Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 15 (2007) 6501–6509
Bioorganic & Medicinal Chemistry

QSAR studies on the activation of the human carbonic anhydrase cytosolic isoforms I and II and secretory isozyme VI with amino acids and amines

Jyoti Singh,[a] Basheerulla Shaik,[a] Shalini Singh,[b] Sarla Sikhima,[c] Vijay K. Agrawal,[a] Padmakar V. Khadikar[c,*] and Claudiu T. Supuran[d,*]

[a]*QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa 486 003, India*
[b]*Department of Chemistry, Bareilly College, Bareilly (UP) 243001, India*
[c]*Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd, 3, Khatipura, Indore 452 007, India*
[d]*Laboratorio di Chimica Bioinorganica, Dipartimento di Chimica, University of Florence, via della Lastruccia, 3, RM-188, Polo Scientifico, 50019 Sesto Fiorentino, Firenze, Italy*

Received 3 May 2007; revised 9 July 2007; accepted 12 July 2007
Available online 25 July 2007

Abstract—The first QSAR study on the activation of the human secretory isoform of the metalloenzyme carbonic anhydrase (CA, EC 4.2.1.1), CA VI, with a series of amines and amino acids is reported. A large set of topological indices have been used to obtain several tri-/tetra-parametric models. We compared the CA VI activating QSAR models with those calculated for activation of the cytosolic human isozymes hCA I and hCA II. In addition, the effect of D- and L-amino acids as activators of hCA I, hCA II and of hCA VI as compared to those of structurally related biogenic amines was investigated for obtaining statistically significant and predictive QSAR equations. The obtained models are discussed using a variety of statistical parameters. The best models were obtained for hCA II activation, followed by hCA I, whereas the QSAR models for the activation of hCA VI were statistically weaker.

1. Introduction

A multitude of physiologically relevant compounds such as amino acids, oligopeptides or small proteins, as well as many biogenic amines (histamine, serotonin and catecholamines among others), were shown to efficiently activate the catalytic activity of many isoforms of the metalloenzyme carbonic anhydrase (CA, EC 4.2.1.1) which acts as an efficient catalyst for the interconversion between carbon dioxide and bicarbonate at neutral pH.[1–3] Activation of the cytosolic, ubiquitous human isoforms hCA I and hCA II was the most investigated such phenomenon, and shown thereafter to constitute a possible therapy for the enhancement of synaptic efficacy, which may represent a conceptually new approach in the treatment of Alzheimer's disease, ageing and some other disease conditions characterized by an eventual loss of memory functions.[1] However, unlike CA inhibitors, widely used clinically for the treatment or prevention of a multitude of diseases,[1] CA activators (CAAs) have been much less investigated.[1–3] Only recently, by means of electronic spectroscopy, X-ray crystallography and kinetic measurements, it has been proved that CAAs bind within the enzyme active cavity (in the case of the physiologically most important isoforms, hCA I and hCA II) at a site distinct of the inhibitor or substrate binding sites, participating thereafter in the rate-determining step of the catalytic cycle, the proton transfer reaction between the active site and the environment.[1–3] Just recently the unique secretory hCA isoform VI (hCA VI), present in saliva and milk, has been cloned and characterized by this group,[2a] and its activation with a set of amino acids and amines also investigated.[2b]

The present paper deals with the estimation/prediction of the activation constant (log K_A) of hCA VI (measured

Keywords: Carbonic anhydrase; Activation; Human secretory isoform (CA VI); Amino acids; Amines; QSAR; Topological index; Regression analysis.
* Corresponding authors. Tel.: +91 731 2531906 (P.V.K.); tel.: +39 055 457 3005; fax: +39 055 4573385 (C.T.S.); e-mail addresses: jyoti_singh07@rediffmail.com; shalini_singh15@yahoo.com; vijay-agrawal@lycos.com; pvkhadikar@rediffmail.com; claudiu.supuran@unifi.it

0968-0896/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2007.07.004

experimentally in the above-mentioned study) as compared to the same data reported/calculated for the cytosolic isoforms hCA I and hCA II. This is the first QSAR report of a CA VI activator series, and the second ever QSAR of any CA activators, as the only other such report is the study performed in 1994 by Clare and Supuran using ab intio calculations.[3b] Consequently, we define the activation constant (K_A) as already mentioned in the original report[2b] similar to the inhibition constant K_I^1. Thus, K_A is obtained by considering the classical Michaelies–Menten equation[1,2] as given below:

$$v = v_{MAX}/\{1 + K_M/\{S\}(1 + [A]_f/K_A)\} \quad (1)$$

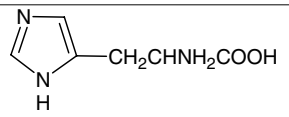
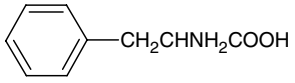
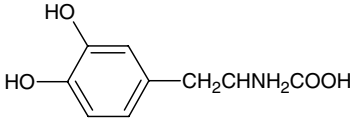
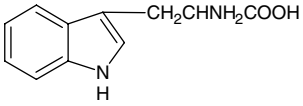
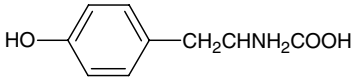
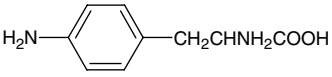
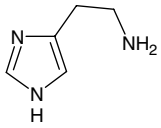
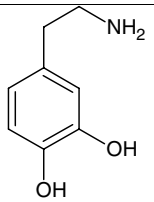
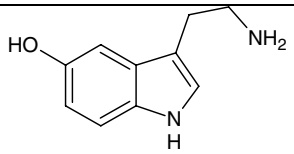
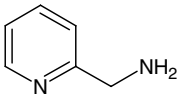
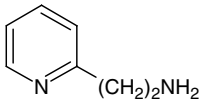
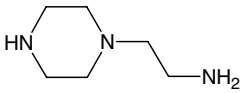
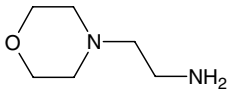
where $[A]_f$ is the free concentration of activator. Working at substrate concentrations considerably lower than K_M ($[S] \ll K_M$), and considering that $[A]_f$ can be represented in the form of the total concentration of the enzyme ($[E]_t$) and activator ($[A]_t$), the obtained competitive steady-state equation for determining the activation constant is given by Eq. 2:[2b]

$$v = v_0 \cdot K_A \Big/ \left\{ K_A + \left([A]_t - 0.5 \left\{ \left([A]_t + [E]_t + K_A \right) - \left([A]_t + [E]_t + K_A \right)^2 - 4[A]_t \cdot [E]t \right)^{1/2} \right\} \right\} \quad (2)$$

where v_0 represents the initial velocity of the enzyme-catalyzed reaction in the absence of activator.[2b]

hCA VI was shown[2] to possess a significant enzymatic activity for the physiological reaction, of the same order of magnitude as the ubiquitous isoform hCA I or the transmembrane, tumour-associated isozyme hCA IX.

Table 1. Structure of compounds **1**–**17** tested as CA activators[2b]

 L-His 1.	D-His 2.	 L-Phe 3.
D-Phe 4.	 L-DOPA 5.	D-DOPA 6.
 L-Trp 7.	D-Trp 8.	 L-Tyr 9.
 4-NH_2-L-Phe 10.	 Histamine 11.	 Dopamine 12.
 Serotonin 13.	 2-Pyridyl-methylamine 14.	 2-(2-Aminoethyl)pyridine 15.
 1-(2-Aminoethyl)piperazine 16.	 4-(2-Aminoethyl)morpholine 17.	

Thus, investigating in detail its inhibition and/or activation may be relevant for a better understanding of this isoform, and for the potential use of modulators of its activity for designing putative medicinal chemistry application.[1–3] In the above-mentioned study[2b] it has been also observed that amino acids and amines of type **1**–**17** (Table 1) act as CA VI activators, with variable efficacies, depending on their chemical structure. For example, L-His, L-Trp and dopamine showed weak CA VI activating effects, whereas D-His, D-Phe, L-DOPA, L-Trp, serotonin and pyridyl-alkyl amines were better activators. The best hCA VI activators were L-Phe, D-DOPA, 4-amino-L-Phe and histamine. All the amino acids and amines used as activators were shown to enhance k_{cat}, having no effect of K_M, and participating thus in the rate-determining step in the catalytic cycle, the proton transfer reactions between the enzyme active sites and the environment.[2b] The aforementioned findings prompted us to undertake the present investigation which is a first report of a QSAR study for CA VI activators. The study is designed by considering the 17 compounds investigated[2b] as hCA I, hCA II and hCA VI activators shown in Table 1.

2. Results and discussion

The structural details of the amino acids and the amines used in the present investigation are shown in Table 1. The D/L-form of the amino acid was designated by an indicator parameter I_1, which was taken as 1 for D-amino acids and 0 for L-amino acids. A comparison was made among the estimation of CA I, CA II and CA VI activation constants based on the obtained QSAR models.

The indicator parameter I, and hCA I, hCA II and hCA VI activation constants for the series of amino acids and amines of the type **1**–**17** are given in Table 2.[2b] The activation constants for the ubiquitous isozymes hCA I and hCA II are also given for comparison, as they will be used to derive QSAR models for the activation of these isoforms too, as only one such report is available in the literature at this moment.[3b] All the CAAs investigated up to now possess protonatable moieties of the primary amino or heterocyclic amines type, being thus able to participate to proton transfer processes leading to the generation of the nucleophilic species of the enzyme, with hydroxide coordinated to the active site zinc cation. It is interesting to mention that the amines used in this study possess amino-ethyl or amino-methyl moieties in addition to aromatic/heterocyclic groups, the last of which usually incorporates nitrogen atoms that can be protonated at pH values in the physiological range. Data of Table 2 show that the activity profile of amines and amino acids investigated here as CAAs is different for the three investigated isozymes, that is, hCA I, hCA II and hCA VI. The following sequence of activation efficiency is observed for CA I, CA II and CA VI:

(i) Activation of isoform hCA I. D-Phe > serotonin > L -Trp > D-Trp > 2-pyridyl-methylamine > dopamine > 2-[2-aminoethyl)pyridine > 1-(2-aminoethyl)-piperazine > D-DOPA > L-DOPA > histamine > 4-NH_2-L-Phe > 4-(2-aminoethyl)-morpholine > D-His > L-Phe > L-His > L-Tyr.

(ii) Activation of isoform hCA II. Histamine > serotonin > D-His > pyridyl-methylamine > L-Trp > 2-[2-aminoethyl)pyridine > D-Trp > L-DOPA > L-His > dopamine > D - DOPA > 1-(2-aminoethyl)-piperazine > 4-(2-aminoethyl)-morpholine > 4-NH_2-L-Phe > D-Phe > L-Phe > L-Tyr.

(iii) Activation of isoform hCA VI. 4-(2-Aminoethyl)-morpholine > D-Trp > L-His > dopamine > serotonin > L-DOPA=2-[2-aminoethyl)pyridine > D-His > 1-(2-aminoethyl)-piperazine > L-Tyr > histamine > 4-NH_2-L-Phe > D-DOPA > L-Phe.

From the above data it is clear that the set of amino acids and amines used here act differently in activating CA I, CA II and CA VI. In case of CA I and CA II, the activation constant for L-Tyr is the smaller one, indicating it to be the strongest activator.

We used here topological modelling of the activation constant (log K_A) for the set of 17 activators of Table 1, by using topological indices, which are numerical representation of structural data, being calculated by using the software developed by Karelson.[4] The structure optimization was made with HyperChem[5] and the ACD LAB[6] softwares. The calculated values of the topological indices are summarized in Tables 3 and 4. Stepwise regression analysis[7] using the method of maximum-R^2 was used for obtaining statistically significant models in each of the three cases (for the investigated CA isoforms). This has been done by using the NCSS software.[8] The results obtained are presented in Tables 5–7 and discussed below.

(i) Modelling activation constants of isoform I, log K_A (hCA I). Inspection of data of Table 5 shows that only three- and four-variable models were found to be statistically significant for modelling log K_A(hCA I) and that the four-variable model was the best one for this purpose. This model is shown below:

Table 2. Activation constants against isoforms hCA I, hCA II and hCA VI with derivatives **1**–**17**,[2b] and the indicator parameter I_1

Compound	log K_A(hCA I)	log K_A(hCA II)	log K_A(hCA VI)	I_1
1	−1.5228	1.0374	1.5051	0
2	−1.10457	1.6334	1.1139	1
3	−1.1549	−1.8860	0.0899	0
4	1.9344	−1.4559	1.2041	1
5	0.4913	1.0569	1.2552	0
6	0.6901	0.8920	0.6608	1
7	1.6434	1.4313	1.1760	0
8	1.6127	1.0791	1.5910	1
9	−1.6989	−1.9586	0.9689	0
10	−0.6197	−0.8239	0.7259	—
11	0.3222	2.0969	0.8129	—
12	1.1303	0.9637	1.3222	—
13	1.6532	1.6989	1.2787	—
14	1.4149	1.5314	1.1461	—
15	1.1139	1.1760	1.2552	—
16	0.8692	0.3617	0.9795	—
17	−0.8538	−0.7212	1.6232	—

Table 3. Calculated topological indices for the compound **1**–**17** investigated as CA activators

Compound	W	$^{0}\chi$	$^{1}\chi$	$^{2}\chi$	$^{3}\chi$	$^{0}\chi^{v}$	$^{1}\chi^{v}$	$^{2}\chi^{v}$	$^{3}\chi^{v}$	$^{1}\chi^{shape}$	$^{2}\chi^{shape}$	$^{3}\chi^{shape}$
1	165	8.2675	5.1983	4.6074	3.3179	5.1581	2.5262	1.5450	0.8037	8.0604	3.3800	2.2271
2	165	8.2675	5.1983	4.6074	3.3179	5.1581	2.5262	1.5450	0.8037	8.0604	3.3800	2.2271
3	212	8.9746	5.6983	4.9610	3.5679	5.7637	2.8818	1.7610	0.9367	8.9025	3.9923	2.7637
4	212	8.9746	5.6983	4.9610	3.5679	5.7637	2.8818	1.7610	0.9367	8.9025	3.9923	2.7637
5	321	10.7151	6.5029	6.0906	4.5499	6.5802	3.2901	2.1693	1.1825	10.8085	4.2961	2.8527
6	321	10.7151	6.5029	6.0906	4.5499	6.5802	3.2901	2.1693	1.1825	10.8085	4.2961	2.8527
7	369	10.8364	7.1815	6.5028	5.2942	7.2109	3.8290	2.5832	1.6339	10.0423	3.9398	1.9812
8	369	10.8364	7.1815	6.5028	5.2942	7.2109	3.8290	2.5832	1.6339	10.0423	3.9398	1.9812
9	268	9.8449	6.0922	5.5829	3.9786	6.1719	3.0859	1.9651	1.0388	9.8549	4.1337	2.9766
10	268	9.8449	6.0922	5.5829	3.9786	6.2109	3.1054	1.9846	1.0485	9.8549	4.1337	2.9766
11	67	5.8199	3.9318	2.9391	2.1716	3.8416	1.8680	1.0208	0.5479	5.4664	2.5916	1.4131
12	160	8.2675	5.2363	4.4223	3.4037	5.2637	2.6318	1.6450	0.9266	8.1990	3.4791	2.0430
13	238	9.2591	6.3088	5.4682	4.4835	6.3026	3.3749	2.2631	1.4801	8.4387	3.2978	1.4904
14	64	5.8199	3.9318	2.9122	2.3020	3.8944	1.9208	1.0590	0.5729	5.2506	2.4257	1.2931
15	94	6.5270	4.4318	3.2927	2.4216	4.3944	2.170	1.1840	0.6413	6.2333	3.1982	1.8163
16	94	6.5270	4.4318	3.2927	2.4216	4.3416	2.0916	1.1180	0.5972	6.9926	3.8211	2.3015
17	94	6.5270	4.4318	3.2927	2.4216	4.3026	2.0526	1.088	0.5787	6.9926	3.8211	2.3015

Table 4. Topological indices for the compounds **1**–**17** investigated as CA activators

Compound	TMSA	PPSA-1	PPSA-2	PPSA-3	PNSA-1
1	389.7038	108.6624	38.1753	5.4955	281.0413
2	401.4055	106.7958	37.5195	5.0843	294.6097
3	340.7005	177.5994	38.3380	1.5272	163.1010
4	340.7005	177.5994	38.3380	1.5272	163.1010
5	416.1916	115.9404	50.2132	3.5049	300.2511
6	430.4977	112.1823	48.5856	3.5345	318.3153
7	395.2202	177.5994	53.1777	2.2489	217.6207
8	412.2568	182.9860	54.7906	2.1040	229.2708
9	379.9520	130.9481	43.5166	2.9273	249.0039
10	473.4406	193.6098	53.6354	4.7728	279.8308
11	313.1795	129.2451	25.2866	5.0327	183.9343
12	382.6761	146.9584	40.4971	3.7388	235.7176
13	397.7086	161.9910	41.5362	3.8727	235.7176
14	288.9116	186.4312	25.2217	4.3376	102.4803
15	322.3225	215.0315	30.0992	4.2896	107.2909
16	399.2768	254.9598	55.5925	9.2665	144.3169
17	329.0612	234.1141	67.2446	11.7879	94.9471

Table 5. Summary of regression models for modelling activation constant (log K_A) of hCA I, hCA II and hCA VI with amino acids and amines **1**–**17**

Activators	Model	Parameters	Se	R^2	R^2_A	F
hCA I	1	$^{3}\chi^{shape}$	3.2814	0.2063	0.1533	3.898
	2	$^{2}\chi^{shape}$, $^{3}\chi^{shape}$	2.8758	0.4310	0.3497	5.302
	3	$^{2}\chi^{shape}$, $^{3}\chi^{shape}$, PPSA-2	2.7015	0.5337	0.4261	4.960
	4	$^{2}\chi^{shape}$, $^{3}\chi^{shape}$, PPSA-1, PPSA-2	2.5655	0.6118	0.4825	4.729
hCA II	5	$^{3}\chi^{shape}$	1.9385	0.5379	0.5071	17.460
	6	$^{3}\chi^{shape}$, PNSA-1	1.3176	0.8007	0.7723	28.128
	7	$^{3}\chi^{shape}$, PNSA-1, PPSA-3	1.3015	0.8195	0.7778	19.669
	8	$^{3}\chi^{shape}$, PNSA-1, $^{1}\chi^{shape}$, $^{2}\chi^{v}$	1.0392	0.8938	0.8583	25.236
hCA VI	9	$^{3}\chi^{shape}$	0.3319	0.1280	0.0699	2.202
	10	$^{3}\chi^{shape}$, PPSA-2	0.3019	0.3264	0.2302	3.393
	11	$^{3}\chi^{shape}$, PPSA-1, PPSA-2	0.3038	0.3668	0.2206	2.510
	12	$^{3}\chi^{shape}$, TMSA, PNSA-1, PPSA-2	0.3110	0.3873	0.1831	1.897

In case of modelling log K_A(hCA I) alone the indicator parameter I_1 was statistically significant and the models improved.

Table 6. Models for log K_A(hCA I) containing the indicator parameter I_1

Activator	Model	Parameters	Se	R^2	R^2_A	F
log K_A(hCA I)	13	PPSA-2, I_1	11.4517	0.5579	0.4105	3.785
	14	$^{2}\chi^{shape}$, $^{3}\chi^{shape}$, I_1	9.9920	0.7195	0.5512	4.275
	15	$^{2}\chi^{shape}$, $^{3}\chi^{shape}$, PPSA-2, I_1	9.7751	0.7852	0.5705	3.656

Table 7. Effect of L-amino acid in modelling activation constant (log K_A) of hCA I, hCA II and hCA VI

Activators	Model	Parameters	Se	R^2	R^2_A	F
hCA I	16	$^3\chi^{shape}$	4.4329	0.4509	0.4009	9.032
	17	$^2\chi^{shape}$, $^3\chi^{shape}$	3.6305	0.6652	0.5982	9.932
	18	$^2\chi^{shape}$, $^3\chi^{shape}$, TMSA	3.6604	0.6932	0.5915	6.793
	19	$^1\chi^v$, $^1\chi^{shape}$, $^3\chi^{shape}$, PPSA-1	2.7788	0.8431	0.7646	10.744
hCA II	20	$^3\chi^{shape}$	1.8698	0.6348	0.6016	19.123
	21	$^3\chi^{shape}$, TMSA	1.5792	0.7632	0.7158	16.115
	22	$^3\chi^{shape}$, TMSA, PPSA-1	1.5166	0.8034	0.7379	12.263
	23	$^3\chi^v$, $^1\chi^{shape}$, $^3\chi^{shape}$, TMSA	1.0772	0.9119	0.8678	20.689
hCA VI	24	$^3\chi^{shape}$	0.3579	0.0978	0.0158	1.192
	25	$^2\chi^{shape}$, $^3\chi^{shape}$	0.3594	0.1730	0.0076	1.046
	26	$^1\chi$, $^1\chi^v$, $^3\chi^{shape}$	0.2879	0.5225	0.3633	3.282
	27	W,$^1\chi$, $^1\chi^v$, $^3\chi^{shape}$	0.2958	0.5525	0.3288	2.470

$$\begin{aligned}\log K_A(\text{hCA I}) = &-4.2817 + 4.1976(\pm 1.2697)^2\chi^{shape} \\ &- 3.7718(\pm 0.9688)^3\chi^{shape} \\ &+ 0.0091(\pm 0.0056)\text{PNSA-1} \\ &- 0.0853(\pm 0.0372)\text{PPSA-2} \\ N = 17, \text{Se} = 2.5655, &R^2 = 0.6118, \\ R^2_A = 0.4825, &F = 4.729\end{aligned} \quad (3)$$

The physical significance of this model will be discussed separately in the following section.

(ii) Modelling activation constants of isoform I, log-K_A(hCA II). Table 5 shows that in modelling the activation constant log K_A(hCA II), again the four-parametric model was the best suited one, as shown below:

$$\begin{aligned}\log K_A(\text{hCA II}) = &\ 2.1997 - 4.8511(\pm 0.1545)^2\chi^0 \\ &- 2.0137(\pm 0.6515)^1\chi^{shape} \\ &- 5.1231(\pm 0.9448)^3\chi^{shape} \\ &+ 0.0061(\pm 0.0029)\text{PNSA-1} \\ N = 17, \text{Se} = 1.0391, &R^2 = 0.8938, \\ R^2_A = 0.8583, &F = 25.236\end{aligned} \quad (4)$$

Once again the physical significance of this model will be discussed in the following section.

(iii) Modelling activation constant of isoform VI, log-K_A(hCA VI). A perusal of data from Table 5 shows that all the four models are statistically inferior, as in each case R^2 is much smaller than 0.7. However, the data do show that a four-variable model gives better results. This model is presented below:

$$\begin{aligned}\log K_A(\text{hCA VI}) = &\ 2.0758 - 0.4411(\pm 0.1912)^3\chi^{shape} \\ &- 0.0044(\pm 0.0044)\text{TMSA} \\ &+ 0.0252(\pm 0.0.0119)\text{PPSA-2} \\ &+ 0.0027(\pm 0.0025)\text{PNSA-1} \\ N = 17, \text{Se} = 0.3110, &R^2 = 0.3873, \\ R^2_A = 0.1831, &F = 1.897\end{aligned} \quad (5)$$

We note that the statistics of this model, expressed by Eq. 5, is not really good.

From the results and discussion made above we conclude that the methodology used here is excellent for modelling log K_A(hCA II), it is little bit inferior for modelling log K_A(hCA I) while it is worse for modelling log K_A(hCA VI). It should be noted that none of the models contain the indicator parameter I_1. When we forced this parameter I_1 to occur in the models, we observed an improvement in the statistics only in modelling log K_A(hCA I) (Table 5). In other cases I_1 had coefficients considerably smaller than its standard deviation (data not shown). Such models are not allowed statistically.[8] It seems that for modelling log K_A(hCA II) and log K_A(hCA I) it does not matter if the amino acid activator is present in the D- or L-enantiomeric form, while it does matter while modelling log K_A(hCA VI). This is in fact in agreement with X-ray crystallographic data on hCA I and hCA II in adducts with amino acid activators, showing that both the L- or D-activators bind efficiently within the enzyme cavity, but in rather different ways.[9]

Data of Table 6 indicate that all the three models contain I_1 as the correlating parameter and that all are statistically significant. The model containing $^2\chi^{shape}$, and $^3\chi^{shape}$, PPSA-2 and I_1 as the correlating parameters is the most appropriate one for calculating log K_A(hCA I). This model is represented by the following equation:

$$\begin{aligned}\log K_A(\text{hCA I}) = &\ 9.5318(\pm 4.0807) \\ &+ 7.8659(\pm 3.8243)^2\chi^{shape} \\ &- 5.6599(\pm 2.8669)^3\chi^{shape} \\ &- 0.1675(\pm 0.1514)\text{PPSA-2} \\ &+ 1.02777(\pm 0.6614)I_1 \\ N = 17, \text{Se} = 0.7751, &R^2 = 0.7852, \\ R^2_A = 0.5708, &F = 3.656\end{aligned} \quad (6)$$

2.1. Effect of amino acids and amine activators

The aforementioned results obtained by using Eq. 6 prompted us to investigate the modelling of activation

Table 8. Effect of D-amino acid in modelling activation constant (log K_A) of hCA I, hCA II and hCA VI

Activators	Model	Parameters	Se	R^2	R^2_A	F
hCA I	28	$^3\chi^v$	1.4844	0.1294	0.0424	1.487
	29	$^2\chi, ^1\chi^v$	1.2114	0.4782	0.3623	4.125
	30	$^0\chi, ^2\chi, ^2\chi^v$	1.1423	0.5876	0.4330	3.800
	31	$^0\chi, ^2\chi, ^2\chi^v, ^1\chi^{shape}$	1.0471	0.6968	0.5235	4.021
hCA II	32	$^2\chi^{shape}$	1.2539	0.4491	0.3940	8.152
	33	$^3\chi, ^0\chi^v$	1.0010	0.6841	0.6138	9.743
	34	$^3\chi, ^1\chi^v, ^2\chi^{shape}$	0.9807	0.7305	0.6294	7.226
	35	$^1\chi, ^1\chi^v, ^3\chi^v, ^2\chi^{shape}$	0.9515	0.7780	0.6511	6.132
hCA VI	36	$^3\chi^v$	0.2786	0.0261	0.0000	0.268
	37	$^2\chi^v, ^3\chi^v$	0.2445	0.3252	0.1753	2.169
	38	$^2\chi^v, ^3\chi^v, ^0\chi^v$	0.2248	0.4930	0.3029	2.593
	39	$^0\chi^v, ^2\chi^v, ^3\chi^v, ^2\chi^{shape}$	0.2255	0.5533	0.2981	2.168

constant more thoroughly, by considering the effect of D- and L-amino acids (individually) together with the activation by amines only. We considered (i) modelling of log K_A, using only L-amino acids and the amines; (ii) modelling of log K_A using only D-amino acids and amines. The obtained results are summarized in Tables 7 and 8, for L-amino and D-amino acids, respectively.

2.2. Effect of L-amino acids and amine activators

We observed (Table 7) considerable improvement in statistics in all the three cases when we have used L-amino acids and amines together. In the case of modelling log K_A(hCA VI) we now obtained two statistically significant models (Nos. 26 and 27).

(a) Modelling of log K_A(hCA I). A perusal of Table 7 shows that there are three statistically significant models, out of which the four-variable model is the best for modelling log K_A(hCA I). This model is shown below:

$$\begin{aligned}\log K_A(\text{hCA I}) = {} & 8.5550 + 2.6052(\pm 0.9091)^1\chi^{shape} \\ & - 4.9903(\pm 1.1298)^3\chi^{shape} \\ & + 0.0176(\pm 0.0046)\text{PPSA-1} \\ & - 2.5851(\pm 1.1381)^1\chi^{shape} \\ N = 13, \text{Se} = 2.7788, & R^2 = 0.8431, \\ R^2_A = 0.7646, & F = 10.744\end{aligned} \quad (7)$$

(b) Modelling of log K_A(hCA II). In the case of modelling log K_A(hCA II) all the four proposed models (Table 7) are statistically significant and the four-variable model yielded excellent results. This model is presented below.

$$\begin{aligned}\log K_A(\text{hCA II}) = {} & 0.3457 - 6.6142(\pm 1.8429)^3\chi^v \\ & + 1.9653(\pm 0.5363)^1\chi^{shape} \\ & - 6.0735(\pm 1.0295)^3\chi^{shape} \\ & + 0.0095(\pm 0.0045)\text{TMSA} \\ N = 13, \text{Se} = 1.0772, & R^2 = 0.9119, \\ R^2_A = 0.8678, & F = 20.689\end{aligned} \quad (8)$$

(c) Modelling of log K_A(hCA VI). Table 7 shows that in modelling log K_A(hCA VI) two statistically significant models were obtained. As in both the cases statistics is more or less similar the former model containing smaller number of variables is considered best for modelling log K_A(hCA VI). This three-variable model is found as below:

$$\begin{aligned}\log K_A(\text{hCA VI}) = {} & -1.8023 - 7.5855(\pm 2.6923)^1\chi^v \\ & - 1.0277(\pm 0.3331)^3\chi^{shape} \\ & + 4.7645(\pm 1.6850)\text{TMSA} \\ N = 13, \text{Se} = 0.2879, & R^2 = 0.5225, \\ R^2_A = 0.3633, & F = 3.282\end{aligned} \quad (9)$$

A close examination of the four-variable model of Table 7 indicated that it contained the Wiener index (W) as one of the correlating parameters, whose coefficient (0.0044) was considerably smaller than its standard deviation (0.0060). Such models are not allowed statistically.[8] This is another supporting result for not considering a four-variable model for modelling log K_A(hCA VI). Hence, as discussed above, it is a three-variable model, the most appropriate one for modelling log K_A(hCA VI).

2.3. Effect of D-amino acids and amine activators

We shall discuss in the following the combination of D-amino acids with amines for modelling the activation constants. The results obtained are presented in Table 8. We observed that except for modelling log K_A(hCA VI), the results are inferior than in the cases discussed above for L-amino acids and amine activators.

(i) Modelling of log K_A(hCA I). Table 8 show that both three- and four-variable models are statistically significant and the latter gives better results according to the following equation:

$$\begin{aligned}\log K_A(\text{hCA I}) = {} & -10.6547 + 6.4971(\pm 2.2468)^1\chi^v \\ & - 1.5611(\pm 0.8867)^1\chi^{shape} \\ & + 6.3564(\pm 2.7914)^0\chi \\ & - 9.9470(\pm 3.5526)^2\chi^v \\ N = 12, \text{Se} = 1.0481, & R^2 = 0.6968, \\ R^2_A = 0.5235, & F = 4.021\end{aligned} \quad (10)$$

(ii) Modelling of log K_A(hCA II). Data of Table 8 also show that there are three statistically significant models, out of which a four-variable model is the best. This model is shown below:

$$\begin{aligned}\log K_{\mathrm{A}}(\text{hCA II}) &= 2.4960 - 16.6716(\pm 7.2341)^{1}\chi^{v} \\ &\quad + 6.5967(\pm 3.9526)^{3}\chi^{v} \\ &\quad - 2.70806(\pm 0.8746)^{2}\chi^{\text{shape}} \\ &\quad + 8.6825(\pm 3.8628)^{1} \\ N = 12, \mathrm{Se} &= 0.9515, R^2 = 0.7780, \\ R^2_{\mathrm{A}} &= 0.6511, F = 6.132\end{aligned} \tag{11}$$

(iii) Modelling of log K_A(hCA VI). Data of Table 8 shows that there are two statistically significant models for modelling log K_A(hCA VI), of which a four-variable model is found to be the best. This model is presented below.

$$\begin{aligned}\log K_{\mathrm{A}}(\text{hCA VI}) &= -2.7253 + 3.3348(\pm 2.2934)^{0}\chi^{v} \\ &\quad - 10.7355(\pm 5.8500)^{2}\chi^{v} \\ &\quad + 6.2781(\pm 2.4137)^{3}\chi^{v} \\ &\quad - 0.5853(\pm 0.6019)^{2}\chi^{\text{shape}} \\ N = 12, \mathrm{Se} &= 0.2255, R^2 = 0.5533, \\ R^2_{\mathrm{A}} &= 0.2981, F = 2.168\end{aligned} \tag{12}$$

We observed that in case of $^{2}\chi^{\text{shape}}$ its coefficient is smaller than the corresponding standard deviation. In view of this we have to consider the smaller parametric model, that is, we need to examine three-variable models for obtaining statistically significant QSARs for this isozyme. Such a three-variable model is shown below (Table 8):

$$\begin{aligned}\log K_{\mathrm{A}}(\text{hCA VI}) &= -4.6720 + 1.2301(\pm 0.7561)^{0}\chi^{v} \\ &\quad - 5.6469(\pm 2.6066)^{2}\chi^{v} \\ &\quad + 4.6730(\pm 1.7551)^{3}\chi^{v} \\ N = 12, \mathrm{Se} &= 0.2248, R^2 = 0.4930, \\ R^2_{\mathrm{A}} &= 0.3029, F = 2.593\end{aligned} \tag{13}$$

These results prove that for modelling log K_A(hCA VI) the presence of D-, L- and/or both enantiomers in the model is beneficial.

We shall examine all the 39 models presented in Tables 5–8. This examination is based on the variation in R^2 and R^2_A as we proceed from one-variable to higher variable models. A close look at Tables 5–8 indicates that (except for the models 27 and 39) in each category, R^2 and R^2_A increase with the increasing number of correlating parameters. Such an increase should be expected, as R^2 always increases with increase in the number of correlating parameters.[10,11] However, this is not the case with R^2_A. It increases only when the added parameter contributes their fair share to the model; otherwise R^2_A declines.[10,11] In the case of our model 26, containing $^{1}\chi$, $^{2}\chi^{v}$ and $^{3}\chi^{\text{shape}}$ as the correlating parameters, R^2_A is 0.3622. When a new parameter, namely the Wiener index (W), was added to this model, the resulting four-variable model 27 has an R^2_A of 0.3288. That is, due to the addition of W, when R^2_A is diminished, signifying that the added parameter does not contribute a fair share to the model. This is also the case with our model 39, when a new parameter $^{2}\chi^{\text{shape}}$ was added to the model; R^2_A being reduced from 0.3029 to 0.2981. This means that addition of $^{2}\chi^{\text{shape}}$ is again not well justified. At this point, it is worth mentioning that R^2_A relates with R by the following expression:

$$R^2_{\mathrm{A}} = 1 - (1 - R^2)\{(n-1)/(n-k-1)\} \tag{14}$$

where n is the sample size and k is the number of independent variables in the models.

Obviously, the R^2_A value takes into account the adjustment of R^2. Therefore, if a variable is added that does not contribute its fare share, the R^2_A will actually decline. R^2_A is particularly important when the number of independent variables is large relative to the sample size.[10,11] When taking into account the relationship between sample size and the number of variables, R^2 may appear artificially high if the number of variables is high compared to the sample size. R^2_A is a measure of % explained variation in the dependent variable that takes into account the relationship between the number of cases (compounds) and number of independent variables in the regression models. Whereas R^2 will show an increase when an independent variable is added, R^2_A will decrease if the added variable does not reduce the unexplained variation enough to offset the loss of degrees of freedom.[10,11] A perusal of Table 5–8 shows that in each of the 39 models, the majority of the correlating parameters are connectivity or connectivity type indices, in which collinearity is a well-documented problem.[3b,10] Therefore, all these models exhibit statistical collinearity defects. Application of ridge statistics has indicated that in each model there are two or more variables whose variable inflation factor (VIF) values are considerably larger than 10, which is indicative of the occurrence of multicollinearity problems. However, the multicollinearity problem will be resolved by using the recommendation of Randic.[11] Randic[11,12] stated that if a descriptor strongly correlates with another descriptor already used in a regression, such a descriptor should be discarded in most studies. For example, among descriptors $^{1}\chi$ and $^{2}\chi$, the first one often strongly correlates with the second one, and as a consequence in QSAR studies $^{2}\chi$ should be discarded.[10,11] Although two highly correlated descriptors depict the same features of molecular structure, it is important to recognize that even highly interrelated descriptors differ in certain structural traits. The difference between them may be relatively small but nevertheless important for structure–property regressions. 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 considered property. If the part in which the second descriptor dif-

fers from the first, regardless of how small it is, is relevant for the property under consideration, then the descriptor should be included. Randic[11] further stated that the selection of descriptors to be used in QSAR 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 the two descriptors parallel one another; that is, duplicate 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.[11,12]

Finally, let us comment on the predictive powers of the obtained models highlighted in Tables 5–8. The easiest way to examine the predictive power is to calculate the Pogliani factor Q. This quality factor Q is defined in the literature[13–15] as the ratio of correlation coefficient R to the stranded error of estimation (Se), for example, $Q = R/\text{Se}$. That means that the higher is the value of R, and lower the value of Se, the larger will be Q and the better will be the predicting power of the model. The calculated Q values for various equations of our work are summarized below.

Model	R	Se	Q	Activity
4	0.7822	2.5655	0.3048	log K_A(hCA I)
8	0.8862	0.9775	0.9060	log K_A(hCA I)
12	0.9182	2.7788	0.3304	log K_A(hCA I)
15	0.8348	1.0471	0.7972	log K_A(hCA I)
19	0.9454	1.0392	0.9097	log K_A(hCA II)
23	0.9549	1.0772	0.8865	log K_A(hCA II)
26	0.8820	0.9515	0.9270	log K_A(hCA II)
31	0.6223	0.3112	1.9997	log K_A(hCA VI)
35	0.7228	0.2879	2.5107	log K_A(hCA VI)
38	0.7014	0.2248	3.1200	log K_A(hCA VI)

From the above data we arrive at the following conclusions:

(i) Among the models for correlating log K_A(hCA I), model 15 has both excellent statistics as well as excellent predictive power.
(ii) Model 35 which correlates log K_A(hCA II) has the highest predictive power, but it has slightly inferior statistics as compared to model 8.
(iii) Among the models for correlating log K_A(hCA VI), model 38 has the highest predictive power.

In order to confirm our results we have calculated the activities using the corresponding models and compared them with the observed activity. The results have shown that the calculated activities in each case were very much nearer to the experimentally determined ones.

3. Conclusions

This is first QSAR for modelling activation constants for a series of CA VI activators. Furthemore, for the same set of CAAs, models were obtained for activation of the cytosolic isoforms CA I and CA II, and such models have been compared between them. This comparative study has shown that the methodology is best suited for modelling log K_A(hCA II), is slightly inferior for modelling log K_A(hCA I) and is worse for modelling log K_A(hCA VI). The scarcity of QSAR models for CA activators (as compared to the large number of such QSARs for CA inhibitors)[16] warrants further studies in this field.

4. Experimental

4.1. Activation constants

All the activation constants were reported earlier as $K_A$values[2b] and were converted to their log units to be used in the present study.[1–3]

4.2. Calculation of topological indices

All the topological indices were calculated using the software of Karelson.[4] The structure optimization was performed using HyperChem[5] and the ACD Labs[6] softwares.

4.3. Regression analysis

Stepwise regression analysis adopting maximum R^2 method[7] was performed using NCSS software.[8]

Acknowledgments

This work was financed in part by an EU project of the 6th framework programme (DeZnIT project, contract No. LSHB-CT-2007-037303, to C.T.S.). One of the authors (S.S.) expresses her thanks to the Department of Science & Technology, the Government of India, New Delhi, for awarding a DST project SR/WOS-A/CS/61/2004 under Woman Scientists scheme, and to the Principal for his personal interest and providing the facility to carry out this work.

References and notes

1. (a) *Carbonic Anhydrase-Its inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC Press: Boca Raton, Fl.N.Y., 2004; pp 1–363; (b) Pastorekova, S.; Parkkila, S.; Pastorek, J.; Supuran, C. T. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 199; (c) Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Expert Opin. Ther. Pat.* **2004**, *14*, 667; (d) Supuran, C. T.; Scozzafava,

A.; Casini, A. *Med. Res. Rev.* **2003**, *23*, 146; (e) Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Expert Opin. Ther. Pat.* **2006**, *16*, 1627.
2. (a) Nishimori, I.; Minakuchi, T.; Onishi, S.; Vullo, D.; Scozzafava, A.; Supuran, C. T. *J. Med. Chem.* **2007**, *50*, 381; (b) Nishimori, I.; Vullo, D.; Innocenti, A.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem.* **2007**, *15*, 5351.
3. (a) Supuran, C. T. Carbonic Anhydrase: Catalytic and Inhibition Mechanism Distribution, and Physiological Roles. In *Carbonic Anhydrase—Its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC Press: Boca Raton, Fl, 2004; pp 1–24; (b) Clare, B. W.; Supuran, C. T. *J. Pharm. Sci.* **1994**, *83*, 768–773.
4. Karelson, M. ChemAxon softwares for the calculation of topological indices (hettp://www.chemaxon.com).
5. HYPERCHEM, Release 7.03 for Windows (2002) Hypercube, Inc., Aorida:USA.
6. ACD-Labs software for calculating the referred physicochemical parameters; Chem Sketch 3.0, www.acdlabs.com.
7. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
8. NCSS, http://www.ncss.com.
9. (a) Temperini, C.; Scozzafava, A.; Puccetti, L.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5136; (b) Temperini, C.; Vullo, D.; Scozzafava, A.; Supuran, C. T. *J. Med. Chem.* **2006**, *49*, 3019; (c) Temperini, C.; Scozzafava, A.; Vullo, D.; Supuran, C. T. *Chemistry* **2006**, *12*, 7057; (d) Temperini, C.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5152.
10. (a) Singh, J.; Lakhwani, M.; Khadikar, P. V.; Balaban, A. T.; Clare, B. W.; Supuran, C. T. *Rev. Roum. Chim.* **2006**, *51*, 691; (b) Agrawal, V. K.; Singh, J.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2044; (c) Agrawal, V. K.; Banaerji, M.; Gupta, M.; Singh, J.; Khadikar, P. V.; Supuran, C. T. *Eur. J. Med. Chem.* **2005**, *40*, 1002; (d) Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2004**, *14*, 5661.
11. Randic, M. *Acta Chem. Slov.* **1998**, *45*, 239.
12. Randic, M. *J. Chem. Inf. Comput. Sci* **1997**, *37*, 672–687.
13. Pogliani, L. *Amino Acids* **1994**, *6*, 141–153.
14. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065–18077.
15. Pogliani, L. *Chem. Rev.* **2000**, *100*, 3827–3858.
16. (a) Clare, B. W.; Supuran, C. T. QSAR Studies of Carbonic Anhydrase Inhibitors. In *Carbonic Anhydrase—Its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC Press: Boca Raton, Fl, 2004; pp 149–182; (b) Hillebrecht, A.; Supuran, C. T.; Klebe, G. *ChemMedChem* **2006**, *1*, 839; (c) Weber, A.; Bohm, M.; Supuran, C. T.; Scozzafava, A.; Sotriffer, C. A.; Klebe, G. *J. Chem. Inf. Model.* **2006**, *46*, 2737; (d) Tuccinardi, T.; Nuti, E.; Ortore, G.; Supuran, C. T.; Rossello, A.; Martinelli, A. *J. Chem. Inf. Model.* **2007**, *47*, 515.