Cent. Eur. J. Chem. • 7(4) • 2009 • 909–922
DOI: 10.2478/s11532-009-0073-4

VERSITA

Central European Journal of **Chemistry**

CP-MLR directed QSAR study of carbonic anhydrase inhibitors: sulfonamide and sulfamate inhibitors

Research Article

Brij Kishore Sharma[1], Pradeep Pilania[1], Prithvi Singh[1], Susheela Sharma[2,*], Yenamandra S. Prabhakar[3]

[1]*Department of Chemistry, S.K. Government College, Sikar-332 001, India*

[2]*Department of Engineering Chemistry, Sobhasaria Engineering College, Sikar-332 021, India*

[3]*Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow-226 001, India*

Received 21 January 2009; Accepted 23 June 2009

Abstract: The inhibition activities of sulfonamide and sulfamate derivatives for human carbonic anhydrases have been quantitatively analyzed using DRAGON descriptors. QSAR models have been obtained through combinatorial protocol-multiple linear regression (CP-MLR) computational procedure. For the hCA I inhibition activity, a higher value of information content index of the 1-order neighborhood symmetry (IC1) and a lower value of the Moran autocorrelations, MATS2v and MATS1p, along with a lower number of sulfur atoms in a molecular structure (nRSR) is beneficial to the activity. A higher number of 5-membered rings (nR05), a bigger distance between nitrogen and sulfur T(N..S), and a higher value of van der Waals volume weighted descriptor (GATS6v), are helpful to improve the hCA II inhibition activity. For the inhibition of hpCA, a lower value of the descriptors Jhetv and PW5, and higher values of the eigenvalue sum from Z weighted distance matrix, SEigZ, the Moran autocorrelation of lag 8 weighted by atomic van der Waals volumes, MATS8v and the Moran autocorrelation of lag 4 weighted by atomic Sanderson electronegativities, MATS4e are favorable. The derived significant models in such descriptors may further be used to synthesize new potential compounds and to decipher the mode of their actions at molecular level.

Keywords: *Quantitative structure-activity relationship (QSAR) • Inhibition activities of sulfonamide and sulfamate derivatives for human carbonic anhydrases (hCA I and hCA II) and α-carbonic anhydrase from Helicobacter pylori (hpCA) • Combinatorial protocol in multiple linear regression (CP-MLR) analysis • DRAGON descriptors*

© Versita Warsaw and Springer-Verlag Berlin Heidelberg.

1. Introduction

Carbonic anhydrases are comprised of a family of monomeric zinc metalloenzymes. These enzymes catalyze the reversible hydration of CO_2 to bicarbonate and a proton [1-5]. Several forms of carbonic anhydrase occur in nature and there are at least five distinct CA families (α, β, γ, δ and ε) reported in the literature. These families possess no significant amino acid sequence similarity. In most cases they are thought to be an illustration of convergent evolution. Presently, 16 isoforms of human carbonic anhydrase (hCA) have been cloned and a variety of compounds were analyzed for their inhibitory effects on most of them [6-8]. Many of these isoenzymes such as hCA II, IV, VA, VB, VII, IX, XII, XIII and XIV represent valid targets for the development of novel antiglaucoma, antitumor, antiobesity or anticonvulsant drugs [9-13].More recently, the representatives of the α- or β-CA class have also been cloned and characterized in other organisms,

* *E-mail: susheela_13@yahoo.co.in*

namely *Plasmodium falciparum* [14], *Mycobacterium tuberculosis* [15], *Cryptococcus neoformans* [16], or *Candida* spp. [17]. These enzymes have been proved to be critical for the growth or virulence of these pathogens. The varieties of these organisms are highly pathogenic and exhibit different degree of resistance to the available drugs targeting them. Thus, the inhibition of their CAs may represent novel approaches to fight such ailments [14-17].

Helicobacter pylori, a Gram-negative neutralophile [18] was shown to be related with chronic gastritis, peptic ulcers and gastric cancer, the second most common tumor in human [19]. *H. pylori* is a widely spread pathogen, causing sometimes severe gastrointestinal diseases leading to a significant morbidity and mortality [20]. The first-line treatment of peptic ulcer disease caused by *H. pylori* involves a therapy with two antibiotics (amoxicillin and clarithromycin) and a proton pump inhibitor (PPI). This treatment may, however, fail due to an increase in the prevalence of antibiotic resistance [21-23]. Thus, an empirical quadruple regimen (PPI, bismuth, tetracycline and metronidazole) has generally been recommended as the second-line therapy. But, several studies have shown that even two successive treatments failed to get rid of *H. pylori* in some cases [21-23]. For this reason, there is an urgent need for alternative therapies that could target the root cause and be devoid of the problems arising with presently available drugs. *H. pylori* may grow and multiply in the stomach, in the highly acidic conditions at pH < 1.4 [24]. Thus, the pathogen has emerged in the specific processes that sustain the cytoplasmic pH nearly 6.4 for survival and growth. In fact, two enzymes are involved in these processes: urease [24] in the cytoplasm and an α-CA (designated as hpCA) in the periplasm [24,25], which separates the outer and inner membranes of this bacterium. The β-CA has also been found in the cytoplasm of *H. pylori*, where it appears to play a crucial role in the urea-bicarbonate metabolism and acid resistance of the pathogen [26]. In fact, the study of Sachs' group [24,25] has provided the proof-of-concept that hpCA may be an attractive drug target for developing anti-*H. pylori* agents.

More recently, Nishimori *et al.* [27] have sequenced hpCA DNAs from a large panel of independent strains of *H. pylori*, which were obtained from patients with a variety of gastric mucosal lesions including gastritis, gastric ulcer and gastric cancer. They have also evaluated the inhibitory effects of a panel of sulfonamides/sulfamates (known inhibitors of other α-CAs) against this enzyme and inferred that effective inhibitors targeting this bacterial CA can be designed. A library of sulfonamides/sulfamates was investigated for the inhibition of hpCA, in addition to new derivatives obtained by attaching 4-*tert*-butyl-phenylcarboxamido/sulfonamide tails to benzenesulfonamide/ 1,3,4-thiadiazole-2-sulfonamide scaffolds. The inhibition data of a total of 47 such compounds against the host isozymes, hCA I and hCA II, and the bacterial enzyme hpCA have been reported. However, the study was directed at the alteration of either substituents in parent moiety or variation of main scaffolds but no rationale was provided to account for the inhibition data against three enzyme systems. The aim for present communication is, therefore, to establish quantitative relationships between inhibition data of these enzymes and descriptors unfolding the structural variations of all reported congeners.

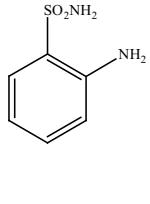
2. Materials and methods

2.1. Data set

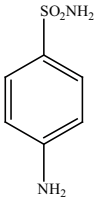
The reported compounds [27] have been represented in Schemes 1, 2 and 3. Compounds (**1-24**) given in Scheme 1, are representative of simple aromatic and heteroaromatic sulfonamides, while compounds (**25-35**) included in Scheme 2 are clinically used drugs. The analogues (**36-47**), incorporated in Scheme 3, are newly synthesized compounds obtained by attaching 4-*tert*-butyl-phenylcarboxamido/sulfonamide tails to benzenesulfonamide/1,3,4-thiadiazole-2-sulfonamide scaffolds. The inhibition activity data, reported in terms of binding constant, K_i, of all 47 compounds were obtained against the host isozymes hCA I and hCA II and the bacterial enzyme hpCA. For the purpose of QSAR study, all analogues have been randomly divided into training and test sets. Out of the 47 analogues, 10 compounds have been placed in the test set for the external validation of derived models. The training and test set compounds along with their biological actions (expressed as pK_i on molar basis to guarantee the linear dependence of dependent variable on structure accounting descriptors) are listed in Tables 1 and 2, respectively.

2.2. Computational procedure

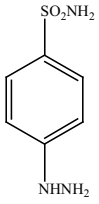
The DRAGON software [28] has been used for the parameterization of the compounds under study. This software offers several hundreds of descriptors from different perspectives relating to empirical, constitutional and topological indices characteristic to the molecules under multi-descriptor class environment. The structures of the compounds under study have been drawn in 2D ChemDraw [29] using the standard procedure. These structures were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D-structures of the



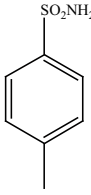
1



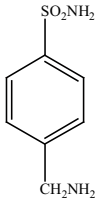
2



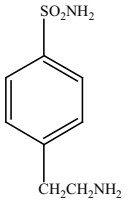
3



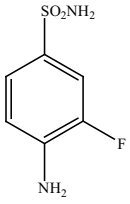
4



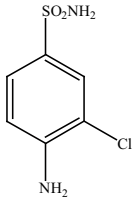
5



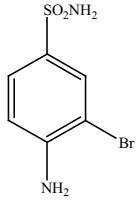
6



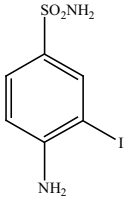
7



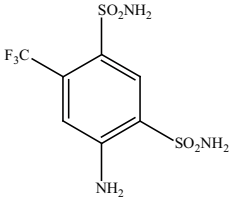
8



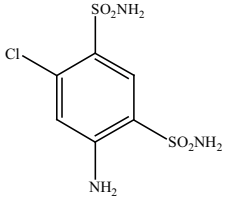
9



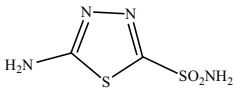
10



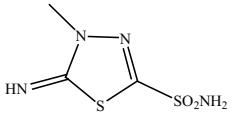
11



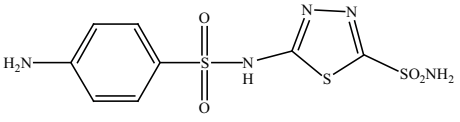
12



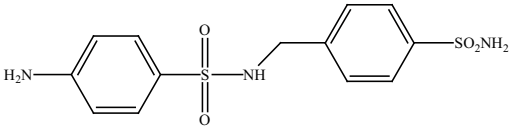
13



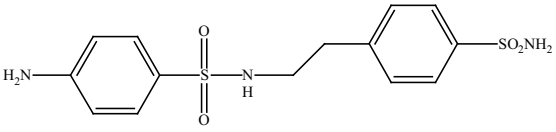
14



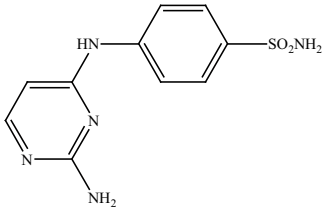
15



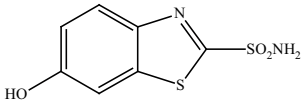
16



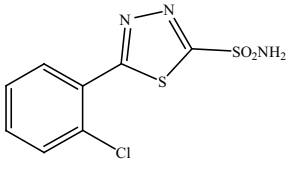
17



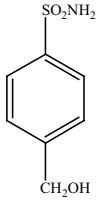
18



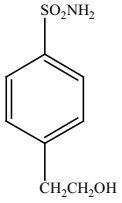
19



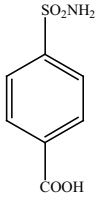
20



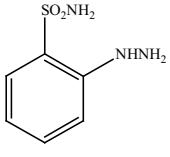
21



22

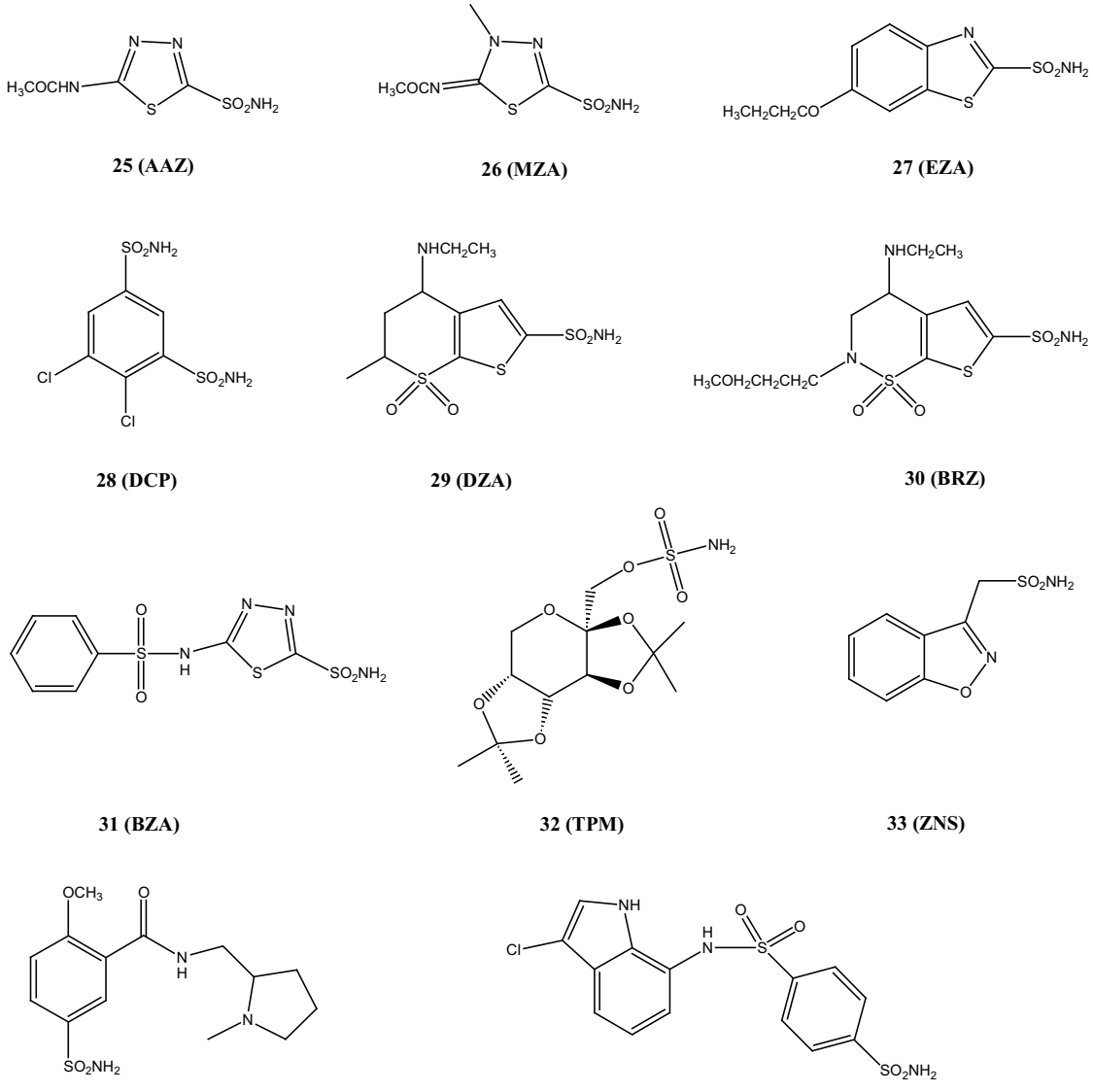


23



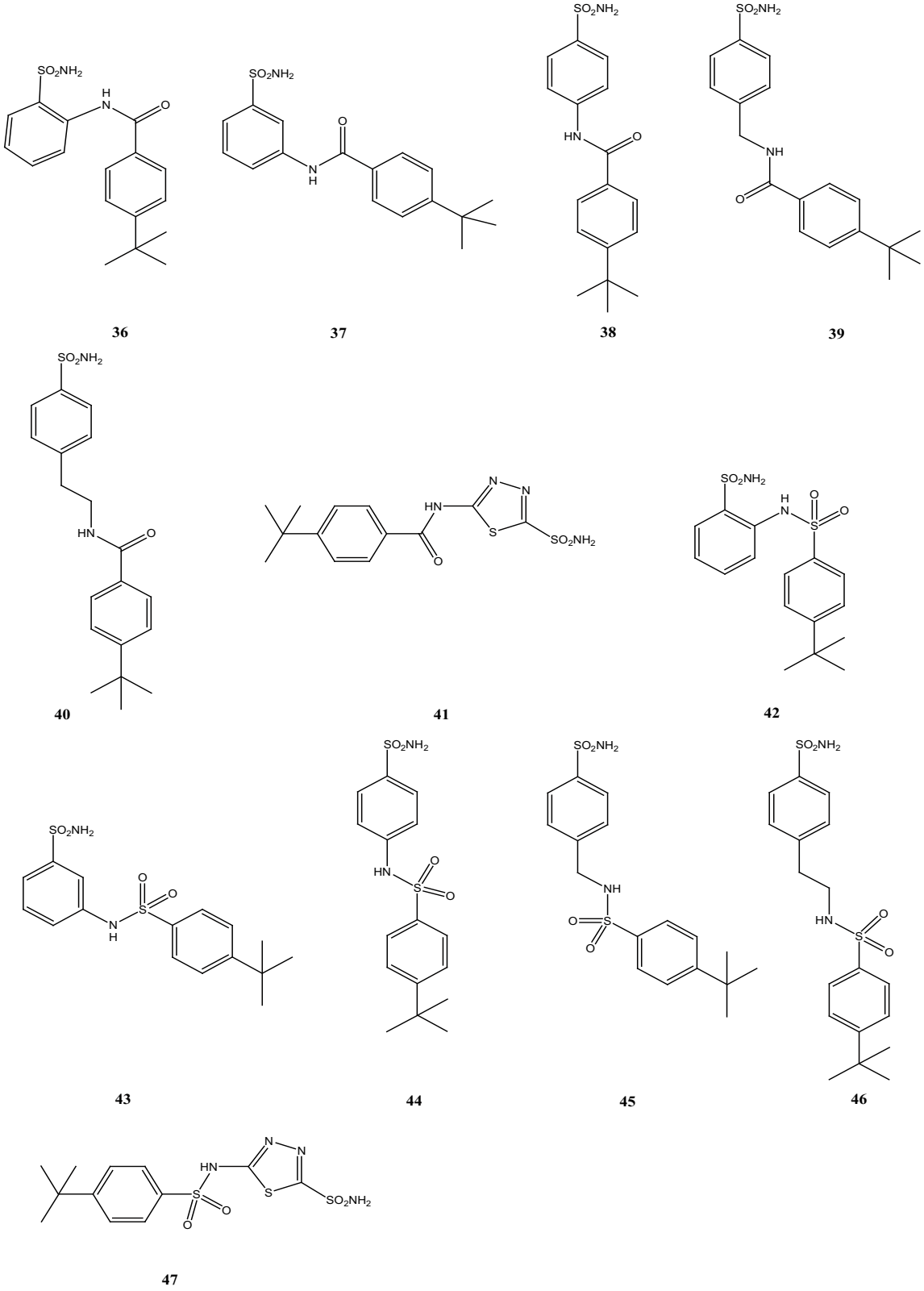
24

Scheme 1. Simple aromatic and heteroaromatic sulfonamides (**1**-**24**).



Scheme 2. Systemically acting carbonic anhydrase inhibitors (**25**-**28**), antiglaucoma (**29**-**31**) and antiepileptic (**32**-**35**) agents. Acetazolamide, AAZ (**25**), methazolamide, MZA (**26**), ethoxzolamide, EZA (**27**), and dichlorophenamide, DCP (**28**), are the classical, systemically acting CA inhibitors. Dorzolamide, DZA (**29**), and brinzolamide, BRZ (**30**) are topically acting antiglaucoma agents, benzolamide, BZA (**31**) is an orphan drug belonging to this class of pharmacological agents, whereas topiramate, TPM (**32**), and zonisamide ZNS (**33**) are widely used antiepileptic drugs. Sulpiride, SLP (**34**), and indisulam IND (**35**) belong to this class of pharmacological agents.

compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This was done to ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software for computing the parameters corresponding to 0D-, 1D-, and 2D-descriptor classes. The descriptor classes considered in the study along with their definitions and scope in addressing the structural features have been presented in Table 3. As the total number of descriptors involved in this study is very large, only the names of descriptor classes and the actual descriptor involved in the models have been listed. The combinatorial protocol-multiple linear regression (CP-MLR) computational procedure of model development is briefly described below.



Scheme 3. Newly synthesized compounds having 4-tert-butyl-phenylcarboxamido/ sulfonamide tails attached to benzenesulfonamide/ 1,3,4-thiadiazole-2-sulfonamide scaffolds.

Table 1. Observed and modeled carbonic anhydrase inhibition activities of training set compounds.

S. No[a].	pK_i hCA I			pK_i hCA II			pK_i hpCA		
	Obsd[b].	Calc. Eq. 2	Pred. LOO	Obsd[b].	Calc. Eq. 3	Pred. LOO	Obsd[b].	Calc. Eq. 5	Pred. LOO
2	4.602	4.805	4.833	6.620	6.808	6.819	6.343	6.484	6.513
3	4.553	4.435	4.418	6.523	6.846	6.867	6.500	6.085	6.008
4	4.105	4.636	4.697	6.495	6.860	6.897	6.347	6.480	6.509
6	4.678	4.778	4.790	6.796	6.853	6.856	5.939	5.985	5.997
8	5.009	5.595	5.625	6.959	7.068	7.083	6.423	6.333	6.311
9	5.187	5.580	5.597	7.398	7.201	7.149	6.345	6.320	6.316
10	5.222	5.702	5.721	7.155	7.294	7.352	6.292	6.337	6.343
11	5.237	5.690	5.731	7.201	7.047	7.037	6.385	6.229	6.164
12	5.076	5.519	5.542	7.125	7.108	7.107	7.310	7.154	7.069
13	5.066	4.916	4.901	7.222	7.469	7.506	6.491	6.734	6.766
14	5.032	5.660	5.791	7.721	7.579	7.566	6.260	6.534	6.554
15	8.222	7.822	7.727	8.699	8.552	8.487	6.572	–[c]	–[c]
16	6.785	5.754	5.664	7.337	7.421	7.434	6.883	6.731	6.699
17	6.733	5.898	5.819	7.301	7.526	7.572	6.943	7.019	7.034
18	6.963	7.410	7.500	7.481	7.397	7.384	7.076	6.730	6.690
19	7.022	7.293	7.350	7.523	7.715	7.739	6.684	6.595	6.565
20	6.161	5.972	5.841	7.921	7.767	7.751	6.979	6.751	6.724
21	7.260	–[c]	–[c]	7.097	6.775	6.749	6.057	6.334	6.361
22	4.678	5.602	5.665	6.903	6.669	6.649	5.945	6.146	6.179
24	4.620	4.435	4.408	6.903	6.617	6.574	6.267	5.935	5.840
25	6.602	6.721	6.730	7.921	7.818	7.811	7.678	7.467	7.426
26	7.301	6.799	6.642	7.854	7.879	7.882	6.648	6.580	6.571
27	7.602	6.728	6.649	8.097	7.655	7.611	6.714	6.162	6.082
29	4.301	4.294	4.290	8.046	8.112	8.119	5.361	6.101	6.267
30	4.347	4.543	4.664	8.523	8.152	8.104	6.678	6.419	6.351
32	6.602	7.061	7.752	8.000	8.184	8.269	6.764	6.763	6.762
34	5.921	6.088	6.102	7.398	7.735	7.756	6.690	6.852	6.863
35	7.509	7.537	7.542	7.824	8.046	8.067	6.384	6.765	6.925
36	4.910	4.777	4.769	6.618	6.632	6.634	6.268	6.816	6.890
37	4.969	4.777	4.765	6.678	6.689	6.690	6.500	6.760	6.783
39	4.877	4.419	4.378	6.896	6.749	6.739	7.208	6.986	6.961
40	4.905	4.590	4.567	6.910	6.798	6.791	7.292	7.141	7.118
41	6.267	5.831	5.778	7.745	7.576	7.561	7.886	7.799	7.775
42	4.833	4.836	4.837	6.451	6.742	6.766	6.194	–[c]	–[c]
44	4.886	4.836	4.831	6.924	6.870	6.867	7.222	7.234	7.235
46	4.919	4.640	4.611	7.027	7.014	7.014	7.569	7.403	7.377
47	6.471	6.194	6.165	7.824	7.894	7.901	7.921	8.089	8.163

[a]Corresponds to the structures represented in Schemes I, II and III; [b]The binding constants K_i were taken from Ref. [27] and are given here as pK_i on molar basis; [c]The 'outlier(s)' of present study.

2.3. Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies [30-34]. The procedure employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure-activity models, each having unique combination of descriptors from the generated data set of the compounds under study. The 'filters' set in CP-MLR are intended at (i) having inter-parameter correlation to a predefined cutoff value (filter-1; default acceptable value ≤ 0.3; (ii) optimize the variable entry to a model through t-value of regression

Table 2. Predicted residual activity of test set compounds and corresponding predictive r^2

S. No[a].	pK_i hCA I			pK_i hCA II			pK_i hpCA		
	Obsd[b].	Calc. Eq. 2	Res[c].	Obsd[b].	Calc. Eq. 3	Res[c].	Obsd[b].	Calc. Eq. 5	Res[c].
1	4.343	4.805	–0.462	6.530	6.410	0.120	6.371	6.259	0.112
5	4.602	4.444	–0.158	6.770	6.851	–0.081	6.059	6.117	–0.058
7	5.081	5.868	–0.787	7.222	6.833	0.389	5.910	6.311	–0.401
23	4.638	6.500	–1.862	6.876	6.670	0.206	5.978	5.982	–0.004
28	5.921	6.140	–0.219	7.420	7.165	0.255	6.423	6.695	–0.272
31	7.824	8.043	–0.219	8.046	7.961	0.085	6.502	7.201	–0.699
33	7.252	7.183	0.069	7.456	7.676	–0.220	6.636	6.639	–0.003
38	4.846	4.777	0.069	6.876	6.716	0.160	7.102	6.985	0.117
43	5.017	4.836	0.181	6.693	6.844	–0.151	6.498	7.024	–0.526
45	4.915	4.480	0.435	6.983	6.936	0.047	7.509	7.204	0.305
	Test set r^2		0.641			0.841			0.570

[a,b]*See foot note under Table 1.* [c]*The difference between observed and calculated pKi value. The number of test set compounds is 10.*

Table 3. Descriptor classes used for the analysis of derivatives of sulfonamide and sulfamate as the inhibitors of carbonic anhydrases, hCA I, hCA II and hpCA.

Descriptor class (acronyms)[a]	Definition and scope
Constitutional **(CONST)**	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Topological **(TOPO)**	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts **(MWC)**	2D-descriptors representing self-returning walks counts of different lengths
Modified Burden eigenvalues **(BCUT)**	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms
Galvez topological charge indices **(GALVEZ)**	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
2D-autocorrelations **(2D-AUTO)**	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)
Functional groups **(FUNC)**	Molecular descriptors based on the counting of the chemical functional groups
Atom centered fragments **(ACF)**	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
Empirical **(EMP)**	1D-descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
Properties **(PROP)**	1D-descriptors representing molecular properties of a molecule

[a]*Ref. [28].*

coefficients (filter-2; default acceptable value ≥ 2.0; (iii) comparability of models (regression equations) with different number of descriptor in terms of square root of adjusted multiple correlation coefficient (filter-3; r-bar, default acceptable value ≥ 0.71), and (iv) addressing the external consistency of the model with leave-one-out (LOO) cross-validation as default option (filter-4; cross-validated Q^2 criteria, default acceptable limits are $0.3 \leq q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to unique solutions.

Further, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test [35,36] by repeated randomization of the biological response. The datasets with randomized response vector have been reassessed by MRA. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This

has been used as a measure to express the percent chance correlation of the model under scrutiny. Thus, the CP-MLR protocol has been applied with default filter thresholds to identify all possible models that could emerge from the descriptors of compounds.

For each model, derived in n data points, a number of statistical parameters were obtained to access its overall statistical significance. These are: the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F), the cross-validated indices, Q^2_{LOO} and Q^2_{L5O} respectively from leave-one-out and leave-five-out procedures. In leave-five-out procedure a group of five compounds is randomly kept outside the analysis each time in such a way that all compounds, for once, become the part of the predictive groups. The robustness of the models was evaluated by Q^2 index. The predictive power of the models was ascertained by test set r^2. A value greater than 0.5 of this index hints at a reasonable sound model.

A number of additional statistical parameters such as the Akaike's information criterion, AIC [37,38], the Kubinyi function, FIT [39,40], and the Friedman's lack of fit, LOF [41] have also been derived to evaluate the best model. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value (Fisher ratio) was proved to be a useful parameter for assessing the quality of the models. The main disadvantage of the F-value is its sensitivity to changes in k (the number of variables in the equation that describe the model), if k is small, and its lower sensitivity if k is large. The FIT criterion has a low sensitivity toward changes in k-values, as long as they are small numbers, and a substantially increasing sensitivity for large k-values. The model that produces the minimum value of AIC and the highest value of FIT is considered potentially the most useful and the best. The LOF takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large numbers of parameters. A minimum LOF value infers that the derived model is statistically sound.

3. Results and discussion

A total number of 510 descriptors from 0D-, 1D-, and 2D-modules of DRAGON software were calculated and subjected to CP-MLR with default filters implemented therein. The results of analysis for inhibition activities of isozymes hCA I and hCA II and the bacterial enzyme hpCA are discussed in the following sections.

3.1. Inhibition activity of hCA I

For inhibition activity of hCA I, the analysis revealed 8 models involving two descriptors with highest r-bar value of 0.733. This r-bar value was then retained as a threshold limit for filter-3 in CP-MLR and again used it to derive higher models. The method yielded 346 models in three descriptors with highest r-bar value of 0.830. Similar steps were followed to derive the models in four descriptors, however, the highest significant model emerged is shown in Eq. 1. The list of participating descriptor, their average regression coefficient and total incidences are included in Table 4.

$$pK_i(\text{hCA I}) = -4.253 + 3.256(0.387)IC1 - 5.384(1.153)MATS2v - 8.318(1.413)MATS1p - 1.150(0.347)nRSR$$

$$n = 37,\ r = 0.879,\ s = 0.568,\ F = 27.282,\ Q^2_{LOO} = 0.708,\ Q^2_{L5O} = 0.697,\ AIC = 0.424,\ FIT = 2.059,\ LOF = 0.455 \quad (1)$$

Where IC1 is the information content index of the neighborhood symmetry of order-1, MATS2v is the Moran autocorrelations of lag 2 weighted by atomic van der Waals volume, MATS1p is the Moran autocorrelations of lag 1 weighted by atomic polarizability and nRSR is the number of sulfurs. The derived F-value for above Equation remained significant at 99% level and the Q^2 index accounted for a robust model but the r^2-value has explained only for 77% of variance in observed activity values. To improve the significance of above equation the compound having highest residual activity was considered as the outlier.

An outlier to a QSAR is identified normally by having a large standard residual activity and can indicate the limits of applicability of QSAR models [42]. There are many reasons for their occurrence in QSAR studies; for example, chemicals might be acting by a mechanism different from that of the majority of the data points. It is also likely that outlier might be a result of a random experimental error that could be significant when analyzing a large data set. Although it is acceptable to remove a small number of outliers from the QSAR [43] but it is not acceptable to remove the outliers repeatedly from a QSAR analysis simply for improving a correlation. In present work, the lone compound **21** have shown highest residual and was considered as an outlier. This compound is ignored in the training set to yield a more significant correlation Eq. 2.

Table 4. Descriptor classes and identified descriptors in modeling the inhibition activities of human carbonic anhydrase, hCA I.

Des. Class	Identified descriptors[a] and their average regression coefficient (incidence)[b]
CONST	nCIC, 0.978(2); nO, 0.298(3)
TOPO	HNar, 5.016(7); TI2, 0.246(2); Jhetv, –1.187(2); Jhete, –0.816(1); Jhetp, –0.759(1); X0A, –26.147(1); DECC, 0.893(2); IC1, 3.116(26); SEigv, –0.185(1); T(O..O), 0.019(2); T(S..S), 0.077(1)
GALVEZ	GGI9, 2.865(2); JGI9, 62.248(2)
2D-AUTO	MATS2v, –5.181(8); MATS1p, –6.919(13); MATS2p, –6.216(7); GATS8v, 1.385(2); GATS2p, 3.608(1)
FUNC	nCq, –0.770(4); nNHR, –0.817(1); nRSR, –1.529(10)
ACF	C-029, –1.459(3)

[a]The descriptors are identified from the four parameter models, emerged from CP-MLR protocol with filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.830, and filter-4 as 0.3 ≤ Q^2 ≤1.0 with a training set of 32 compounds. CONST: nCIC, number of rings; nO, number of oxygen atoms; TOPO: HNar, the Narumi harmonic topological indices; TI2, second Mohar index; Jhetv, Jhete and Jhetp, Balaban-type index from van der Waals, electronegativity and polarizability weighted distance matrix respectively; X0A, average connectivity index chi-0, DECC, eccentric; IC1, information content index of 1-order neighborhood symmetry; SEigv, eigenvalue sum from van der waals weighted distance matrix; T(O..O) and T(S..S), sum of topological distances between O..O and S..S respectively; GALVEZ: GGI9 and JGI9, topological and mean topological charge indices of order 9; 2D-AUTO: MATSkw and GATSkw are the Moran and Geary auto-correlations of lag k, weighted by some physical property w such as atomic van der Waals volume (v), atomic polarizability (p), atomic mass (m) etc.; FUNC: nCq, number of total quarternary C(sp3); nNHR, number of secondary aliphatic amines; nRSR, number of sulfurs; ACF: C-029, R-CX-X. [b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

$$pK_i(\text{hCA I}) = -4.504 + 3.312(0.324)IC1 - 5.119(0.968)\ MATS2v - 8.691(1.188)MATS1p - 1.116(0.291)nRSR$$
$$n = 36,\ r = 0.915,\ s = 0.476,\ F = 39.640,\ Q^2_{LOO} = 0.771,\ Q^2_{L5O} = 0.732,\ AIC = 0.300, FIT = 3.049,\ LOF = 0.323 \quad (2)$$

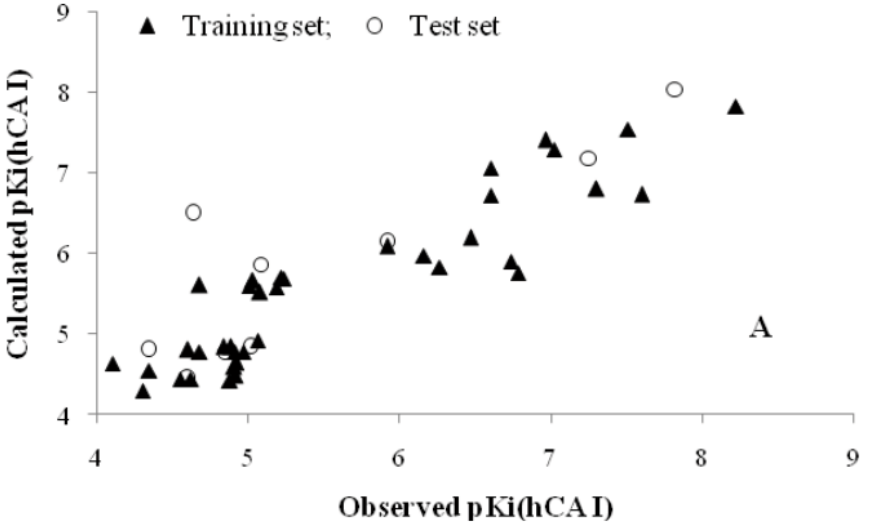
The statistical parameters of Eq. 2 have now improved over to that of Eq. 1. The r^2-value has explained for 84% of variance in observed activity values and Q^2 index has accounted comparatively for a better robust model. The decreased values of parameters AIC and LOF and increased value of FIT have further shown the superiority of this model over to that of the model in Eq. 1. Eq. 2 was also subjected to randomization process, where 100 simulations were carried out but none of the identified models has shown any chance correlation. The pK_i values of training set compounds calculated using Eq. 2 and predicted from LOO procedure have been included in Table 1. The model (2) has also validated with an external test set of ten compounds listed in Table 2. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 value and the same is reported in Table 2. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Fig. 1A.

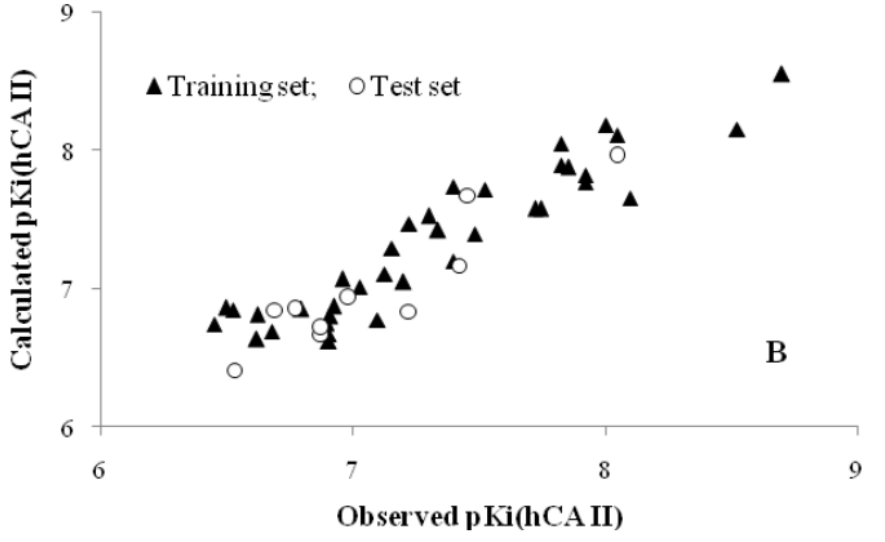
From Eq. 2, it appeared that the descriptor IC1 has contributed positively while the descriptors MATS2v, MATS1p and nRSR have added negatively to the inhibition activity. This implies that a higher value of IC1 (a measure of structural complexity per vertex) would augment the activity of a compound. The role of atomic van der Waals volumes and polarizability in respective lags may be seen in Moran autocorrelations. A lower (or negative) value of the Moran autocorrelations, MATS2v and MATS1p, is beneficial in improving the activity. Similarly a lower number of sulfur atoms in a molecular structure (nRSR) would enhance the activity of a compound.

3.2. Inhibition activity of hCA II

The inhibition data of 37 training set compounds pertaining to the isoform hCA II were subjected to CP-MLR analysis, setting the filter-3 value at 0.71, using 510 initial DRAGON descriptors as independent variables. The analysis resulted into 806 models in two descriptors but the highest significant model yielded the r-bar value of 0.886. This r-bar value was then considered as the threshold value for filter-3 in CP-MLR and the models involving three descriptors were derived. A total number of 60 descriptors belonging to different classes have participated in 106 models with their r-bar values ≥ the threshold limit set for filter-3. The complete list of participating descriptors, their average regression coefficients and total incidences are included in Table 5. The highest significant model, involving one descriptor each from the CONST, TOPO and 2DAUTO classes, that was obtained is shown in Eq. 3

re 1. The variations of observed versus calculated pK$_i$ values. Plots A, B and C correspond to Eqs. 2, 3 and 5, respectively.





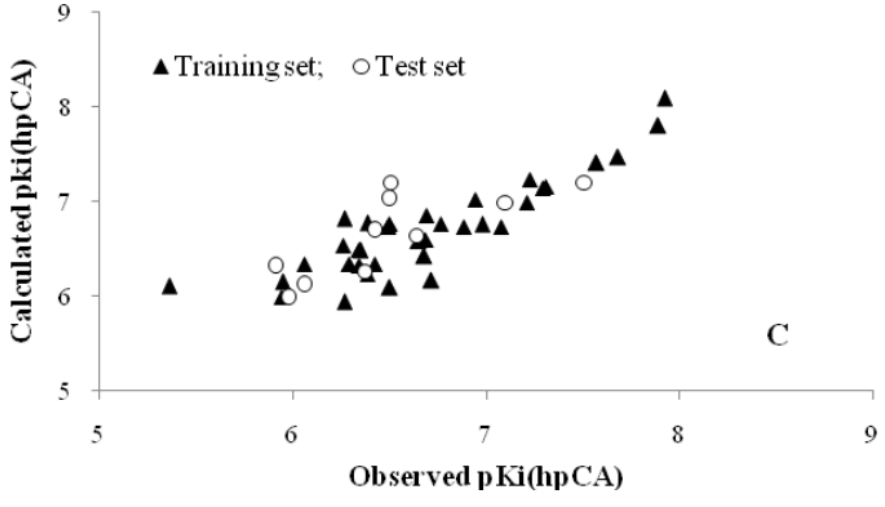


Table 5. Descriptor classes and identified descriptors in modeling the inhibition activities of human carbonic anhydrase, hCA II.

Des. Class	Identified descriptors[a] and their average regression coefficient (incidence)[b]
CONST	AMW, 0.061(4); Mv, 3.518(4); Mp, 3.268(6); nAB, –0.056(1); nR05, 0.740(28)
TOPO	AAC, 0.480(1); HNar, 1.528(3); GNar, 2.211(3); X0A, –8.901(3); X1Av, 5.388(1); X5sol, 0.209(2); BLI, 1.804(1); IVDE, –1.970(3); IC1, 1.394(6); SIC1, 1.342(4); CIC1, –0.171(3); BIC1, 1.567(5); IC2, 1.129(1); SIC2, 3.850(10); CIC2, –0.603(7); BIC2, 4.811(13); SIC3, 3.238(1); BIC3, 3.972(2); SIC5, 2.248(1); CIC5, –0.335(1); BIC5, 4.530(2); SEigZ, 0.568(4); SEigm, 0.578(3); MPC06, 0.013(1); MPC08, 0.011(2); D/Dr05, 0.015(3); T(N..N), 0.020(4); T(N..O), 0.011(19); T(N..S), 0.023(40); T(O..S), 0.020(1)
MWC	SRW07, 0.006(21); SRW09, 0.001(13)
BCUT	BEHm2, 0.519(2)
GALVEZ	JGI10, –25.742(1)
2D-AUTO	ATS2m, 0.021(1); MATS1m, 8.153(5); MATS2m, 20.324(4); MATS6m, –3.712(3); MATS1v, –2.541(2); MATS6v, –0.674(8); MATS1p, –2.800(1); MATS6p, –0.538(4); GATS6v, 0.619(6); GATS2e, 1.458(6); GATS6p, 0.269(3)
FUNC	nROR, 0.160(3)
ACF	C-006, 0.143(3); C-026, –0.211(3); C-032, 0.512(1); H-053, 0.063(3); O-059, 0.160(3)
EMP	Ui, –0.370(1)
PROP	MLOGP, –0.225(1); PSA, 0.016(26)

[a]The descriptors are identified from the three parameter models, emerged from CP-MLR protocol with filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.886, and filter-4 as 0.3 ≤ Q^2 ≤1.0 with a total number of 32 compounds. CONST: AMW, average molecular weight; Mv and Mp, mean atomic van der Waals volume and polarizability (scaled on carbon atom); nAB and nR05, number of aromatic bonds 5-membered rings respectively; TOPO: AAC, mean information index on atomic composition; HNar and GNar, the Narumi harmonic and geometric topological indices; X0A, average connectivity index chi-0; X1Av, average valence connectivity index chi-1; X5sol, salvation connectivity index chi-5; BLI, Kier benzene-likeliness index; IVDE, mean information content vertex degree equality; IC1 and IC2, information content indices of 1- and 2-order neighborhood symmetry; SICk, BICk and CICk, structural, bond and complementary information contents of neighborhood symmetry of k-order; SEigZ and SEigm, the eigenvalue sum from Z and mass weighted distance matrix; MPC06 and MPC08; molecular path count of 6-and 8-order; D/Dr05, distance/detour ring index of order 5; T(A..B), sum of topological distances between A..B; MWC: SRW07 and SRW09, self-returning walk count of 7- and 9-order; BCUT: BEHm2, highest eigen value n.2 of Burden matrix weighted by atomic masses; GALVEZ: JGI10, mean topological charge index of 10-order; 2D-AUTO: ATSkw, MATSkw and GATSkw, the Broto-Moreau, Moran and Geary auto-correlations of lag k, weighted by some physical property w such as atomic van der Waals volume (v), atomic polarizability (p), atomic mass (m) etc.; FUNC: nROR number of ethers (aliphatic); C-006, CH2RX; C-026, R-CX-R; C-032, X-CX-X; H-053, H attached to C0(sp3) with 2X attached to next C; O-059, Al-O-Al; EMP: Ui, unsaturation index; PROP: MLOGP, Moriguchi octanol-water partition coefficient (logP); PSA, fragment-based polar surface area. [b]See footnote under Table 4.

pK_i(hCA II) = 6.114 + 0.791(0.065)nR05 + 0.022(0.003) T(N..S) +0.602(0.141)GATS6v

n = 37, r = 0.933, s = 0.216, F = 73.450, Q^2_{LOO} = 0.835, Q^2_{L5O} = 0.845, AIC = 0.058, FIT = 4.790, LOF = 0.059 (3)

Where the descriptors nR05 and T(N..S), account for the number of 5-membered rings and the topological distance between N and S respectively in a compound. The descriptor, GATS6v is the Geary autocorrelation of lag 6, weighted by atomic van der Waals volume. The statistical parameter r has accounted for 87% of variance in observed activities while F-value remained significant at 99% level. These are in support of highly significant results. Similarly, the lower values of AIC and LOF and the higher value of FIT are in favor of a potentially useful and the best model. In randomization process (100 simulations) the identified model has not shown any chance correlation. The sufficiently high value of Q^2 index have accounted for an internally robust model. The pK_i(hCA II) values, calculated using Eq. 3 and predicted using LOO procedure, were found in close agreement with the observed ones (Table 1). The predictions based on external test set are within the reasonable limits (Table 2). The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Fig. 1 (B). From Eq. 3, it appeared that a higher number of 5-membered rings and a bigger distance between N and S along with a higher value of van der Waals volume weighted parameter of lag 6 are helpful in improving the activity of a compound.

3.3. Inhibition activity of bacterial enzyme hpCA

The inhibition activity of bacterial enzyme hpCA seems to be highly structurally sensitive as no significant correlation was obtained when 37 training set compounds were subjected to CP-MLR analysis in two descriptors (filter-1 = 0.3 and filter-3 = 0.71). Therefore, in successive analysis a liberated value of 0.79 for filter-1 has been considered. The models were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. Finally 9 models, sharing 18 descriptors as relevant ones, in five parameters were obtained through CP-MLR. The participating descriptors along with their average regression coefficients and total incidences are listed in Table 6 and their physical meaning is provided as foot note under this Table. The highest significant model in five parameters is given below as Eq. 4

pK_i(hpCA) = 9.645 – 0.823(0.174)Jhetv – 21.256(4.973) PW5 + 0.248(0.096)SEigz + 0.800(0.214)MATS8v + +1.823(0.450)MATS4e
n = 37, r = 0.826, s = 0.340, F = 13.318, Q^2_{LOO} = 0.563, Q^2_{L5O} = 0.529, AIC = 0.161,FIT = 1.074, LOF = 0.182 (4)

Where Jhetv is the Balaban-type index from van der Waals volume weighted distance matrix, PW5 is the path/walk 5- Randic shape index and SEigZ is the eigenvalue sum from Z weighted distance matrix. The other participating descriptors, MATS8v and MATS4e are the Moran autocorrelation of lag 8 weighted by atomic van der Waals volumes and Moran autocorrelation of lag 4 weighted by atomic Sanderson electronegativities. However, the derived F-value for above equation remained significant at 99% level but the Q^2 index accounted for a slightly inferior model and the r^2-value has explained only for 68% of variance in observed activity values. To improve the significance of above equation the compound having very high residual activities were considered as the outliers. For the inhibition activity of bacterial enzyme hpCA, the compounds **15** and **42** have shown high residuals and were considered as outliers. The model obtained on removal of these compounds from the training set is shown in Eq. 5.

pK_i(hpCA) = 9.683 – 0.808(0.150)Jhetv – 24.095(4.386) PW5 + 0.339(0.088)SEigz + 0.720(0.186)MATS8v + +1.853(0.390)MATS4e
n = 35, r = 0.880, s = 0.293, F = 19.987, Q^2_{LOO} = 0.677, Q^2_{L5O} = 0.683, AIC = 0.121,FIT = 1.666, LOF = 0.139 (5)

The statistical parameters of Eq. 5 have now improved over to that of Eq. 4. The r^2-value has explained for 77% of variance in observed activity values and Q^2 index has accounted comparatively for a better robust model. The decreased values of parameters AIC and LOF and increased value of FIT have further shown the superiority of this model over to that of the model in Eq. 4. The activity data have been randomized to generate a new model without altering the original

Table 6. Descriptor classes and identified descriptors in modeling the inhibition activities of human carbonic anhydrase, hpCA.

Des. Class	Identified descriptors[a] and their average regression coefficient (incidence)[b]
CONST	RBF, 6.939(2); nN, 0.196(3)
TOPO	Xt, –33.812(1); J, –1.188(2); Jhetv, –0.860(4); PW4, –22.751(1); PW5, –17.503(4); SEigZ, 0.248(1); T(N..O), .011(2); T(O..O), 0.012(1)
GALVEZ	JGI2, –6.175(1); JGI4, –13.934(2)
2D-AUTO	MATS2m, 27.024(4); MATS8v, 0.834(6); MATS4e, 1.868(7); GATS8m, –105.226(1)
FUNC	nNHR, –0.566(1)
ACF	C-001, 0.209(2)

[a]The descriptors are identified from the one and two parameter models, emerged from CP-MLR protocol with filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.71, and filter-4 as 0.3 ≤ Q^2 ≤1.0 with 24, 11 and 12 compounds of Table 1 for inhibition of hpCA; CONST: RBF, rotatable bond fraction; nN, number of nitrogen atoms; TOPO: Xt, total structure connectivity index; J, Balaban J index; Jhetv, Balaban-type index from van der Waals volume weighted distance matrix; PW4 and PW5, path/walk 4- and 5- Randic shape indices; SEigZ, eigenvalue sum from Z weighted distance matrix; T(A..B), sum of topological distances between A..B; GALVEZ: JGI2 and JGI4, mean topological charge index of 2- and 4-order; 2D-AUTO: MATSkw and GATSkw, the Moran and Geary auto-correlations of lag k, weighted by some physical property w such as atomic van der Waals volume (v), atomic polarizability (p), atomic mass (m) etc.; FUNC: nNHR, number of secondary aliphatic amines; ACF: C-001, CH3R/CH4. [b]See footnote under Table 4.

descriptors matrix used for the derivation of Eq. 5. In 100 such iterations, none of the emerged model could yield a correlation better than the correlation shown by Eq. 5. This has justified that model in Eq. 5 is not being an outcome of any chance correlation. The pK_i values of training set compounds calculated using Eq. 5 and predicted from LOO procedure have been included in Table 1. Further, the model (4) is validated with an external test set of ten compounds listed in Table 2. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 value and the same is reported in Table 2. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Fig. 1C. From Eq. 5, it appeared that a compound will be more active provided its Balaban-type index from van der Waals volume weighted distance matrix, Jhetv and path/walk 5-Randic shape index, PW5 both have lower values. On the other hand higher values of the eigenvalue sum from Z weighted distance matrix, SEigZ, the Moran autocorrelation of lag 8 weighted by atomic van der Waals volumes, MATS8v and the Moran autocorrelation of lag 4 weighted by atomic Sanderson electronegativities, MATS4e is favorable in improving its activity pertaining to hpCA.

4. Conclusions

The inhibition activities of sulfonamide and sulfamate derivatives related to human carbonic anhydrases (hCA I and hCA II) and α-carbonic anhydrase from *Helicobacter pylori* (hpCA) have been quantitatively expressed in terms of 0D-, 1D- and 2D-descriptors of DRAGON software. For the inhibition activity of hCA I, a higher value of information content index of the neighborhood symmetry of order-1 (descriptor, IC1) and a lower (or negative) value of the Moran autocorrelations, MATS2v and MATS1p, along with a lower number of sulfur atoms in a molecular structure (nRSR) is beneficial in improving the activity. A higher number of 5-membered rings (nR05), a bigger distance between nitrogen and sulfur T(N..S), and a higher value of van der Waals volume weighted Geary autocorrelation of lag 6 (GATS6v) are helpful in improving the inhibition activity of a compound pertaining to hCA II. For the inhibition of bacterial enzyme, hpCA, a lower value of both the descriptors Jhetv and PW5 is beneficial to activity. On the other hand the sign of regression coefficients of descriptors SEigZ, MATS8v and MATS4e advocates that a higher value of these descriptors would be favorable in improving its hpCA inhibition activity. The derived significant models in such descriptors may further be used to synthesize new potential compounds and to decipher the mode of their actions at molecular level.

Acknowledgements

We are thankful to our Institutions for providing necessary facilities to complete this work. CDRI communication no. 7770.

References

[1] C.T. Supuran, A. Scozzafava, J. Conway, Carbonic Anhydrase-Its Inhibitors and Activators (CRC Press: Boca Raton, FL, 2004) 1
[2] W.S. Sly, P.Y. Hu, Annu. Rev. Biochem. 64, 375 (1995)
[3] S. Pastorekova, S. Parkkila, J. Pastorek, C.T. Supuran, J. Enzy. Inhibn. Med. Chem. 19, 199 (2004)
[4] C.T. Supuran, A. Scozzafava, A. Casini, Med. Res. Rev. 23, 146 (2003)
[5] A. Scozzafava, A. Mastrolorenzo, C.T. Supuran, Expert Opin. Ther. Pat. 14, 667 (2004)
[6] J. Lehtonen et al., J. Biol. Chem. 279, 2719 (2004)
[7] K. Fujikawa-Adachi, I. Nishimori, T. Taguchi, S. Onishi, J. Biol. Chem. 274, 21228 (1999)
[8] K. Fujikawa-Adachi, I. Nishimori, T. Taguchi, S. Onishi, Genomics 61, 74 (1999)
[9] (a) I. Nishimori et al., J. Med. Chem. 48, 7860 (2005); (b) D. Vullo et al., J. Med. Chem. 47, 1272 (2004); (c) C.T. Supuran, Expert Opin. Ther. Pat. 13, 1545 (2003)
[10] (a) A. Weber et al., J. Med. Chem. 47, 550 (2004); (b) S. Pastorekova et al., Bioorg. Med. Chem. Lett. 14, 869 (2004)
[11] (a) D. Vullo et al., Bioorg. Med. Chem. Lett. 15, 971 (2005); (b) G. De Simone et al., Bioorg. Med. Chem. Lett. 15, 2315 (2005)
[12] D. Vullo et al., Bioorg. Med. Chem. Lett. 15, 963 (2005)
[13] (a) E. Svastova et al., FEBS Lett. 577, 439 (2004); (b) A. Cecchi et al., J. Med. Chem. 48, 4834 (2005); (c) V. Menchise et al., J. Med. Chem. 48, 5721 (2005)
[14] J. Krungkrai et al., Bioorg. Med. Chem. 13, 483 (2005)

[15] (a) A. Suarez Covarrubias et al., J. Biol. Chem. 280, 18782 (2005); (b) A. Suarez Covarrubias, T. Bergfors. T.A. Jones, M. Hogbom, J. Biol. Chem. 281, 4993 (2006)
[16] (a) T. Klengel et al., Curr. Biol. 15, 2021 (2005); (b) E.G. Mogensen et al., Eukaryot. Cell 5, 103 (2006)
[17] Y.S. Bahn, G.M. Cox, J.R. Perfect, J. Heitman, Curr. Biol. 15, 2013 (2005)
[18] (a) B.J. Marshall, J.R. Warren, Lancet 16, 1311 (1984); (b) B.J. Marshall, Am. J. Gastroenterol. 89, S116 (1994)
[19] (a) J. Personnet et al., N. Engl. J. Med. 325, 1131 (1991); (b) S.F. Moss, M.J. Blaser, Nat. Clin. Pract. Oncol. 2, 90 (2005); (c) J. Kountouras, C. Zavos, D. Chatzopoulos, Hepatogastroenterology 52, 1305 (2005); (d) B.J Marshall, H.M. Windsor, Med. Clin. North Am. 89, 313 (2005)
[20] (a) R. O'Mahony, D. Vaira, J. Holton, C. Basset, Sci. Prog. 87, 269 (2004); (b) F. Megraud, B.J. Marshall, Gastroenterol. Clin. North Am. 29, 759 (2000)
[21] J.P. Gisbert, J.M. Pajares, Helicobacter. 10, 363 (2005)
[22] Y. Nakayama, D.Y. Grahm, Expert Rev. Anti-Infect. Ther. 2, 599 (2004)
[23] (a) F. Megraud, Drugs 64, 1893 (2004); (b) M.F. Loughlin, Expert Opin. Ther. Targets 7, 725 (2003)
[24] G. Sachs et al., Physiology (Bethesda) 20, 429 (2005)
[25] E.A. Marcus, A.P. Moshfegh, G. Sachs, D.R. Scott, J. Bacteriol. 187, 729 (2005)
[26] F.N. Stahler, L. Ganter, K. Lederer, M. Kist, S. Bereswill, FEMS Immunol. Med. Microbiol. 44, 183 (2005)
[27] I. Nishimori et al., J. Med. Chem. 49, 2117 (2006)
[28] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON software version 3.0-2003, Milano (Italy)
[29] ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation: Cambridge, USA
[30] Y.S. Prabhakar, QSAR Comb. Sci. 22, 583 (2003)
[31] Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, S.B. Katti, QSAR Comb. Sci. 23, 234 (2004)
[32] Y.S. Prabhakar, Internet Electron. J. Mol. Des. 3, 150 (2004)
[33] M.K. Gupta, R. Sagar, A.K. Shaw, Y.S. Prabhakar, Bioorg. Med. Chem. 13, 343 (2005)
[34] S. Sharma, Y.S. Prabhakar, P. Singh, B.K. Sharma, Euro. J. Med. Chem. 43, 2354 (2008)
[35] S.S. So, M. Karplus, J. Med. Chem. 40, 4347 (1997)
[36] Y.S. Prabhakar, R.K. Rawal, M.K. Gupta, V.R. Solomon, S.B. Katti, Comb. Chem. High Through. Screen. 8, 431 (2005)
[37] H. Akaike, In: B.N. Petrov, F. Csaki (Eds.), Second international symposium on information theory (Akademiai Kiado Budapest, 1973) 267
[38] H. Akaike, IEEE Trans Autom. Control AC-19, 716 (1974)
[39] H. Kubinyi, Quant. Struct.-Act. Relat. 13, 285 (1994)
[40] H. Kubinyi, Quant. Struct.-Act. Relat. 13, 393 (1994)
[41] J. Friedman, Technical Report No 102, Laboratory for computational statistics (Stanford University, Stanford, 1990)
[42] R.L. Lipnick, Sci. Total Environ. 109, 131 (1991)
[43] J. Devillers, R.L. Lipnick, In: K. Karcher, J. Devillers (Eds.), Practical applications of quantitative structure-activity relationships (QSAR) in environemntal chemistry and toxicology (Kluwer, Dordrecht, 1990) 129