.8	186.04	186	186.23
4.	200.25	52.07	154.6	421.1	1.587	54.6	1.293	20.64	200.06	200	200.26
5.	190.19	42.92	124.8	348.3	1.603	60.5	1.523	17.01	190.02	190	190.19
6.	206.64	47.63	132.6	378.0	1.637	66.0	1.558	18.88	205.99	206	206.65
7.	251.10	50.52	136.8	392.0	1.660	67.3	1.834	20.03	249.94	250	251.10
8.	298.10	55.73	147.0	414.9	1.709	71.3	2.088	22.09	297.92	298	298.10
9.	340.41	85.62	245.9	671.3	1.613	56.5	1.384	33.94	340.05	340	340.42
10.	354.44	90.25	262.4	711.4	1.603	53.9	1.350	35.77	354.07	354	354.44
11.	222.24	45.95	127.3	400.7	1.64	97.9	1.744	18.21	221.98	222	222.24
12.	236.27	53.66	131.9	394.6	1.737	80.0	1.790	21.03	236.00	236	236.27
13.	305.15	61.28	171.2	494.4	1.634	69.7	1.782	24.29	303.91	304	305.15
14.	322.44	75.48	211.0	609.0	1.634	69.9	1.530	29.29	324.01	324	324.40

Where W = Molecular weight; MR = Molar refractivity; MV = Molecular volume; α = polarizability; η = index of refraction; γ = surface tension; d = density; PC = Parachor; MIM = Monoisotopic Mass; NM = Nominal Mass; AM = Average Mass. [50]

Results and Discussion

Since the objective of our investigation is to work out relative correlation potential of topological indices at one hand and the structural descriptor at the other hand for modeling inhibitory activity, $logK_i$ (mCAXIII), we have attempted QSAR study under following three headings:

(1) QSAR study for modeling $logK_i$ (mCAXIII) using distance-based topological indices including Balaban and Balaban type indices,

(2) QSAR study for modeling $logK_i$ (mCAXIII) using structural descriptors, and

(3) QSAR study for modeling $logK_i$ (mCAXIII) based on combinations of topological and structural descriptors.

We now discuss these three types of QSAR studies.

(1) Topological modeling of $logK_i$ (mCAXIII)

The preliminary regression analysis has indicated that the inhibitory activity of mCAXIII i.e. $logK_i$ (mCAXIII) can be successfully modeled even in mono-parametric regression using J_{hetv} as the correlating parameter. Other topological indices are incapable of modeling this activity. This mono-parametric model is found as below:

$$logK_i \text{ (mCA XIII)} = -0.744 + 0.773\ (\pm 0.151)\ J_{hetv} \quad (1)$$

$$N = 14,\ Se = 0.283,\ R = 0.829,\ F = 26.353,\ Q = 2.930$$

The positive coefficient of J_{hetv} indicates that increase in van der Waals weighted distance is favorable for the exhibition of the activity.

Step-wise regression has indicated that addition of Wiener index, W; in the above model [eq. (1)] yields a model with dramatically improved statistics. The resulting bi-parametric model is found as below:

$$logK_i \text{ (mCA XIII)} = 1.116 + 0.228\ (\pm 0.107)\ J_{hetv} - 9.331 \times 10^{-4}\ (\pm 1.390 \times 10^{-4})\ W \quad (2)$$

$$N = 14,\ Se = 0.131,\ R = 0.969,\ R^2_A = 0.927,\ F = 84.111,\ Q = 7.397$$

This means that the two variable regressions yielded an excellent model. The negative sign of W may probably due to high collearnity between J_{hetv} and W. Such problems of colleanirity and how to deal with them are discussesd separately in the following section. However, occurrence of W in the above model does indicate that size, shape, and branches have significant effect on the exhibition of $logK_i$ (mCA XIII).

Looking to the sample size (**14** compounds) and following the rule of thumb [35,36] we can at the most go for tri-parametric regression analysis. In doing so, we observed that addition of $^1\chi^v$

(first-order valance connectivity index) to the above model [eq. (2)] gave a tri-parametric model as below:

$$\log K_i \text{ (mCA XIII)} = 0.645 + 0.259\ (\pm 0.093)\ J_{hetv}$$
$$- 1.391 \times 10^{-3}\ (\pm 2.370 \times 10^{-4})\ W + 0.112\ (\pm 0.068)\ {}^{1}\chi^{v} \quad (3)$$
$$N = 14,\ Se = 0.112,\ R = 0.979,\ R^2_A = 0.947,\ F = 79.142,\ Q = 8.741$$

The physical significances of J_{hetv} and W terms involved in eq. (3) are the same as discussed above for eq. (2). The positive coefficient of ${}^{1}\chi^{v}$ in this eq. (3) indicates that the presence of heteroatom and first-order branching is favorable for the exhibition of $\log K_i$ (mCAXIII). The sample size (**14** compounds) did not permit us to go to higher parametric regression analysis. However, when we did so following tetra-parametric model resulted by the addition J_{hetp}:

$$\log K_i \text{ (mCA XIII)} = 0.013 + 1.341\ (\pm 0.446)\ J_{hetv}$$
$$- 1.336 \times 10^{-3}\ (\pm 1.943 \times 10^{-4})\ W + 0.181\ (\pm 0.057)\ {}^{1}\chi^{v}$$
$$- 0.798\ (\pm 0.324)\ J_{hetp} \quad (4)$$
$$N = 14,\ Se = 0.091,\ R = 0.988,\ R^2_A = 0.965,\ F = 89.813,\ Q = 10.858$$

This four-parametric model for a set of **14** compounds made us to critically examine the rule of thumb [35,36]. This rule argues that multiple regression analysis generally requires significantly more compounds than parameters; a useful rule of thumb is three to six times the number of parameters under consideration. Hence, in case the lower limit of the rule of thumb is in favor of this four-parametric model.

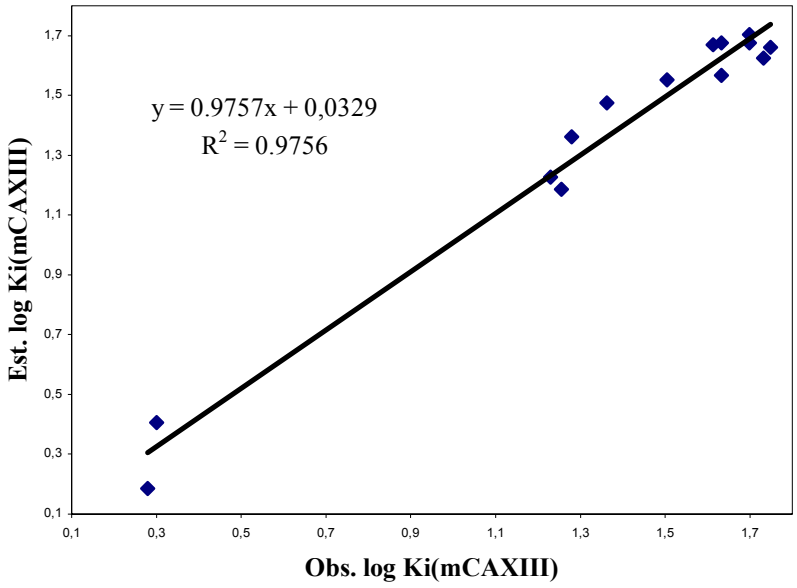


Figure1. Correlation of observed and calculated $\log K_i$ (mCAXIII) using eq (4)**.**

The physical significances of J_{hetv}, W, and $^1\chi^v$ terms involved in eq. (4) are the same as above [eq. (3)]. The negative coefficient of J_{hetp} indicates that polarizable weighted distance has a negative effect on the exhibition of $logK_i$ (mCAXIII).

The aforementioned results prompted us to comment on the intercept (i.e. constant terms), which is approaching zero as we go from mono- to tri-parametric regressions. The constant term gradually approaches to its ideal value of zero. This means that no systematic error occurred in the calculation of logK (mCAXIII). The constant term approaching the ideal value of zero also indicates that sufficient range of compounds is used in obtaining the model. We now proceed to discuss the modeling of $logK_i$ (mCA XIII) using structural descriptors.

(2) Modeling of $logK_i$ (mCA XIII) using structural descriptors

Preliminary regression analysis using structural descriptors indicated that MW is the promising parameter to be used in multi-parametric regression analysis. Out of the several structural parameters used only MW yields a sufficiently good model:

$$logK_i\ (mCAXIII) = 2.686\text{-}5.389\times10^{-3}\ (\pm1.509\times10^{-3})MW \quad (5)$$
$$N = 14,\ Se = 0.353,\ R = -0.718,\ F = 12.750,\ Q = -2.034$$

The physical significance of this and following models will be discussed separately in the latter part of this modeling.

Addition of d parameter to the above model [eq. (5)] resulted into yet another model with dramatically improved statistics. This model is found as below:

$$logK_i\ (mCAXIII) = 0.273 - 6118 \times 10^{-3}\ (\pm\ 1.196 \times 10^{-3})\ MW + 0.987\ (\pm\ 0.330)\ d \quad (6)$$
$$N = 14,\ Se = 0.273,\ R = 0.856,\ R^2_A = 0.684,\ F = 15.075,\ Q = 3.136$$

Further addition of η term to the above model [eq. (6)] a model with still improved quality resulted:

$$logK_i\ (mCAXIII) = 6.522 - 6.229 \times 10^{-3}\ (\pm\ 1.176 \times 10^{-3})\ MW + 1.528\ (\pm\ 0.0.557)\ d - 3.690\ (\pm\ 3.090)\ \eta \quad (7)$$
$$N = 14,\ Se = 0.268,\ R = 0.875,\ R^2_A = 0.696,\ F = 10.914,\ Q = 3.265$$

Finally, addition of α term to the above model resulted into the model:

$$logK_i\ (mCAXIII) = -1.462 + 0.028\ (\pm\ 0.013)\ MW - 3.588\ (\pm\ 1.969)\ d + 5.817\ (\pm\ 4.319)\ \eta - 0.356\ (\pm\ 0.134)\ \alpha \quad (8)$$
$$N = 14,\ Se = 0.211,\ R = 0.932,\ R^2_A = 0.881,\ F = 14.958,\ Q = 4.417$$

Here also, the constant term goes on decreasing as we go from mono- to higher-parametric regressions. This again indicates that no systematic error occurred in modeling the activity and that sufficient range of compound is used in obtaining the model.

(3) Modeling of $logK_i$ (mCA XIII) based on the combination of topological and structural descriptors

The present investigation will not be complete and justified unless we make further investigation to model $logK_i$ (mCAXIII) using combinations of topological and structural descriptors. Out of several such attempts we observed that the following tetra-parametric model yielded best results:

$$logK_i\ (mCAXIII) = 0.161 + 1.319(0.572)\ JhetV - 0.867(\pm 0.429)\ JhetP$$
$$- 1.4020^{-3}\ (\pm 3.505\times 10^{-4})\ W + 0.037\ (\pm 0.0\ 21)\ \alpha \qquad (9)$$
$$N = 14,\ Se = 0.115,\ R = 0.981,\ R^2_A = 0.944,\ F = 56.027,\ Q = 8.530$$

A comparison of this model [eq. [(9)] with those models represented by eq. 4 and 8 expressed that it is better than the model expressed by eq. (8) but is slightly worse than the model expressed by eq. (9). This means that in comparison to structural descriptors the topological descriptors are better suited for modeling $logK_i$ (mCAXIII).

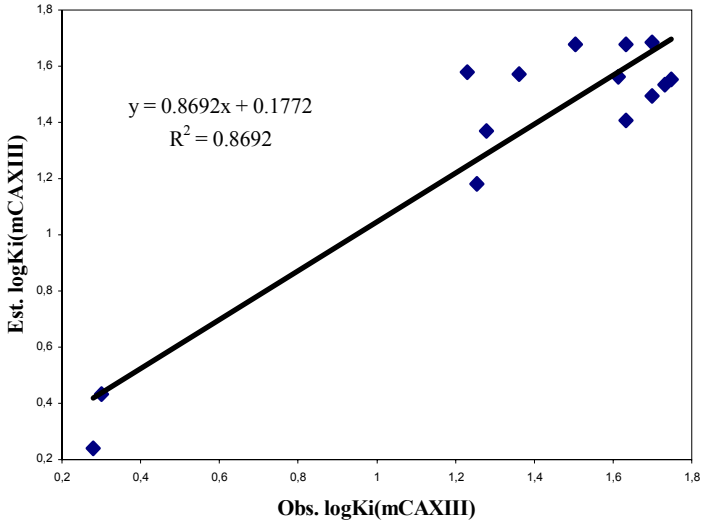


Figure 2. Correlation of observed and calculated $logK_i$ (mCAXIII) using eq. (8).

Problem of colinearity

To arrive at the final conclusion it is necessary to examine the presence of co linearity, if any, in the proposed models. The simplest way is to obtain correlation matrix in each case. The perusal of Table 6 indicates that all the proposed models suffer from the defect due to massive co linearity. Another way is to examine the Durbin-Watson test [37,38]. For this the obtained Durbin-Watson D term is used to obtain lower and upper d values i.e. to obtain the values of dl and du from the Durbin- Watson parameter D (Table 7). This can be done using some standard statistics book [34]. The results are summarized below.

Thus, the Durbin-Watson test fail to give a definite conclusion regarding the presence / absence of multi-collinearly in the proposed modes. Therefore, we have to make use of the recommendations made by Randic [39,40] for resolving the problem of co linearity.

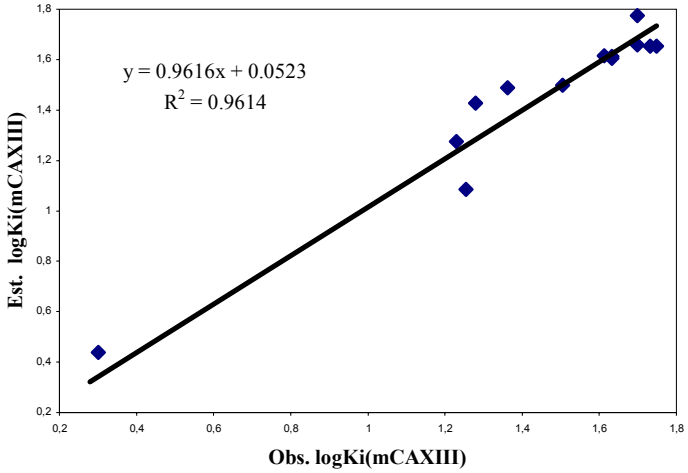


Figure 3. Correlation of observed and calculated $\log K_i$ (mCAXIII) using eq. (9).

Table 5. Regression parameters and quality of correlation

Model	Parameters	Se	R	R^2A	F	Q
	(i) Balaban Type indices and topological descriptors					
1.	Jhetv	0.283	0.829	-	26.355	2.930
2.	Jhetv, W	0.131	0.969	0.927	84.111	7.397
3.	Jhetv, $^2\chi$	0.165	0.950	0.885	50.871	5.758
4.	Jhetv, $^1\chi^v$	0.226	0.904	0.785	24.725	4.000
5.	Jhetv, W, $^2\chi$	0.137	0.969	0.920	51.051	7.073
6.	Jhetv, W, $^1\chi^v$	0.112	0.979	0.947	78.142	8.741
7.	Jhetv, $^2\chi$, $^1\chi^v$, Jhetp	0.166	0.959	0.884	25.770	5.778
8.	Jhetv, W, $^1\chi^v$, Jhetp	0.091	0.988	0.965	89.813	10.858
9.	Jhetv, W, $^2\chi$, $^1\chi^v$	0.102	0.985	0.956	72.248	9.657
	(ii) Mixed model: Balaban Type indices, topological descriptors and physicochemical parameters					
10.	Jhetv, Jhetp, W, α	0.115	0.981	0.944	56.027	8.530
11.	Jhetv, $^2\chi$, $^1\chi^v$, α	0.150	0.967	0.905	32.060	6.447
12.	Jhetv, Jhetp, $^2\chi$, α	0.181	0.951	0.862	21.230	5.255
	(iii) Physicochemical parameters					
13.	MW	0.353	-0.718	-	12.756	-2.034
14.	MW, d	0.273	0.856	0.684	15.075	3.136
15.	MW, α	0.229	0.902	0.779	23.959	3.939
16.	MW, η	0.339	0.768	0.515	7.909	2.266
17.	MW, d, α	0.220	0.918	0.796	17.883	4.173
18.	MW, d, η	0.268	0.875	0.696	10.914	3.265
19.	MW, d, η, α	0.211	0.932	0.811	14.958	4.417

Table 6. Correlation matrix for the best tri-and tetra-parametric models

	$logK_i$(mCAXIII)	W	$^1\chi^v$	Jhetv
$logK_i$ (mCAXIII)	1.000			
W	-0.956	1.000		
$^1\chi^v$	-0.860	0.941	1.000	
Jhetv	0.829	-0.759	-0.747	1.000

	$logK_i$ (mCAXIII)	MW	d	η
$logK_i$ (mCAXIII)	1.000			
MW	-0.718	1.000		
d	0.310	0.204	1.000	
η	0.184	0.121	0.816	1.000

	$logK_i$ (mCAXIII)	Jhetv	W	$^1\chi^v$	Jhetp
$logK_i$ (mCAXIII)	1.000				
Jhetv	0.829	1.000			
W	-0.956	-0.759	1.000		
$^1\chi^v$	-0.860	-0.747	0.941	1.000	
Jhetp	0.773	0.990	-0.704	-0.690	1.000

	$logK_i$ (mCAXIII)	MW	d	η	α
$logK_i$ (mCAXIII)	1.000				
MW	-0.718	1.000			
d	0.310	0.204	1.000		
η	0.184	0.121	0.816	1.000	
α	-0.878	0.914	-0.191	-0.140	1.000

Table 7. Durbin-Watson test

Model	Parameters used	D	dl	du
1 [eq. (2)]	3	1.6585	0.82	1.75
2 [eq. (4)]	3	1.2043	0.82	1.37
3 [eq. (8)]	4	1.8014	0.69	1.97
4 [eq. (9)]	4	1.5625	0.69	1.97

Randic [39,40] stated that if a descriptor strongly correlates with another descriptor already used in a regression, such a descriptor in most studies should be discarded. For example $^1\chi$ and $^2\chi$, $^1\chi$ often strongly correlates and in many structure-property-activity studies $^2\chi$ has been discarded. This is not theoretically justified and despite the widespread practice should be stopped. Although two highly correlated descriptors overall depict the same features of molecular

structure, it is important to recognize that even highly interrelated descriptors differ in some other structural traits. The difference between them may be relatively small but nevertheless very important for structure-property regression.

The criteria for inclusion or exclusion of descriptors should not be based on parallelism between descriptors even if overwhelming, but should be based on whether the part in which two descriptors disagree is or is not relevant for the characterization of the property considered .If the part in which the second descriptor differ from the first, regardless of how small it is, is relevant for the property under consideration, then the descriptor should be included. Randic [39,40] further stated that the selection of descriptors to be used in structure-property-activity studies should not be delegated solely to computers, although statistical criteria will continue to be useful for preliminary screening of 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 whether two descriptors parallel one another; i. e. duplicates much of the same structural information, but whether they are complementary in those parts that are important for structure-property-activity correlations. Hence, the residual of the correlation between two descriptors should be examined and kept or discarded depending on how well it can improve the correlation based on already selected descriptors.

Predictive power of the model

The predictive power of the model is judged by obtaining predictive correlation coefficients R^2_{pred}. This is done by plotting a graph between observed and calculated $logK_i$ (mCAXIII). We have chosen models represented by eqns.4, 8 and 9. No outlier existed in any of these models. In case of models based on topological indices only, the R^2_{pred} was found to be 0.976. while for the model based on structural descriptors only the R^2_{pred} was 0.8692. The R^2_{pred} value based on the combination of both the types if descriptors was found to be 0.9614.Once again we observed that topological indices exhibit better predictive power compared to structural descriptors. The predictive power is further confirmed by calculating Pogliani's quality factor (Q) [41-43]. The Q values are reported under each of the proposed models indicating that the predictive power goes on increasing as we go from mono- to tetra-parametric models and is highest for the latter.

Model validation

With this much discussion focusing on the process of solving the problem of interactive between variables and co linearity we now discuss model validation. This validation is required to avoid the possibility of a chance correlation. Such validation is normally done by experimental as well as regression method. In experimental validation the results are analyzed by using the model itself. The high correlation coefficient the lowest standard deviation and F values significantly grater than 90% are enough to validate the model. If the model satisfies all this requirements then it needs to be further validated using cross-validated parameters.

The estimation of probable error of coefficient of correlation (PE) is the first requirement for validating the method. This is defined as below:

$$PE = \frac{2}{3}\frac{1-r^2}{\sqrt{n}}$$

Where r (or R in multiple correlation) is the correlation coefficient and n is the number of compounds under study. It is argued that:

(i) if r (or R) < PE, then r (or R) is not significant;

(ii) if r > PE, several times; at least 3-times grater correlation is indicated, and

(iii) if r (or R) > 6PE, then the correlation is definately good.

The 6 PE data presented in Table 9 indicate that all the proposed correlations are good.

Table 8. Actual and predicted values of logKi (mCAXIII) their residue

		$logK_i$ (mCAXIII)					
		Equation-4		Equation- 8		Equation- 9	
Compd. No	Actual	Pred.	Res.	Pred.	Res.	Pred.	Res.
1.	1.633	1.566	0.067	1.678	-0.045	1.613	0.020
2.	1.505	1.551	-0.046	1.678	-0.173	1.499	0.006
3.	1.613	1.670	-0.057	1.563	0.050	1.614	-0.001
4.	1.633	1.676	-0.043	1.407	0.226	1.606	0.027
5.	1.699	1.703	-0.004	1.684	0.015	1.657	0.042
6.	1.748	1.662	0.086	1.553	0.195	1.652	0.096
7.	1.732	1.625	0.107	1.535	0.197	1.654	0.078
8.	1.699	1.675	0.024	1.495	0.204	1.774	-0.075
9.	0.301	0.406	-0.105	0.433	-0.132	0.438	-0.137
10.	0.279	0.186	0.093	0.240	0.039	0.191	0.088
11.	1.230	1.226	0.004	1.579	-0.349	1.275	-0.045
12.	1.279	1.361	-0.082	1.369	-0.090	1.426	-0.147
13.	1.362	1.475	-0.113	1.572	-0.210	1.488	-0.126
14.	1.255	1.186	0.069	1.182	0.073	1.085	0.170

In cross-validation method validation is carried out on the basis of cross-validated parameters: PRESS (Predicted residual sum of squares), SSY (Sum of the squares of the response value), r^2_{cv} (overall predictive ability), S_{press} or S_{cv} (uncertainty of prediction), and PSE or S_{pred} (predictive square error). The calculated values of parameters are shown in Table 9. We observed that in all the cases PRESS < SSY indicating that the models predicts better than chance and can be considered statistically significant. Except model 1 and 6, all other models have the ratio of PRESS/SSY smaller than 0.4 indicated them to be quite good models. Furthermore, for models-2, 3, 4 and 9 this ratio is much smaller than 0.1 that indicates all these models are excellent models. This is further confined by the values of R^2cv, S_{PRESS}, and PSE. It is important to mention that PSE is more directly related to the uncertainty of the prediction and is important in those cases in that S_{PRESS} coincide to Se. Finally, we will like to make comments

on R^2A. It takes into account of the adjustment of R^2. If a variable is added that does not contribute its fair share, then the R^2A value declines. It is a measure of the % explained variation in the dependent variable that takes into account the relation between the number of compound and the number of independent variables in the regression model. R^2A will decrease if the added variable doesn't reduce the unexplained variation enough to set the loss of degrees of freedom. A perusal of Table 5 shows that in all cases discussed above R^2A goes on increasing with the added variables.

Table 9. Cross-validated parameters

Model	Parameters	PRESS	SSY	PRESS/SSY	R^2cv	S_{PRESS}	PSE	6PE
1.	Jhetv	0.963	2.115	0.455	0.547	0.263	0.262	0.335
2.	Jhetv, W	0.189	2.889	0.065	0.935	0.126	0.116	0.066
3.	Jhetv, W, $^1\chi^v$	0.126	2.952	0.043	0.957	0.107	0.095	0.045
4.	Jhetv, W, $^1\chi^v$, Jhetp	0.075	3.003	0.025	0.975	0.087	0.073	0.013
5.	MW	1.492	1.586	0.940	0.060	0.339	0.326	0.163
6.	MW, d	0.823	2.255	0.365	0.635	0.262	0.242	0.080
7.	MW, d, η	0.720	2.358	0.305	0.695	0.258	0.227	0.250
8.	MW, d, η, α	0.402	2.676	0.150	0.850	0.200	0.169	0.141
9.	Jhetv, Jhetp, W, α	0.119	2.959	0.040	0.960	0.109	0.092	0.040

Conclusions

From the results and discussion made above we conclude the following:

(1) The development of the QSAR model on murine recombinant isoenzyme mCAXIII is rigorous and formally unexceptional, especially the choice of descriptors used by us appears particularly appropriate. The final model is predictive and the analysis can give precious hints for the understanding of is enzyme inhibition mechanisms.

(2) In spite of fact the isoenzyme m CAXIII is available in guile-limited amount, and that database is rather limited, this first QSAR study based of **14** compounds could be very useful for examining its inhibitory power;

(3) $\log K_i$ (mCA XIII) could be best modeled by topological indices, both in simple as well as multiple regression analysis;

(4) The structural descriptors can also be used successfully for modeling $\log K_i$ (mCA XIII). However, the resulting models are inferior to those obtained using topological indices;

(5) The combinations of topological and structural descriptors do not yield models better than those, which are obtained using topological indices alone;

(6) In topological modeling of $\log K_i$ (mCAXIII) Balaban type indices play a dominating role. They in combination with W and $^1\chi^v$ yield excellent multi-parametric models for modeling $\log K_i$ (mCAXIII), and

(6) In modeling $\log K_i$ (mCAXIII) using structural descriptors, MW played a dominating role for modeling $\log K_i$ (mCAXIII).

Experimental Section

(1) Inhibition constant of mCA XIII

The inhibition constant of mCA XIII, as reported earlier [30,31] were used after converting them into their logarithm, i.e., used as $\log K_i$ (mCA XIII).

(2) Topological indices

All the topological indices were calculated from the hydrogen suppressed graphs. These graphs were obtained after deleting all the carbon-hydrogen as well as heteroatom-hydrogen bonds from the molecular structures of the compounds used. The DRAGON Software [32] was used for calculating these indices. The structure optimization for using DRAGON Software was made by ACD Lab's Software [33].

(3) Regression analysis

Regression analysis was made using maximum R^2 method following stepwise regression analysis. The Regress-1 Software provided by Istvan Lukovits was used for making regression analysis.

Acknowledgements

Authors are thankful to Prof. Istvan Lukovits for providing Regress-1 Software, which is used for regression analysis. Authors are thankful to Prof. A.D.N. Bajpai, Vice-Chancellor of APS University, Rewa, India for providing research facilities and encouragement. Authors (J. Singh and VKA) are also thankful to UGC New Delhi, India for providing financial support through project No F.31-97/2005(SR).

References

1. Agrawal, V. K.; Banerji, M.; Gupta, M.; Singh, J.; Khadikar, P. V.; Supuran, C. T. *Eur.J.Med.Chem.* **2005**, *40*, 1002.
2. Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem. Letters* **2005,** *15, 203.*
3. Mandloi, D.; Joshi, S.; Khadikar, P. V.; Khosla, N. *Bioorg. Med. Chem. Letters* **2005**, *15*, 405.
4. Khadikar, P. V.; Sharma, V.; Karmarkar, S.; Supuran, C. T. *Bioorg. Med. Chem. Letters* **2005**, *15,* 931.

5. Khadikar, P. V.; Sharma, V.; Karmarkar, S.; Supuran, C. T. *Bioorg. Med. Chem. Letters* **2005***, 15,* 923.
6. Balaban, A. T.; Khadikar, P. V.; Supuran, C. T; Thakur, A.; Thakur, M.; *Bioorg. Med. Chem. Letters* **2005,** In Press.
7. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T., *Bioorg. Med. Chem. Letters* **2004,** *14*, 5661.
8. Jaiswal, M.; Khadikar, P. V.; Scozzafava, A.; Supuran, C. T., *Bioorg. Med. Chem. Letters* **2004,** *14*, 3283.
9. Agrawal, V. K.; Bano, S.; Supuran, C. T.; Khadikar, P. V. *Eur. J. Med. Chem.* **2004,** *39*, 593.
10. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2004,** *12,* 2477.
11. Saxena, A.; Agrawal, V. K.; Khadikar, P. V. *Oxid. Commun.* **2003**, *26*, 9.
12. Agrawal, V. K.; Shrivastava, S.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *11*, 5353.
13. Agrawal, V. K.; Khadikar, P. V., *Bioorg. Med. Chem. Letters*, **2003**, *13*, 447.
14. Agrawal, V. K.; Sharma, R.; Khadikar, P. V., *Bioorg. Med. Chem.* **2002**, *10*, 2993.
15. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 171.
16. Khadikar, P. V.; Clare, W. B.; Balaban, A. T.; Supuran, C. T.; Agarwal, V. K.; Singh, J.; Joshi, A. K.; Lakhwani, M. *Romania Rev.* **2005**. Accepted
17. Agrawal, V. K.; Singh, J.; Khadikar, P. V. Supuran, C. T. *Bioorg. Med. Chem*. *Letters* **2006,** *16,* 2044.
18. Khadikar, P. V.; Deeb, O.; Jaber, A.; Singh, J.; Lakhwani, M.; Agrawal, V. K. *Letters in Drug Design & Discovery* **2006,** Accepted.
19. Singh, J.; Lakhwani, M.; Khadikar, P. V.; Agrawal, V. K.; Balaban, A. T.; Clare, W. B. *Romania Rev.* **2006,** Commu.
20. *Carbonic Anhydrase-Its Inhibitors and Activators;* Supuran, C. T.; Scozzafava, A.; Conway, J.; Eds. CRC: Boca Raton, 2004; pp 1-373
21. Pastorekova, S.; Parkkila, S.; Pastorek, J.; Supuran, C. T.; *J. Enzym. Inhibin. Med. Chem.* **2004,** *19,* 199.
22. Supuran, C. T.; Vullo, D.; Manole, G.; Casini, A.; Scozzafava, A. *Curr. Med. Chem*.-*Cardiovasc. Hematol. Agents* **2004,** *2*, 49.
23. Supuran, C. T.; Scozzafava, A. *Expert opin. Ther. Pat.* **2000,** *10,* 575.
24. Supuran, C. T.; Scozzafava, A. *Expert Opin. Ther. Pat.* **2002,** *12,* 217.
25. Supuran, C. T.; Scozzafava, A. *Curr. Med. Chem.-Imm. Endocrinol. Metab.Agents* **2001,** *1,* 61.
26. Supuran, C. T.; Scozzafava, A.; Casini, A. *Med. Res. Rev.* **2003,** *23,* 146.
27. Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Expert Opin. Ther. Pat.* **2004,** *14,* 667.
28. Supuran, C. T. *Expert Opin. Ther. Pat.* **2003,** *13,* 1545.
29. Vullo, D.; Voipio, J.; Innocenti, A.; Rivera, C.; Ranki, H.; Scozzafava, A.; Kaila, K.; Supuran, C. T. *Bioorg. Med. Che. Letters* **2005**, *15,* 971; and references there in related to mCAXIII.

30. Lehtonen, J.; Shen, B.; Vihinen, M.; Casini, A.; Scozzafava, A.; Supuran, C. T.; Parkkila, A. K.; Saarnio, J.; Kivela, A. J.; Washeed, A, Sly, W. S.; Parkkila, S. *J. Biol. Chem*. **2004**, *279*, 2719.
31. Lehtonen, J.; Parkkila, S.; Vullo, D.; Casini, A.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004,** *14,* 3757.
32. Dragon software for calculation of Balaban type and other indices, www.disat.unimib.it
33. ACD-Lab software for calculating the referred physicochemical parameters; Chem .Sketch 3.O, www.acdlabs.com
34. Chaterjee, S., Hadi, A. S.; Price, B., *Regression Analysis by Examples* ,3rd Ed Wiley: New York, 2000.
35. Tute, M. S. *History and Objectives of Quantitative Drug Design in Advances in Drug Research,* Harter, N. J., Simmord, A. B., Eds., Vol. 6, Academic Press: London, 1971, pp 1.
36. Crown, H., Ed. *Comprehensive Drug Design;* Pergamon Press: New York 1990, p 19.
37. Durbin, J.; Wastson, G. S. Testing for-Serial correlation in Least Square Regression*, Biometrika*, **1950**, *37*, 49.
38. Durbin, J.; Watson, G.S. *Testing for-Serial correlation in Least Square Regression, II,* **1951**, *38*, 159.
39. Randic, M. *Acta Chem. Slov.* **1998**, *45,* 239.
40. Randic, M. *J. Chem. Inf. Comput. Sci.* **1997,** *37,* 672.
41. Pogliani, L. *Amino Acids.* **1994,** *6,* 141.
42. Pogliani, L. *J. Phys. Chem.* **1996,** *100,* 18065.
43. Pogliani, L. *Chem. Rev.* **2000,** *100,* 3827.
44. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
45. Randic, M. *J. Am. Chem. Soc.* **1975**, *97,* 6609.
46. Randic, M. *J. Mol. Graph Model* **2001**, *20*, 19.
47. Kier, L. B.; Hall, L. H.; Murray, W. J.; Randic, M. *J. Pharm. Sci.* **1975**, *64,* 1971.
48. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press: New York, 1976.
49. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89,* 399.
50. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley- VCH: Weinheim (GER), 2000.