Estimation of Human Carbonic Anhydrase II Inhibition Using Topological Indices and their Combination with Quantum-Theoretical Descriptors

Jyoti Singh[1], Shalini Singh[2], Rajika Mishra[3], Padmakar V. Khadikar[4,*], Claudiu T. Supuran[5], Brian W. Clare[6] and Meenakshi Lakhwani[7]

[1]QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa-486 003, India; 2Department of Chemistry, Bareilly College, Bareilly (U.P); [3]Department of Health Engineering and Environment Manegement , MANIT , Bhopal, India; [4,]Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd., 3, Khatipura, Indore 452 007, India; [5]Laboratio di Chimica Bioinorganica, Diparmento di Chimica, University of Florence, via della Lastruccia, 3, Rm-188,Polo Scientifico, 50019 SestoFlorentino,Fireze, Italy; [6]School of Biomedical and Chemical Sciences, University of Western, Australia ,35,Stirling Highway Crawltywa 6009, Australia; [7]Sanghvi Institute of Management & Sciencs, Indor (M.P.), India*

Abstract: Mathematical models were developed for the estimation of human carbonic anhydrase (CA) II inhibition. A large set of 95 CA inhibitors incorporating diverse aromatic rings were used for this purpose. The numerical descriptors used were distance- and connectivity- based indices, quantum -theoretical descriptors and Balaban and Balaban type descriptors of molecular structure. After descriptor generation, multiple linear regression analysis was performed to find superior models for estimation. The obtained results indicate that: (i) models based on topological indices are superior to those based on quantum -theoretical descriptors; (ii) combinations of topological and quantum-theoretical descriptors improves the quality of regression; (iii) in both cases involvement of Balaban and Balaban type indices is beneficial. The results are described critically based on variety of statistical parameters.

Key Words*:* Carbonic anhydrase, Balaban indices, QSAR, topological index, quantum-theoritical descriptor, human CA II.

1. INTRODUCTION

Many different isoforms of the zinc enzyme carbonic anhydrase (CA, EC 4.2.1.1) enzyme are found in the mammalian body, each having specific physiologic functions [1,2]. Diseases caused by problematic acid-base secretion chemistry in the body, particularly in the eye, have been linked to the dysfunctional activities of several types of carbonic anhydrases [1-3]. Conditions such as macular edema and open-angle glaucoma can be treated by employing drugs which reduce the rate of formation of aqueous humor, i.e., sulfonamide CA inhibitors. It was demonstrated that certain CA enzymes contribute to the creation of eye humor through the production of bicarbonate ions [1, 2]. Drugs inhibiting the activity of the CA isozymes that exist in the eye have been and are successful in relieving symptoms and treating such widespread ophthalmologic diseases.

Carbonic anhydrase II (CA II) [E.C. 4.2.1.1] is a ubiquitous and physiologically highly relevant isoform. It is a highly efficient catalyst for the reversible hydration of carbon dioxide through a two-step, zinc-hydroxide mechanism described by equations *A* and *B* below [2,4]:

$$Zn^{2+}\text{-}OH^- + CO_2 \overset{H_2O}{\leftrightarrow} Zn^{2+}\text{-}HCO_3^- \leftrightarrow Zn^{2+}\text{-}H_2O \qquad (A)$$

$$Zn^{2+}\text{-}H_2O \leftrightarrow Zn^{2+}\text{-}OH^- + H^+ \qquad (B)$$

CA II can also hydrate aldehydes and hydrolyze some esters. It is a well-characterized enzyme whose three-dimensional structure has been determined by X-ray crystal lography in the absence and presence of inhibitors [1]. In addition, structure-activity-relationships of various sulfonamide CA II inhibitors have been studied [9-14]. CA II inhibitors have found a wide range of application as diuretics, antiepileptics, as agents for the treatment of glaucoma and modulators of cancer chemotherapy [1-4]. The development of topical CA inhibitors for the treatment of glaucoma, dorzolamide and brinzolamide, has renewed the pharmacological interest for this enzyme [2].

The active site of human CA II (hCA II) contains a catalytically essential zinc ion in tetrahedral geometry. The metal ion is coordinated by three imidazolic nitrogen atoms belonging to His94, His96 and His119 and one oxygen atom from a water molecule / hydroxide ion [1]. At physiological *pH*, aromatic and heterocyclic unsubstituted sulfonamides ($R\text{-}SO_2NH_2$), which are known to inhibit CAs, have an ionized sulfonamido group (pK_a 6 ~ 10). Upon binding, the sulfonamido group displaces the water from the zinc coordination sphere. Substitution of the RSO_2NH_2 hydrogen substantially decreases the CA inhibitory activity [15,16] due to steric hindrance. The aromatic side chains of sulfonamide inhibitors interact with many amino acid residues in the binding site (e.g., Phe131, Leu141, Val143, Ala145) and stabilize the interaction [1]. Unsubstantiated amides ($R\text{-}CONH_2$) such as urethane, phenylcarbamate are a second, albeit much less potent, class of known CA-II inhibitors. It contrast to sulfonamides, the CA inhibitors anions such as SCN^-, ClO_4^-, I^- are also weak CA-II inhibitors with Ki (binding constant) values of 8 ~ 30 μM [17, 18], since they only coordinate zinc and lack other stabilizing interactions. In

*Address correspondence to this author at the Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd., 3, Khatipura, Indore 452 007, India; Tel: +91-731-2531906; E-mail: pvkhadikar@rediffmail.com

1573-4064/08 $55.00+.00

summary, a negativity charged Zn ion coordinator and suitable hydrophobic moieties are thought to be the important structural requirement of CA II inhibitors.

Quantitative structure-activity-relationship (QSAR) methodology can be helpful in screening a large library of possible drug candidates for selectivity and potency. Data base mining methods which also includes QSAR type [18] approaches [9-14], play an important role in contemporary computer-assisted drug design (CADD) and lead compound discovery. Classical QSAR analyses are used to predict compound activities, define pharmacophore models, guide lead optimization and deduce mechanistic details of protein-ligand / inhibitor interactions [18]. In conventional 2D QSAR analysis, biological activities are quantitatively expressed as a function of the presence or absence of specific structural features (indicator variable or Free-Wilson approach), values of physico-chemical properties (Hansch analysis), or a combination of both. Molecular structure is encoded through the generation of descriptors, which numerical values are corresponding to topological, geometric, or electronic features such descriptors having some specific numerical values are called topological indices. The aim of the present study is to employ distance- and connectivity-based biological indices, quantum-theoretical descriptors, Balaban and Balaban type indices and their combinations to identify CA-II inhibitors and derive predictive models such that the resulting models can be applied to rapidly screen large data bases [19-23].

In view of the above and in continuation to our earlier work [24-31] we have chosen ring system **1-25** to which tails **A-D** were attached at the amino / hydroxyl functionality by means of amide / ester bonds, as given in (Table **1**). The compounds **1-24, 25-48, 49-72,** and **73-95** contain tail **A-D** respectively. The log K_i (nm) values for this set **95** CA II inhibitors along with the assumed indicator parameters are summarized in Table **2** and they were initially reported in Ref. [32]. The list of the variety of descriptors used in the present study is given in Table **3**. The calculated values of distance- and connectivity- based indices along with the Balaban and Balaban type indices are shown in (Table **4**), while the quantum- theoretical descriptors are given in (Table **5**). All these descriptors are calculated using DRAGON [40] and MOPAC [41] soft wares. The structure optimization was made using Hyperchem [42] software. The statistical calculations were done with MARTHA [43], ORIGIN [44] and NCSS [45] soft wares. All the variables were entered in the beginning of the regression analysis, and the variable selection was performed following variable selection in multiple regression analysis. The final equations (models) ware evaluated with the multiple linear regression facility of the statistical package NCSS [45]. The results are discussed below.

2. RESULTS AND DISCUSSION

As mentioned above the objective of the present study is three-fold, that is, to investigate modeling of CA II inhibition using (i) topological indices; (ii) quantum-theoretical descriptors, and (iii) combination of topological and quantum-theoretical descriptors. Thus, we describe these results as below:

(i) Topological Modeling of CA-II Inhibition

The preliminary regression analysis indicated that statistically significant models start pouring with three variables in the regression analysis. Following maximum-R^2 method [46-48] several three to nine variable regression analyses were performed and the best models using distance-based and Balaban indices are given in (Table **6**). A perusal of this (Table **6**) shows that in all the five models Balaban type index (s) is / (are) present in all the proposed models and also that Wiener index (W) is also present in these proposed models. The results also show that the quality of the model goes on increasing with further addition of other types of Balaban indices. The quality of the model is further improved with the addition of topological indices, in particular first-order connecting index ($^1\chi$). The six and seven parametric models appear to be the best to model CA II inhibition. We have, therefore, to make a better choice in between these two models.

The six-parametric model containing W, J_{hetm}, J_{hete}, BAC, $^1\chi$ and Sz as the correlating parameters is found as below:

$$\log K_i\ (\text{hCA-II}) = 2.144 + 7.79 \times 10^{-4} (\pm 4.600 \times 10^{-4})\, W - 1.5933 (\pm 0.1596)\, Jhetm + 4.4678 (\pm 0.3407)\, Jhete - 0.0117 (\pm 0.0032)\, BAC - 0.9062 (\pm 0.1297)\, {}^1\chi + 0.0012 (\pm 3.370 \times 10^{-4})\, Sz \quad (\mathbf{1})$$

$n = 95$, $Se = 0.490$, $R = 0.913$, $R^2_A = 0.823$, $F = 73.775$, $Q = 1.865$

The other seven-parametric model is found as below:

$$\log K_i\ (\text{hCA-II}) = 2.550 + 5.79 \times 10^{-4} (\pm 4.604 \times 10^{-4})\, W - 1.819 (\pm 0.187)\, Jhetm + 4.665 (\pm 0.346)\, Jhete - 0.0113 (\pm 0.0032)\, BAC - 0.9077 (\pm 0.1270)\, {}^1\chi + 0.0013 (\pm 3.343 \times 10^{-4})\, Sz - 0.3577 (\pm 0.1637)\, I_4 \quad (\mathbf{2})$$

$n = 95$, $Se = 0.480$, $R = 0.918$, $R^2_A = 0.830$, $F = 63.631$, $Q = 1.914$

Both these equations are statistically sound. However, there are no significant changes in the statistics when we go from six- to seven-parametric model. Furthermore, no other higher parametric regression gave still better model. Thus, our results go in favor of six-parametric model i.e. eq. (**1**). However, the final choice can be made by estimating $\log K_i$ (hCA-II) from both the models and investigate their residual properties.. The calculation of predictive correlation coefficient, R^2_{pred}, will help in deciding the problem. This R^2_{pred} is calculated from the plot of observed and calculated (estimated) $\log K_i$ (hCA-II) for these models.

The observed and calculated values obtained from eqs. (1) are shown in Table **7**. Using eq. (1), the R^2_{pred} was found out to be 0.8342 ($R = 0.9134$) (Fig. **1**).

However, this model contains compounds **7, 11, 56** and **57** as outliers. Deleating these outliers from regression procedure yielded the following model with much improved statistics:

Table 1. Structural Details of Carbonic anhysrase Used in Present in Investigation

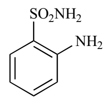
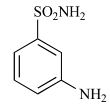
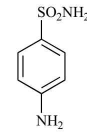
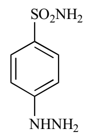
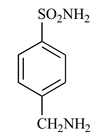
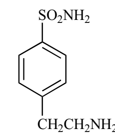
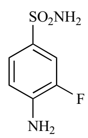
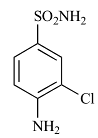
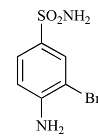
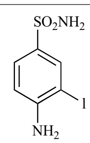
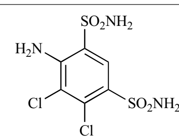
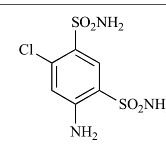
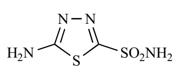
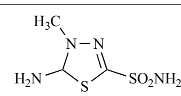
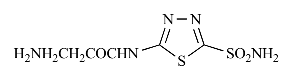
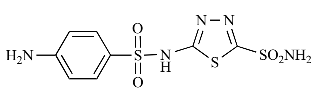
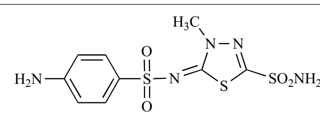
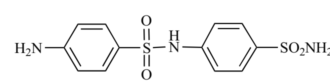
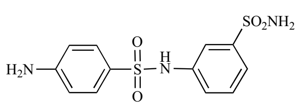
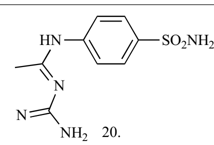
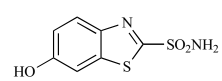
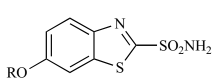
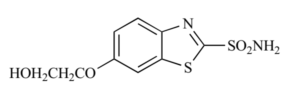
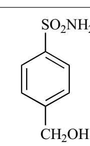
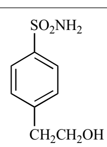
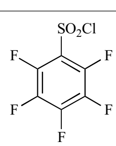
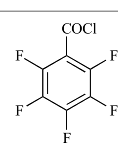
 1.	 2.	 3.
 4.	 5.	 6.
 7.	 8.	 9.
 10.	 11.	 12.
 13.	 14.	 15.
 16.	 17.	 18.
 19.	 20.	 21.
 22.	 23.	 24.
 25.	CF_3SO_2Cl A.	n-$C_4F_9SO_2Cl$ B.
 C	 D	

Table 2. The Inhibition Activity: log Ki (hCAII) and Values of Indicator Parameters (I_1, I_2, I_3 & I_4) Used in Present Study

Compd. No.	log Ki (hCAII)	I_1	I_2	I_3	I_4
1.	4.311	1	0	0	0
2.	4.272	1	0	0	0
3.	4.037	1	0	0	0
4.	4.170	1	0	0	0
5.	3.699	1	0	0	0
6.	2.778	1	0	0	0
7.	2.699	1	0	0	0
8.	2.863	1	0	0	0
9.	3.017	1	0	0	0
10.	2.633	1	0	0	0
11.	1.954	1	0	0	0
12.	2.000	1	0	0	0
13.	1.380	1	0	0	0
14.	1.114	1	0	0	0
15.	0.477	1	0	0	0
16.	0.699	1	0	0	0
17.	1.322	1	0	0	0
18.	1.362	1	0	0	0
19.	1.398	1	0	0	0
20.	-0.046	1	0	0	0
21.	-0.046	1	0	0	0
22.	0.000	1	0	0	0
23.	3.708	1	0	0	0
24.	2.740	1	0	0	0
25.	2.398	0	1	0	0
26.	2.230	0	1	0	0
27.	2.204	0	1	0	0
28.	2.255	0	1	0	0
29.	2.176	0	1	0	0
30.	2.176	0	1	0	0
31.	1.991	0	1	0	0
32.	2.628	0	1	0	0
33.	2.686	0	1	0	0
34.	1.708	0	1	0	0
35.	0.903	0	1	0	0
36.	0.301	0	1	0	0

(Table 2. Contd….)

Compd. No.	log Ki (hCAII)	I_1	I_2	I_3	I_4
37.	0.301	0	1	0	0
38.	0.477	0	1	0	0
39.	0.301	0	1	0	0
40.	0.602	0	1	0	0
41.	1.176	0	1	0	0
42.	1.301	0	1	0	0
43.	1.301	0	1	0	0
44.	-0.301	0	1	0	0
45.	-0.301	0	1	0	0
46.	0.000	0	1	0	0
47.	2.663	0	1	0	0
48.	2.585	0	1	0	0
49.	1.380	0	0	1	0
50.	1.000	0	0	1	0
51.	1.000	0	0	1	0
52.	1.204	0	0	1	0
53.	1.176	0	0	1	0
54.	1.176	0	0	1	0
55.	0.954	0	0	1	0
56.	2.041	0	0	1	0
57.	2.097	0	0	1	0
58.	1.176	0	0	1	0
59.	0.699	0	0	1	0
60.	-0.523	0	0	1	0
61.	-0.523	0	0	1	0
62.	-0.398	0	0	1	0
63.	0.000	0	0	1	0
64.	0.176	0	0	1	0
65.	0.903	0	0	1	0
66.	0.903	0	0	1	0
67.	1.041	0	0	1	0
68.	-0.699	0	0	1	0
69.	-0.523	0	0	1	0
70.	-0.301	0	0	1	0
71.	1.602	0	0	1	0
72.	1.544	0	0	1	0

(Table 2. Contd….)

Compd. No.	log Ki (hCAII)	I_1	I_2	I_3	I_4
73.	1.544	0	0	0	1
74.	1.279	0	0	0	1
75.	1.230	0	0	0	1
76.	1.362	0	0	0	1
77.	1.301	0	0	0	1
78.	1.230	0	0	0	1
79.	1.176	0	0	0	1
80.	2.097	0	0	0	1
81.	2.193	0	0	0	1
82.	1.580	0	0	0	1
83.	1.079	0	0	0	1
84.	0.301	0	0	0	1
85.	0.176	0	0	0	1
86.	0.301	0	0	0	1
87.	0.903	0	0	0	1
88.	1.255	0	0	0	1
89.	1.556	0	0	0	1
90.	1.431	0	0	0	1
91.	-0.301	0	0	0	1
92.	-0.222	0	0	0	1
93.	-.1550	0	0	0	1
94.	1.732	0	0	0	1
95.	1.699	0	0	0	1

I_1 – Indicator parameter for presence (=1) or absence (=0) of Structure A.
I_2 – Indicator parameter for presence (=1) or absence (=0) of Structure B.
I_3 – Indicator parameter for presence (=1) or absence (=0) of Structure C.
I_4 – Indicator parameter for presence (=1) or absence (=0) of Structure F.

Table 3. List of Descriptors Used in the Present Study

S.N.	Index	Meaning	Ref.
(i) Topological indices.			
1.	*W*	Wiener index	32
2.	*Sz*	Szeged index	33-36
3.	$^1\chi$	First-order connectivity index	37
(ii) Balaban and Balaban type indices.			
4.	*J*	Balaban distance connectivity index	38
5.	*Jhetz*	Balaban-type index from z-weighted distance matrix (Barysz matrix)	39
6.	*Jhetm*	Balaban-type index from mass weighted distance matrix	39

(Table 3. Contd….)

S.N.	Index	Meaning	Ref.
7.	*Jhetv*	Balaban-type index from van der Waals weighted distance matrix	39
8.	*Jhete*	Balaban-type index from electro negativity weighted distance matrix	39
9.	*Jhetp*	Balaban-type index from polarizability weighted distance matrix	39
10.	*BAC*	Balaban centric index	39
(iii) Quantum-theoretical descriptors.			
11.	ϕH	Angle between node in highest occupied π orbital and SO_2NH_2 group, DFT (^{0})	39
12.	ϕL	Angle between node in lowest unoccupied π orbital and SO_2NH_2 group, DFT (^{0})	39
13.	E_H	Energy of highest occupied π orbital, (HOPO) DFT (eV)	39
14.	E_{SH}	Energy of second highest occupied π orbital, (SHOPO) DFT (eV)	39
15.	E_L	Energy of lowest unoccupied π orbital, (LUPO) DFT (eV)	39
16.	E_{SL}	Energy of second lowest unoccupied π orbital, (SLUPO) DFT (eV)	39
17.	Q_O	Mulliken charge on sulfonamide O, DFT	39
18.	Q_N	Mulliken charge on sulfonamide N, DFT	39
19.	Q_C	Mulliken charge on C attached to sulfonamide, DFT	39
20.	Q_H	Mulliken charge on sulfonamide H, DFT	39
(iv) Indicator parameters.			
21.	I_1	when tail A is present = 1; otherwise zero	Present work
22.	I_2	when tail B is present = 1; otherwise zero	Present work
23.	I_3	when tail C is present = 1; otherwise zero	Present work
24.	I_4	when tail F is present = 1; otherwise zero	Present work

Table 4. Various Topological Descriptors Used in the Present Study and Their Values

Compd. No.	*W*	*J*	*Jhetz*	*Jhetm*	*Jhetv*	*Jhete*	*Jhetp*	*BAC*	$^1\chi$	*Sz*
1.	572	2.804	5.314	5.340	2.584	3.780	2.770	71	7.950	739
2.	604	2.637	4.899	4.920	2.491	3.580	2.661	71	7.933	803
3.	636	2.502	4.577	4.596	2.411	3.414	2.566	71	7.933	867
4.	778	2.387	4.167	4.181	2.049	3.247	2.039	72	8.433	1030
5.	778	2.387	3.939	3.952	2.313	3.111	2.442	72	8.433	1030
6.	939	2.291	3.504	3.513	2.232	2.881	2.340	73	8.933	1212
7.	710	2.651	4.800	4.825	2.498	3.623	2.615	88	8.344	966
8.	710	2.651	4.859	4.880	2.569	3.612	2.741	88	8.344	966
9.	710	2.651	4.895	4.918	2.584	3.605	2.756	88	8.344	966
10.	1192	3.222	6.137	6.161	3.229	4.387	3.548	181	10.394	1598
11.	1096	3.074	5.855	5.877	3.064	4.186	3.364	156	9.966	1468
12.	525	2.569	5.678	5.711	2.166	3.275	2.428	71	7.433	586
13.	595	2.705	5.556	5.585	2.186	3.375	2.390	88	7.844	662

(Table 4. Contd….)

Compd. No.	*W*	*J*	*J*hetz	*J*hetm	*J*hetv	*J*hete	*J*hetp	*BAC*	$^{1}\chi$	*Sz*
14.	1245	2.293	3.528	3.536	1.915	2.718	1.979	92	9.827	1336
15.	2069	1.842	3.119	3.124	1.781	2.359	1.936	107	12.117	2718
16.	1796	1.908	3.909	3.919	2.022	2.582	2.300	107	11.644	2406
17.	2334	1.815	3.217	3.223	1.846	2.440	2.006	107	12.617	3336
18.	2262	1.870	3.337	3.343	1.885	2.507	2.052	107	12.617	3192
19.	1678	1.804	2.904	2.910	1.670	2.493	1.640	71	11.383	2430
20.	945	1.948	3.699	3.711	1.916	2.594	2.095	71	9.400	1298
21.	945	1.948	3.821	3.834	1.731	2.677	1.878	71	9.400	1298
22.	1564	1.712	2.841	2.846	1.354	2.258	1.372	74	10.900	2007
23.	778	2.387	4.058	4.071	2.080	3.215	2.187	72	8.433	1030
24.	939	2.291	3.591	3.600	2.030	2.963	2.122	73	8.933	1212
25.	1670	3.233	5.008	5.039	2.758	3.942	2.809	206	11.700	1954
26.	1738	3.103	4.788	4.817	2.693	3.812	2.740	206	11.683	2090
27.	1806	2.989	4.60	4.627	2.634	3.696	2.677	206	11.683	2226
28.	2092	2.858	4.352	4.374	2.328	3.559	2.270	207	12.183	2533
29.	2092	2.858	4.190	4.210	2.542	3.453	2.582	207	12.183	2533
30.	2406	2.744	3.866	3.883	2.461	3.253	2.498	208	12.683	2868
31.	1952	3.084	4.752	4.781	2.694	3.823	2.723	235	12.094	2406
32.	1952	3.084	4.778	4.806	2.731	3.818	2.786	235	12.094	2406
33.	1952	3.084	4.794	4.823	2.739	3.815	2.793	235	12.094	2406
34.	2830	3.483	5.652	5.681	3.18	4.358	3.326	376	14.144	3488
35.	2662	3.372	5.449	5.477	3.065	4.214	3.199	339	13.716	3277
36.	1578	3.065	5.079	5.114	2.501	3.619	2.611	206	11.183	1693
37.	1720	3.148	5.517	5.555	2.684	3.888	2.814	235	11.594	1841
38.	2973	2.695	3.820	3.834	2.214	3.087	2.239	239	13.577	3118
39.	4400	2.009	3.387	3.398	1.855	2.500	1.986	266	15.867	5535
40.	4700	2.037	3.540	3.552	1.929	2.598	2.074	299	16.278	5901
41.	4836	1.981	3.212	3.221	1.901	2.525	2.004	266	16.367	6486
42.	4728	2.022	3.286	3.296	1.927	2.570	2.033	266	16.367	6270
43.	3694	2.026	3.059	3.069	1.775	2.613	1.739	206	15.133	4914
44.	2412	2.160	3.482	3.500	1.957	2.658	2.043	206	13.150	3035
45.	2412	2.160	3.570	3.587	1.792	2.730	1.867	206	13.150	3035
46.	3571	1.896	2.872	2.881	1.442	2.361	1.441	209	14.650	4284
47.	2092	2.858	4.282	4.303	2.336	3.543	2.373	207	12.183	2533
48.	2406	2.744	3.943	3.960	2.271	3.331	2.306	208	12.683	2868
49.	1382	2.213	3.962	3.980	2.170	2.996	2.294	102	11.376	1982
50.	1442	2.117	3.751	3.767	2.109	2.880	2.223	102	11.359	2102

(Table 4. Contd….)

Compd. No.	*W*	*J*	*J*hetz	*J*hetm	*J*hetv	*J*hete	*J*hetp	*BAC*	$^{1}\chi$	*Sz*
51.	1502	2.035	3.575	3.589	2.053	2.779	2.160	102	11.359	2222
52.	1755	1.940	3.317	3.328	1.759	2.643	1.752	102	11.859	2514
53.	1755	1.940	3.173	3.184	1.949	2.554	2.041	102	11.859	2514
54.	2034	1.859	2.872	2.880	1.862	2.374	1.942	102	12.359	2832
55.	1631	2.103	3.680	3.696	2.086	2.870	2.176	123	11.770	2401
56.	1631	2.103	3.680	3.696	2.086	2.870	2.176	123	11.770	2401
57.	1631	2.103	3.680	3.696	2.086	2.870	2.176	123	11.770	2401
58.	2416	2.387	4.328	4.344	2.428	3.233	2.62	230	13.819	3468
59.	2265	2.308	4.180	4.194	2.345	3.129	2.527	201	13.392	3258
60.	1301	2.091	4.085	4.105	1.950	2.731	2.125	102	10.859	1728
61.	1301	2.091	4.085	4.105	1.950	2.731	2.125	102	10.859	1728
62.	2541	1.825	2.804	2.811	1.626	2.221	1.678	123	13.253	3088
63.	3829	1.539	2.814	2.821	1.528	2.020	1.677	146	15.542	5360
64.	4100	1.561	2.947	2.954	1.591	2.103	1.752	171	15.953	5709
65.	4226	1.516	2.638	2.644	1.565	2.037	1.685	146	16.042	6254
66.	4126	1.549	2.707	2.713	1.589	2.077	1.714	146	16.042	6054
67.	3195	1.549	2.523	2.529	1.469	2.132	1.458	102	14.808	4779
68.	2039	1.661	3.006	3.016	1.685	2.216	1.818	102	12.825	2998
69.	2039	1.661	3.093	3.103	1.524	2.284	1.635	102	12.825	2998
70.	3075	1.451	2.374	2.379	1.178	1.920	1.193	102	14.325	4178
71.	1755	1.940	3.254	3.265	1.767	2.629	1.847	102	11.859	2514
72.	2340	1.780	2.867	2.873	1.434	2.406	1.433	102	12.859	3177
73.	1282	2.106	2.955	2.965	2.013	2.770	1.910	83	11.052	1851
74.	1338	2.014	2.830	2.829	1.957	2.665	1.858	83	11.036	1963
75.	1394	1.935	2.722	2.730	1.906	2.573	1.811	83	11.036	2075
76.	1634	1.849	2.585	2.592	1.643	2.456	1.510	83	11.536	2354
77.	1634	1.849	2.492	2.499	1.817	2.375	1.734	83	11.536	2354
78.	1899	1.775	2.312	2.317	1.742	2.218	1.669	83	12.036	2658
79.	1518	2.005	2.831	2.840	1.976	2.660	1.883	102	11.446	2248
80.	1518	2.005	2.831	2.840	1.976	2.660	1.883	102	11.446	2248
81.	1518	2.005	2.840	2.850	1.983	2.657	1.888	102	11.446	2248
82.	2274	2.296	3.393	3.403	2.291	3.040	2.248	201	13.496	3280
83.	2128	2.217	3.262	3.271	2.207	2.935	2.161	174	13.068	3076
84.	1203	1.986	3.001	3.012	1.804	2.521	1.771	83	10.536	1606
85.	1203	1.986	3.001	3.012	1.804	2.521	1.771	83	10.536	1606
86.	2383	1.749	2.308	2.312	1.531	2.088	1.475	102	12.929	2906
87.	3619	1.490	2.350	2.354	1.457	1.925	1.494	123	15.219	5078

(Table 4. Contd….)

Compd. No.	*W*	*J*	*J*hetz	*J*hetm	*J*hetv	*J*hete	*J*hetp	*BAC*	$^1\chi$	*Sz*
88.	4000	1.470	2.232	2.236	1.497	1.945	1.509	123	15.719	5938
89.	3904	1.503	2.285	2.289	1.520	1.984	1.533	123	15.719	5746
90.	3014	1.494	2.091	2.095	1.396	2.018	1.298	83	14.485	4528
91.	1907	1.594	2.352	2.358	1.583	2.078	1.554	83	12.502	2818
92.	1907	1.594	2.432	2.438	1.431	2.142	1.385	83	12.502	2818
93.	2892	1.401	1.996	1.999	1.121	1.822	1.062	83	14.002	3947
94.	1634	1.849	2.569	2.575	1.645	2.446	1.553	83	11.536	2354
95.	2190	1.711	2.348	2.352	1.347	2.258	1.250	83	12.536	2988

W- Wiener index; *J*- Balaban distance connectivity index; *J*hetz-Balaban-type index from *z*-weighted distance matrix (Barysz matrix); *J*hetm- Balaban-type index from mass weighted distance matrix; *J*hetv- Balaban-type index from van der Waals weighted distance matrix; *J*hete- Balaban-type index from electro negativity weighted distance matrix; *J*hetp- Balaban-type index from polarizability weighted distance matrix; *BAC*- Balaban centric index; $^1\chi$-First order Randic connectivity index and *Sz*- Szeged index.

Table 5. Various Quantum-Theoretical Descriptors Used in the Present Study and Their Values

Compd. No.	ϕ_H	ϕ_L	E_H	E_{SH}	E_L	E_{SL}	Q_O	Q_N	Q_C	Q_H
1.	149.6	34.8	-6.944	-7.285	-1.170	-0.719	-1.071	-0.845	-0.168	0.806
2.	88.9	55.0	-7.188	-7.450	-1.156	-0.820	-1.073	-0.845	-0.132	0.826
3.	88.7	44.1	-6.541	-7.417	-0.923	-0.612	-1.08	-0.847	-0.117	0.820
4.	89.7	45.4	-6.377	-7.248	-0.769	-0.425	-1.084	-0.848	-0.120	0.818
5.	92.2	42.4	-7.003	-7.281	-1.057	-0.504	-1.078	-0.847	-0.115	0.821
6.	91.4	43.0	-6.789	-7.166	-0.878	-0.367	-1.083	-0.848	-0.117	0.818
7.	76.0	39.7	-6.699	-7.355	-1.165	-0.630	-1.071	-0.846	-0.110	0.824
8.	75.0	43.9	-6.777	-7.316	-1.284	-0.853	-1.071	-0.845	-0.105	0.826
9.	74.7	44.5	-6.742	-7.259	-1.269	-0.882	-1.071	-0.845	-0.105	0.826
10.	71.3	27.1	-7.181	-7.270	-1.981	-1.512	-1.036	-0.846	-0.099	0.842
11.	117.7	31.6	-7.192	-7.479	-1.854	-1.315	-1.036	-0.849	-0.100	0.836
12.	112.8	59.8	-6.973	-9.247	-2.002	-0.315	-1.020	-0.822	-0.172	0.849
13.	119.4	58.0	-6.947	-9.256	-2.016	-1.231	-1.001	-0.826	-0.152	0.854
14.	111.8	60.5	-6.597	-8.716	-1.864	-0.546	-1.035	-0.828	-0.189	0.839
15.	114.1	57.5	-6.487	-8.707	-1.744	0.053	-1.030	-0.828	-0.177	0.842
16.	118.5	54.1	-6.413	-8.407	-1.754	-1.245	-1.013	-0.830	-0.153	0.847
17.	88.9	46.5	-6.149	-7.238	-0.598	-0.385	-1.089	-0.850	-0.119	0.814
18.	31.0	43.0	-6.282	-7.038	-0.893	-0.351	-1.081	-0.848	-0.123	0.821
19.	89.0	44.3	-6.036	-7.155	-1.314	-0.348	-1.085	-0.849	-0.117	0.816
20.	113.8	51.5	-6.488	-6.920	-1.812	-0.613	-1.025	-0.826	-0.147	0.844
21.	123.2	52.3	-6.737	-7.013	-1.901	-0.756	-1.023	-0.826	-0.146	0.845
22.	114.3	49.1	-6.147	-6.833	-1.654	-0.434	-1.03	-0.828	-0.149	0.840
23.	93.0	42.5	-7.023	-7.310	-1.029	-0.512	-1.077	-0.847	-0.113	0.822

(Table 5. Contd….)

Compd. No.	ϕ_H	ϕ_L	E_H	E_{SH}	E_L	E_{SL}	Q_O	Q_N	Q_C	Q_H
24.	91.2	43.7	-6.838	-7.187	-0.931	-0.384	-1.082	-0.849	-0.117	0.818
25.	150.3	33.1	-6.927	-7.286	-1.214	-0.777	-1.071	-0.843	-0.170	0.802
26.	36.1	50.7	-6.803	-7.202	-1.067	-0.776	-1.074	-0.845	0-.126	0.825
27.	90.4	46.2	-6.605	-7.382	-1.096	-0.554	-1.079	-0.848	-0.115	0.819
28.	90.4	44.2	-6.406	-7.243	-0.793	-0.420	-1.082	-0.848	-0.119	0.817
29.	91.8	42.6	-7.011	-7.287	-1.078	-0.517	-1.077	-0.847	-0.115	0.821
30.	88.3	47.3	-6.797	-7.160	-0.867	-0.372	-1.083	-0.849	0-.118	0.818
31.	76.2	39.8	-6.695	-7.353	-1.227	-0.629	-1.071	-0.846	-0.110	0.824
32.	75.5	43.3	-6.752	-7.299	-1.252	-0.829	-1.071	-0.846	-0.105	0.825
33.	75.6	43.8	-6.719	-7.250	-1.280	-0.87	-1.071	-0.846	-0.105	0.824
34.	71.9	27.9	-7.167	-7.266	-2.000	-1.522	-1.035	-0.846	-0.099	0.842
35.	114.9	28.6	-7.139	-7.242	-1.784	-1.308	-1.036	-0.849	-0.102	0.835
36.	112.8	58.6	-6.84	-9.260	-1.888	-0.300	-1.021	-0.823	-0.176	0.848
37.	119.3	59.6	-6.969	-9.369	-2.068	-1.295	-1.001	-0.826	-0.152	0.855
38.	111.8	60.8	-6.616	-9.101	-1.894	-0.615	-1.034	-0.827	-0.188	0.839
39.	114.1	57.6	-6.492	-8.724	-1.746	0.054	-1.029	-0.826	-0.177	0.842
40.	118.5	54.1	-6.424	-8.836	-1.756	-1.269	-1.012	-0.830	-0.154	0.846
41.	88.8	45.8	-6.152	-7.240	-0.767	-0.388	-1.089	-0.850	-0.119	0.814
42.	31.0	43.1	-6.292	-7.038	-0.896	-0.362	-1.081	-0.848	-0.123	0.821
43.	88.9	44.3	-6.031	-7.151	-1.308	-0.346	-1.085	-0.849	-0.117	0.816
44.	113.4	51.4	-6.491	-6.910	-1.79	-0.596	-1.025	-0.825	-0.147	0.844
45.	123.4	52.2	-6.726	-6.992	-1.877	-0.632	-1.023	-0.825	-0.146	0.845
46.	114.0	49.1	-6.121	-6.824	-1.644	-0.417	-1.030	-0.828	-0.149	0.840
47.	94.6	42.3	-7.037	-7.303	-1.271	-0.523	-1.077	-0.847	-0.113	0.822
48.	91.3	43.5	-6.833	-7.182	-0.937	-0.379	-1.082	-0.849	-0.117	0.818
49.	148.0	36.0	-6.745	-7.201	-1.075	-0.579	-1.081	-0.863	-0.155	0.819
50.	33.4	42.3	-6.554	-7.136	-0.956	-0.547	-1.077	-0.848	-0.123	0.821
51.	88.7	47.4	-6.386	-7.314	-0.803	-0.450	-1.085	-0.848	-0.116	0.817
52.	87.8	43.8	-6.141	-7.210	-0.737	-0.428	-1.087	-0.848	-0.120	0.816
53.	86.6	48.0	-6.907	-7.195	-0.946	-0.425	-1.081	-0.848	-0.115	0.819
54.	87.8	46.8	-6.749	-7.149	-0.867	-0.350	-1.084	-0.849	-0.117	0.817
55.	77.8	44.2	-6.478	-7.272	-0.981	-0.539	-1.077	-0.847	-0.111	0.822
56.	78.6	48.8	-6.525	-7.225	-1.034	-0.742	-1.077	-0.847	-0.106	0.822
57.	78.2	51.4	-6.506	-7.178	-1.038	-0.770	-1.077	-0.847	-0.106	0.822
58.	86.2	30.8	-6.847	-7.237	-1.898	-1.473	-1.039	-0.848	-0.096	0.838
59.	93.6	29.3	-6.729	-7.249	-1.661	-1.339	-1.047	-0.850	-0.111	0.833
60.	113.1	58.7	-6.736	-9.238	-1.933	-0.171	-1.011	-0.826	-0.177	0.845
61.	118.7	59.3	-6.689	-8.931	-1.958	-1.425	-1.009	-0.829	-0.152	0.850

(Table 5. Contd….)

Compd. No.	ϕ_H	ϕ_L	E_H	E_{SH}	E_L	E_{SL}	Q_O	Q_N	Q_C	Q_H
62.	111.8	60.7	-6.600	-8.709	-1.879	-0.599	-1.035	-0.828	-0.188	0.838
63.	114.6	60.8	-6.498	-7.032	-2.042	-0.118	-1.027	-0.826	-0.170	0.844
64.	118.0	54.9	-6.855	-8.395	-1.775	-1.167	-1.014	-0.830	-0.153	0.847
65.	88.9	45.7	-6.135	-7.239	-0.722	-0.381	-1.089	-0.850	-0.119	0.813
66.	31.7	35.8	-6.231	-7.098	-0.860	-0.320	-1.08	-0.848	-0.123	0.820
67.	88.7	44.7	-5.957	-7.079	-1.237	-0.315	-1.087	-0.849	-0.117	0.815
68.	110.0	51.5	-6.239	-6.851	-1.756	-0.502	-1.028	-0.827	-0.148	0.842
69.	117.4	50.9	-6.549	-6.899	-1.765	-0.567	-1.027	-0.827	-0.146	0.842
70.	113.7	49.1	-6.108	-6.816	-1.646	-0.413	-1.031	-0.828	-0.149	0.840
71.	85.6	47.4	-6.904	-7.218	-0.990	-0.429	-1.081	-0.848	-0.113	0.818
72.	91.9	43.1	-6.811	-7.164	-0.880	-0.359	-1.084	-0.849	-0.116	0.817
73.	152.5	43.2	-6.546	-7.143	-0.954	-0.508	-1.081	-0.847	-0.193	0.804
74.	31.0	40.8	-6.406	-7.075	-0.855	-0.490	-1.081	-0.848	-0.120	0.820
75.	87.9	47.1	-6.346	-7.253	-0.748	-0.382	-1.086	-0.849	-0.114	0.816
76.	87.8	46.2	-6.124	-7.174	-0.656	-0.350	-1.088	-0.849	-0.121	0.814
77.	87.2	47.4	-6.731	-7.196	-0.868	-0.398	-1.084	-0.848	-0.117	0.818
78.	91.0	43.2	-6.675	-7.140	-0.815	-0.332	-1.085	-0.849	-0.117	0.817
79.	76.8	43.7	-6.439	-7.226	-1.751	-0.469	-1.078	-0.847	-0.109	0.821
80.	77.8	46.5	-6.490	-7.190	-1.800	-0.678	-1.077	-0.847	-0.104	0.821
81.	77.4	46.8	-6.468	-7.136	-1.800	-0.712	-1.077	-0.847	-0.104	0.821
82.	77.9	36.0	-7.005	-7.135	-2.041	-1.477	-1.036	-0.849	-0.095	0.836
83.	100.7	35.1	-6.877	-7.302	-2.022	-1.369	-1.043	-0.851	-0.101	0.832
84.	111.9	64.0	-6.672	-8.765	-2.204	-1.407	-1.032	-0.828	-0.187	0.840
85.	117.3	65.8	-6.585	-8.899	-2.115	-1.387	-1.012	-0.831	-0.175	0.846
86.	111.8	60.5	-6.589	-8.700	-1.876	-0.574	-1.036	-0.829	-0.188	0.838
87.	112.9	57.3	-6.525	-8.971	-1.785	-0.032	-1.030	-0.829	-0.179	0.840
88.	88.8	45.5	-6.140	-7.246	-0.726	-0.375	-1.090	-0.850	-0.118	0.813
89.	30.5	42.6	-6.228	-7.058	-0.883	-0.304	-1.082	-0.848	-0.121	0.820
90.	88.8	44.4	-5.960	-7.190	-1.200	-0.306	-1.088	-0.850	-0.117	0.815
91.	110.7	53.3	-6.165	-6.805	-1.894	-0.431	-1.029	-0.827	-0.147	0.841
92.	120.7	49.8	-6.585	-6.869	-1.699	-0.507	-1.026	-0.827	-0.146	0.842
93.	113.6	49.0	-6.091	-6.812	-1.643	-0.407	-1.031	-0.832	-0.149	0.840
94.	93.7	42.4	-6.825	-7.199	-0.912	-0.385	-1.083	-0.849	-0.114	0.818
95.	88.1	46.9	-6.797	-7.155	-0.858	-0.350	-1.084	-0.849	-0.117	0.817

Φ_H - Angle between node in highest occupied π orbital and SO_2NH_2 group, DFT (^{0});
Φ_L - Angle between node in lowest unoccupied π orbital and SO_2NH_2 group, DFT (^{0});
E_H – Energy of highest occupied π orbital, (HOPO) DFT (eV);
E_{SH} -Energy of second highest occupied π orbital, (SHOPO) DFT (eV);
E_L -Energy of lowest unoccupied π orbital, (LUPO) DFT (eV);
E_{SL} -Energy of second lowest unoccupied π orbital, (SLUPO) DFT (eV);
Q_O – Mulliken charge on sulfonamide O, DFT; Q_N – Mulliken charge on sulfonamide N, DFT;
Q_C – Mulliken charge on C attached to sulfonamide, DFT; QH – Mulliken charge on sulfonamide H, DFT;

Table 6. The Best Variable Modeling of CA-II Inhibition Using Topological Indices Including Balaban and Balaban Type Indices

Model No.	TI used	*Se*	*R*	R^2A	*F*	*Q*
1.	*W*, Jhete, *BAC*	0.728	0.789	0.609	49.824	1.084
2.	*W*, Jhetm, Jhete, *BAC*	0.605	0.860	0.729	64.250	1.421
3.	*W*, Jhetm, Jhete, *BAC*, $^1\chi$	0.518	0.901	0.801	76.863	1.738
4.	*W*, Jhetm, Jhete, *BAC*, $^1\chi$, *Sz*	0.490	0.913	0.823	73.775	1.865
5.	*W*, Jhetm, Jhete, *BAC*, $^1\chi$, *Sz*, I_4	0.480	0.918	0.830	66.631	1.914
6.	*W*, Jhetm, Jhete, *BAC*, $^1\chi$, *Sz*, I_4	0.435	0.934	0.863	95.200	2.147

Table 7. Calculated and Observed logKi (hCAII) Using esq. (1) and (3)

Compd. No.	logKi (hCAII)				
	Equation-1		Equation-3		
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
1.	4.311	3.793	0.518	4.031	0.280
2.	4.272	3.683	0.589	3.863	0.409
3.	4.037	3.557	0.480	3.688	0.349
4.	4.170	3.308	0.862	3.413	0.757
5.	3.699	3.065	0.634	3.129	0.570
6.	2.778	2.609	0.169	2.632	0.146
7.	2.699	3.727	-1.028	-	-
8.	2.863	3.590	-0.727	3.747	-0.884
9.	3.017	3.498	-0.481	3.653	-0.636
10.	2.633	3.174	-0.541	3.390	-0.757
11.	1.954	3.183	-1.229	-	-
12.	2.000	1.200	0.800	1.282	0.718
13.	1.380	1.419	-0.039	1.499	-0.119
14.	1.114	1.196	-0.082	1.203	-0.089
15.	0.477	0.245	0.232	0.202	0.275
16.	0.699	-0.173	0.872	-0.201	0.900
17.	1.322	0.920	0.402	0.911	0.411
18.	1.362	0.804	0.558	0.811	0.551
19.	1.398	1.631	-0.233	1.637	-0.239
20.	-0.046	0.716	-0.762	0.647	-0.693
21.	-0.046	0.891	-0.937	0.847	-0.893
22.	0.000	0.506	-0.506	0.447	-0.447
23.	3.708	3.340	0.368	3.436	0.272
24.	2.740	2.837	-0.097	2.885	-0.145

(Table 7. Contd….)

Compd. No.	**logKi (hCAII)**				
	Equation-1		**Equation-3**		
	Actual	***Est.***	***Res.***	***Est.***	***Res.***
25.	2.398	2.285	0.113	2.413	-0.015
26.	2.230	2.284	-0.054	2.377	-0.147
27.	2.204	2.280	-0.076	2.341	-0.137
28.	2.255	2.186	0.069	2.261	-0.006
29.	2.176	1.974	0.202	2.016	0.160
30.	2.176	1.771	0.405	1.812	0.364
31.	1.991	2.212	-0.221	2.257	-0.266
32.	2.628	2.150	0.478	2.193	0.435
33.	2.686	2.110	0.576	2.152	0.534
34.	1.708	1.600	0.108	1.566	0.142
35.	0.903	1.727	-0.824	1.724	-0.821
36.	0.301	0.816	-0.515	0.845	-0.544
37.	0.301	0.886	-0.585	0.944	-0.643
38.	0.477	0.670	-0.193	0.731	-0.254
39.	0.301	0.270	0.031	0.240	0.061
40.	0.602	0.363	0.239	0.322	0.280
41.	1.176	1.654	-0.478	1.655	-0.479
42.	1.301	1.401	-0.100	1.408	-0.107
43.	1.301	1.394	-0.093	1.414	-0.113
44.	-0.301	-0.479	0.178	-0.635	0.334
45.	-0.301	-0.296	-0.005	-0.430	0.129
46.	0.000	0.144	-0.144	0.118	-0.118
47.	2.663	2.228	0.435	2.297	0.366
48.	2.585	1.997	0.588	2.062	0.523
49.	1.380	1.065	0.315	1.096	0.284
50.	1.000	1.087	-0.087	1.088	-0.088
51.	1.000	1.106	-0.106	1.079	-0.079
52.	1.204	0.998	0.206	0.978	0.226
53.	1.176	0.830	0.346	0.783	0.393
54.	1.176	0.644	0.532	0.596	0.580
55.	0.954	1.032	-0.078	0.995	-0.041
56.	2.041	1.032	1.009	-	-
57.	2.097	1.032	1.065	-	-
58.	1.176	0.362	0.814	0.261	0.915
59.	0.699	0.501	0.198	0.427	0.272

(Table 7. Contd….)

Compd. No.	**logKi (hCAII)**				
	Equation-1		**Equation-3**		
	Actual	***Est.***	***Res.***	***Est.***	***Res.***
60.	-0.523	-0.208	-0.315	-0.255	-0.268
61.	-0.523	-0.208	-0.315	-0.255	-0.268
62.	-0.398	-0.290	-0.108	-0.306	-0.092
63.	0.000	0.095	-0.095	0.090	-0.09
64.	0.176	0.205	-0.029	0.201	-0.025
65.	0.903	1.346	-0.443	1.365	-0.462
66.	0.903	1.105	-0.202	1.129	-0.226
67.	1.041	1.070	-0.029	1.089	-0.048
68.	-0.699	-0.505	-0.194	-0.621	-0.078
69.	-0.523	-0.339	-0.184	-0.435	-0.088
70.	-0.301	0.009	-0.31	0.000	-0.301
71.	1.602	1.036	0.566	1.012	0.590
72.	1.544	0.986	0.558	1.006	0.538
73.	1.544	1.958	-0.414	1.959	-0.415
74.	1.279	1.894	-0.615	1.867	-0.588
75.	1.230	1.814	-0.584	1.762	-0.532
76.	1.362	1.570	-0.208	1.528	-0.166
77.	1.301	1.356	-0.055	1.289	0.012
78.	1.230	1.052	0.178	0.987	0.243
79.	1.176	1.731	-0.555	1.672	-0.496
80.	2.097	1.731	0.366	1.672	0.425
81.	2.193	1.702	0.491	1.642	0.551
82.	1.580	1.302	0.278	1.203	0.377
83.	1.079	1.397	-0.318	1.319	-0.240
84.	0.301	0.893	-0.592	0.820	-0.519
85.	0.176	0.893	-0.717	0.820	-0.644
86.	0.301	0.116	0.185	0.089	0.212
87.	0.903	0.485	0.418	0.484	0.419
88.	1.255	1.604	-0.349	1.626	-0.371
89.	1.556	1.396	0.160	1.423	0.133
90.	1.431	1.335	0.096	1.345	0.086
91.	-0.301	0.131	-0.432	0.005	-0.306
92.	-0.222	0.289	-0.511	0.184	-0.406
93.	-0.155	0.282	-0.437	0.265	-0.420
94.	1.732	1.553	0.179	1.507	0.225
95.	1.699	1.333	0.366	1.337	0.362

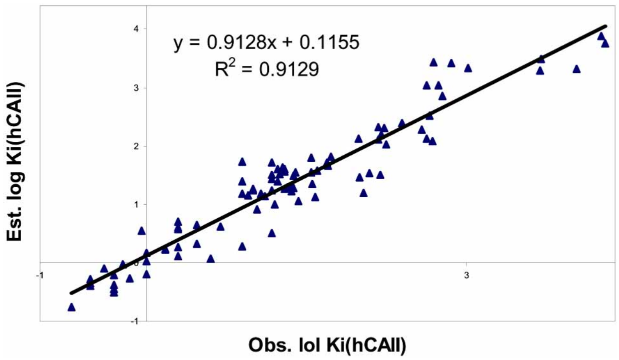


Fig. (1). Observed and calculated logKi (hCAII) using eq. (15).

$\log K_i$ (hCA-II) = 1.292 + 1.223 × 10^{-4} (± 4.148 × 10^{-4}) *W* – 1.596 (± 0.143) *Jhetm*

+ 4.780 (± 0.310) *Jhete* – 0.015 (± 2.930 × 10^{-3}) *BAC* – 0.901

(± 0.117) $^1\chi$ + 1.024 × 10^{-3} (± 3.016 × 10^{-4}) *Sz* **(3)**

$n = 91$, $Se = 0.435$, $R = 0.934$, $R^2_A = 0.863$, $F = 95.200$, $Q = 2.147$

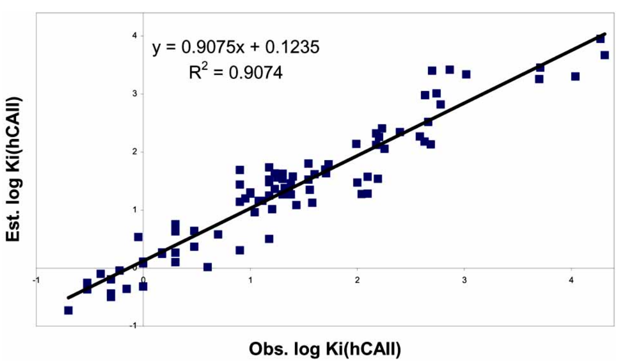


Fig. (2). Observed and calculated logKi (hCAII) using eq. (16).

The observed and calculated $\log K_i$ (hCAII) from eq (3) are given in Table **7** and the correlation between observed and calculated $\log K_i$ (hCA-II) yielded R^2_{pred} as 0.8718 (R = 0.9337).

In case of 7-parametric model eq. (**2**) (see Table **8** for observed and calculated values) the correlation of observed and calculated $\log K_i$ (hCA-II) (Fig. **3**) gave R^2_{pred} as 0.8428 (R = 0.9180). In this case we observed four compounds: **7, 11, 12** and **21** as outliers. The deletion of these compounds from the regression procedure yielded the following model with much improved statistics:

$\log K_i$ (hCA-II) = 1.761 + 5.930 × 10^{-4} (± 4.138 × 10^{-4}) *W* – 1.951 (± 0.190) *Jhetm*

+ 5.056 (± 0.343) *Jhete* – 0.013 (± 2.920 x 10^{-4}) *BAC* – 0.875 (± 0.117) $^1\chi$

+ 1.286 × 10^{-3} (± 3.003 × 10^{-4}) *Sz* – 0.438 (± 0.153) I_4 **(4)**

$n = 91$, $Se = 0.428$, $R = 0.936$, $R^2_A = 0.866$, $F = 84.048$, $Q = 2.187$

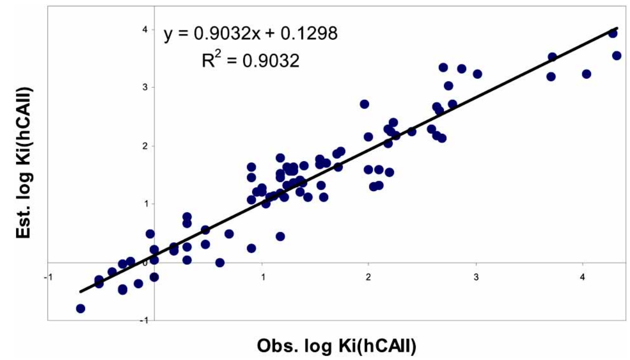


Fig. (3). Observed and calculated logKi (hCAII) using eq. (17).

The correlation of observed and calculated $\log K_i$ (hCA-II) from eqn. (4) (see Table **8** for observed and calculated values) gave R^2_{pred} equal to 0.8764 (R = 0.9362). The above results clearly indicate that only a marginal improvement in statistics occurred as we go from 6- to 7-parametric regression analysis. It means that the 6-parametric eq. (**1**) and its modification after deleting the four outliers i.e. eq. (**3**) are the most appropriate regression expressions for modeling $\log K_i$ (hCA-II).

In eq. (**1**) and (**3**) the six parameters involved are *W, Jhetm, Jhete, BAC,* $^1\chi$ and *Sz*. The changes in the sign of the coefficients of these parameters may probably be due to possible co- linearity existing among them. This co- linearity aspect is dealt with separately in the following discussion. At this stage it is enough to state that t-values of these correlating parameters are 2.707, -11.150, 15.444, -4.979, -7.672 and 3.394 respectively for *W, Jhetm, Jhete, BAC,* $^1\chi$ and *Sz*. All these indices are statistically significant according to values at all level of $p < 0.0001$. Both these models (eq.(**1**) and (**2**)) produces standard error of only 0.490 and 0.435 and explains more than 93.88% of the variance in the experimental $\log K_i$(hCA-II) for the compounds under present study.

Predictive Power of eq. (1) and (3)

We now investigate the predictive power of the proposed models (eq. (**1**) and (**3**)). The initial way to investigate predictive power is to calculate Pogalani's quality factor *Q* [49-51]. This quality factor *Q* is defined as the ratio of correlation coefficient to the standard error of estimation i.e., $Q = R/Se$. This means that higher the *R*, lower the *Se*, the larger will be the value of *Q* and the better will be the predictive power. In case of eq. (**1**) the *Q* value is found to be 1. 914, which is improved in eq. (**3**) yielding Q as 2.147. This improvement is due to the removal of four outliers. This is further supported by their predictive correlation coefficients R^2_{pred} as discussed above.

Cross-Validation

In principal, cross-validation is a practical and reliable method for testing the significance of a model. Hence, to validate the final models generated individually for different activities / properties, leave-one-out method is used to do cross-validation. The leave-one-out method consists of developing a number of models with one compound omitted at the time after developing each model. The omitted sample

Table 8. Observed and Calculated logKi (hCAII) Using eqs. (2) and (4)

Compd. No.	logKi (hCAII)				
	Equation-2			Equation-4	
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
1.	4.311	3.725	0.586	3.851	0.460
2.	4.272	3.672	0.600	3.775	0.497
3.	4.037	3.587	0.450	3.670	0.367
4.	4.170	3.387	0.783	3.478	0.692
5.	3.699	3.169	0.530	3.237	0.462
6.	2.778	2.755	0.023	2.810	-0.032
7.	2.699	3.749	-1.050	-	-
8.	2.863	3.597	-0.734	3.705	-0.842
9.	3.017	3.496	-0.479	3.595	-0.578
10.	2.633	3.055	-0.422	3.207	-0.574
11.	1.954	3.084	-1.130	-	-
12.	2.000	0.941	1.059	-	-
13.	1.380	1.209	0.171	1.110	0.270
14.	1.114	1.261	-0.147	1.250	-0.136
15.	0.477	0.326	0.151	0.303	0.174
16.	0.699	-0.206	0.905	-0.270	0.969
17.	1.322	1.010	0.312	1.034	0.288
18.	1.362	0.880	0.482	0.910	0.452
19.	1.398	1.820	-0.422	1.911	-0.513
20.	-0.046	0.767	-0.813	0.704	-0.750
21.	-0.046	0.931	-0.977	-	-
22.	0.000	0.640	-0.64	0.619	-0.619
23.	3.708	3.438	0.270	3.531	0.177
24.	2.740	2.979	-0.239	3.055	-0.315
25.	2.398	2.282	0.116	2.417	-0.019
26.	2.230	2.308	-0.078	2.423	-0.193
27.	2.204	2.325	-0.121	2.422	-0.218
28.	2.255	2.238	0.017	2.337	-0.082
29.	2.176	2.041	0.135	2.121	0.055
30.	2.176	1.847	0.329	1.914	0.262
31.	1.991	2.250	-0.259	2.342	-0.351
32.	2.628	2.181	0.447	2.268	0.360
33.	2.686	2.136	0.550	2.219	0.467

(Table 8. Contd….)

Compd. No.	logKi (hCAII)				
	Equation-2		Equation-4		
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
34.	1.708	1.541	0.167	1.558	0.150
35.	0.903	1.681	-0.778	1.717	-0.814
36.	0.301	0.723	-0.422	0.700	-0.399
37.	0.301	0.745	-0.444	0.734	-0.433
38.	0.477	0.646	-0.169	0.639	-0.162
39.	0.301	0.223	0.078	0.117	0.184
40.	0.602	0.294	0.308	0.168	0.434
41.	1.176	1.672	-0.496	1.633	-0.457
42.	1.301	1.408	-0.107	1.372	-0.071
43.	1.301	1.493	-0.192	1.543	-0.242
44.	-0.301	-0.418	0.117	-0.511	0.210
45.	-0.301	-0.240	-0.061	-0.317	0.016
46.	0.000	0.190	-0.190	0.136	-0.136
47.	2.663	2.292	0.371	2.395	0.268
48.	2.585	2.070	0.515	2.158	0.427
49.	1.380	1.134	0.246	1.214	0.166
50.	1.000	1.183	-0.183	1.248	-0.248
51.	1.000	1.224	-0.224	1.274	-0.274
52.	1.204	1.128	0.076	1.184	0.020
53.	1.176	0.975	0.201	1.015	0.161
54.	1.176	0.801	0.375	0.835	0.341
55.	0.954	1.146	-0.192	1.197	-0.243
56.	2.041	1.146	0.895	1.197	0.844
57.	2.097	1.146	0.951	1.197	0.900
58.	1.176	0.405	0.771	0.408	0.768
59.	0.699	0.553	0.146	0.569	0.130
60.	-0.523	-0.231	-0.292	-0.292	-0.231
61.	-0.523	-0.231	-0.292	-0.292	-0.231
62.	-0.398	-0.216	-0.182	-0.233	-0.165
63.	0.000	0.131	-0.131	0.111	-0.111
64.	0.176	0.222	-0.046	0.193	-0.017
65.	0.903	1.447	-0.544	1.490	-0.587
66.	0.903	1.196	-0.293	1.241	-0.338
67.	1.041	1.241	-0.200	1.344	-0.303

(Table 8. Contd….)

Compd. No.	**logKi (hCAII)**				
	Equation-2		**Equation-4**		
	Actual	***Est.***	***Res.***	***Est.***	***Res.***
68.	-0.699	-0.392	-0.307	-0.421	-0.278
69.	-0.523	-0.233	-0.29	-0.247	-0.276
70.	-0.301	0.128	-0.429	0.144	-0.445
71.	1.602	1.178	0.424	1.236	0.366
72.	1.544	1.126	0.418	1.197	0.347
73.	1.544	1.852	-0.308	1.919	-0.375
74.	1.279	1.799	-0.520	1.845	-0.566
75.	1.230	1.725	-0.495	1.750	-0.52
76.	1.362	1.471	-0.109	1.491	-0.129
77.	1.301	1.262	0.039	1.263	0.038
78.	1.230	0.948	0.282	.935	0.295
79.	1.176	1.636	-0.460	1.663	-0.487
80.	2.097	1.636	0.461	1.663	0.434
81.	2.193	1.604	0.589	1.628	0.565
82.	1.580	1.158	0.422	1.167	0.413
83.	1.079	1.257	-0.178	1.274	-0.195
84.	0.301	0.716	-0.415	0.658	-0.357
85.	0.176	0.716	-0.540	0.658	-0.482
86.	0.301	-0.078	0.379	-0.138	0.439
87.	0.903	0.252	0.651	0.202	0.701
88.	1.255	1.422	-0.167	1.427	-0.172
89.	1.556	1.208	0.348	1.217	0.339
90.	1.431	1.224	0.207	1.279	0.152
91.	-0.301	0.005	-0.306	-0.051	-0.250
92.	-0.222	0.159	-0.381	0.117	-0.339
93.	-0.155	0.112	-0.267	0.078	-0.233
94.	1.732	1.455	0.277	1.474	0.258
95.	1.699	1.206	0.493	1.228	0.471

data is predicted and the difference between observed and predicted values (activities) is calculated. The predictive ability of the model is quantified in terms of the corresponding leave-one-out cross-validated parameters. The cross-validated parameters often used being *PRESS* (Predicted residual sum of squares), *SSY* (Sum of the squares of the response value), r^2_{cv} (overall predictive ability), S_{press} or S_{cv} (uncertainty of prediction), and *PSE* or S_{pred} (predictive square error). These parameters are defined as below:

$$PRESS = \sum_y (Y_{est} - Y_{obs})^2 \quad \textbf{(5)}$$

$$SSY = \sum_y (Y_{\text{obs}} - Y_{\text{mean}})^2 \quad \textbf{(6)}$$

$$r^2_{cv} = q^2 = 1.0 - \frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{\sum_{i=1}^{n} (Y_{obs} - Y_{mean})^2} \quad (7)$$

$$S_{PRESS} = S_{CV} = \sqrt{\frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{N - M - 1}} \quad (8)$$

$$PSE = S_{pred} = \sqrt{\frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{N}} \quad (9)$$

Here, Y_{obs} and Y_{est} are the experimental and predictive values of the activity respectively. Y_{mean} is the mean value of Y_{obs}. N is the number of compounds used, M is the number of parameters (descriptors) used in the model. For a reliable model, the r^2_{cv} (or q^2) values should be > 0.6. The model is considered to be excellent if r^2_{cv} (or q^2) is $\geq$ 0.9. The actual predictive ability (predictive power) of the model is validated using an external prediction set. The performance of the model (its predictive ability or predictive power) can be given by *PSE* (or S_{pred}).

The aforementioned cross-validated parameters calculated for the models discussed above are summarized in Table **9.** The data shows that except for the 3-variable model all other models are reliable models. Also, that S_{press} (or S_{cv}) is found to be equal to *Se*. Thus, S_{press} is not a good cross-validated parameter to discuss the uncertainty in prediction. In the present situation, therefore, *PSE* (or S_{pred}) is a better parameter for investigating the predictive uncertainty of the model. The lower the value of *PSE* the better will be the predictive ability of the model. A perusal of (Table **6**) shows that *PSE* goes on decreasing as we pass from 3- to 7-variable models and that it is lowest for the modified 6-parametric model (after detecting four outliers). Hence, the most appropriate model for modeling logK_i (hCA-II) is this six-parametric model.

It is argued that *PRESS* is a good estimate of the real predictive error of the model. If *PRESS* is smaller than *SSY*, the model predicts better than chance and can be considered statistically significant. The ratio *PRESS* / *SSY* can be used to calculate approximate confidence intervals of prediction of new observations (compounds). To be a reasonable *QSAR* model, *PRESS* / SSY should be smaller than 0.4 and the value of this ratio smaller than 0.1 indicates an excellent model. A perusal of (Table **9**) shows that except for the tri-parametric model all other models have *PRESS* / *SSY* < 0.4 indicating thereby they to be reasonable models. This ratio for models 4-6 is more or less nearer to 0.1 indicating them to be appreciably good than the remaining model.

(ii) Modeling of logK_i (hCA-II) Using Quantum-Theoretical Descriptors

We now discuss the modeling of logK_i (hCA-II) using quantum-theoretical parameters. The statistical parameters and quality of correlations of variety of regressions attempted are shown in (Table **10**).

A perusal of (Table **10)** shows that : (i) statistically significant model starts pouring with two parametric regression analysis, (ii) all the regression models containing 6 or more correlating parameters have coefficients of Q_N and I_4 terms considerably smaller than their respective standard deviation. Such models are not allowed statistically. In case of model 17, one more term Q_C has coefficient smaller than its standard deviation. It means that only models 7, 8, 9 and 10 (Table **10**) are allowed statistically for modeling logK_i (hCA-II) and that the 5-parametric model 10 gives the best results:

logK_i (hCA-II) = -46.958 – 1.5160 ($\pm$ 0.2115) E_H – 25.143 ($\pm$ 4.634) Q_O

– 13.783 ($\pm$ 12.643) Q_N+ 0.872 ($\pm$ 0.140) I_1 – 0.469 ($\pm$ 0.140) I_3 **(10)**

n = 95, Se = 0.554, R = 0.887, R^2_A = 0.774, F = 65.303, Q = 1.601

We observed that all the three quantum-theoretical descriptors (E_H, Q_O, Q_N) have negative signs. It means that the decrease in the magnitude of these parameters is favorable for the exhibition of logK_i (hCA-II). In addition coefficients of the indicator parameter I_3 is also negative.

It is interesting to mention that the indicator parameter I_1 has positive coefficient.

The fact that in all the higher parametric models (with correlating parameters 6 or more) have the coefficients of Q_N and I_4 significantly smaller than their respective standard deviation means that these parameters are not good and are not favorable for modeling logK_i(hCA-II) in these higher parametric regression analysis. We have, therefore, attempted further regression analysis deleting Q_N and I_4. The results are summarized in (Table **11**).

This table shows that better results are obtained by deleting Q_N and I_4. However, the model 21 to 24 (Table 11) exhibit that the standard error estimations are higher than 0.525. Also, that R^2_A is lower than 0.797. It means that we have to make a choice in between models 19 and 20(Table 11). The model 19 (Table 11) is found as below:

logK_i (hCA-II) = -41.800 – 1.306 ($\pm$ 0.208) E_H – 32.105($\pm$ 2.182) Q_O – 0.485

($\pm$ 0.179) E_{SL} + 1.032 ($\pm$ 0.160) I_1 + 0.286 ($\pm$ 0.159) I_2

- 0.312 ($\pm$ 0.155) I_3 **(11)**

n = 95, Se = 0.531, R = 0.897, R^2_A = 0.792, F = 60.553, Q = 1.689

This model (Eq. (**11**)) (see Table **12** for observed and calculated values) consists of compounds **1, 3, 4, 17, 21** and **87** as outliers, reason being their estimated values yield residues twice the standard deviation. Deletion of these six compounds from the regression procedures (see Table **12** for observed calculated values of logKi (hCAII) using eq (**12**) yields the following regression expressions having much better statistics:

Table 9. Cross-Validated Parameters for the Five Models Mentioned in Table 6, in that Only Topological Indices are Used

Model No.	Number of parameters used	*PRESS* / *SSY*	r^2_{cv}	S_{PRESS}	PSE
1.	3 (95)	0.608	0.391	0.728	0.712
2.	4 (95)	0.350	0.650	0.605	0.590
3.	5 (95)	0.231	0.769	0.518	0.501
4.	6 (95)	0.199	0.801	0.490	0.471
5.	7 (95)	0.187	0.813	0.480	0.459
6.	6 (91)	0.147	0.853	0.435	0.417

Table 10. Statistical Parameters and Quality of Variety of Statistics Attempted for Modeling log Ki (hCAII) Using Quantum-Theoretical Descriptors

Model No.	*QTD used*	*Se*	*R*	R^2A	*F*	*Q*
7.	Q_N, I_1	0.737	0.780	0.599	71.148	1.059
8.	Q_N, I_1, E_H	0.670	0.825	0.670	64.340	1.231
9.	Q_N, I_1, E_H, Q_O	0.585	0.871	0.748	70.726	1.489
10.	Q_N, I_1, E_H, Q_O, I_3	0.554	0.887	0.774	65.303	1.601
11.	$Q_N, I_1, E_H, Q_O, I_3, E_{SL}$	0.539	0.895	0.786	58.508	1.685
12.	$Q_N, I_1, E_H, Q_O, I_3, E_{SL}, I_2, I_4$	0.533	0.898	0.790	51.694	1.685
13.	$Q_N, I_1, E_H, Q_O, I_3, E_{SH}, E_{SL}, I_2, I_4$	0.528	0.901	0.795	46.472	1.703
14.	$Q_N, I_1, E_H, Q_O, I_3, E_{SH}, E_{SL}, I_2, I_4, E_L$	0.529	0.901	0.794	41.139	1.703
15.	$Q_N, I_1, E_H, Q_O, I_3, E_{SH}, E_{SL}, I_2, I_4, E_L, Q_H$	0.530	0.902	0.793	36.848	1.701
16.	$Q_N, I_1, E_H, Q_O, I_3, E_{SH}, E_{SL}, I_2, I_4, E_L, Q_H, Q_C$	0.530	0.901	0.794	33.769	1.700
17.	$Q_N, I_1, E_H, Q_O, I_3, E_{SH}, E_{SL}, I_2, I_4, E_L, Q_H, Q_C, \phi_L$	0.533	0.904	0.794	30.597	1.696

Table 11. Modified Regression Analysis by Deleting Q_N and I_4 and Using Quantum - Theoretical Descriptors

Model No.	*QTD* used	*Se*	*R*	R^2A	*F*	*Q*
18.	E_H, Q_O, E_{SL}, I_1, I_3	0.538	0.893	0.786	70.247	1.660
19.	$E_H, Q_O, E_{SL}, I_1, I_2, I_3$	0.531	0.897	0.792	60.553	1.690
20.	$E_H, Q_O, E_{SH}, E_{SL}, I_1, I_2, I_3$	0.525	0.901	0.797	53.579	1.716
21.	$E_H, Q_O, E_{SH}, E_{SL}, E_L, I_1, I_2, I_3$	0.526	0.902	0.795	46.693	1.714
22.	$E_H, Q_O, E_{SH}, E_{SL}, E_L, Q_H, I_1, I_2, I_3$	0.528	0.902	0.794	41.313	1.708
23.	$E_H, Q_O, E_{SH}, E_{SL}, E_L, Q_H, Q_C, I_1, I_2, I_3$	0.529	0.903	0.794	37.123	1.706
24.	$E_H, Q_O, E_{SH}, E_{SL}, E_L, Q_H, Q_C, \phi_L, I_1, I_2, I_3$	0.529	0.902	0.792	34.583	1.705

Table 12. Observed and Calculated logKi (hCAII) Using Esq. (12) and (15)

Compd. No.	logKi (hCAII)				
	Equation-12			Equation-15	
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
1.	4.311	-	-	3.754	0.557
2.	4.272	3.334	0.938	3.876	0.396
3.	4.037	-	-	3.314	0.723
4.	4.170	-	-	-	-
5.	3.699	3.086	0.613	3.292	0.407
6.	2.778	2.902	-0.124	2.847	-0.069
7.	2.699	2.560	0.139	3.432	-0.733
8.	2.863	2.781	0.082	3.415	-0.552
9.	3.017	2.753	0.264	3.330	-0.313
10.	2.633	2.549	0.084	3.032	-0.399
11.	1.954	2.453	-0.499	-	-
12.	2.000	1.125	0.875	1.455	0.545
13.	1.380	1.005	0.375	1.273	0.107
14.	1.114	1.258	-0.144	1.135	-0.021
15.	0.477	0.633	-0.156	0.315	0.162
16.	0.699	0.729	-0.030	-	-
17.	1.322	-	-	1.298	0.024
18.	1.362	2.203	-0.841	1.227	0.135
19.	1.398	2.022	-0.624	1.540	-0.142
20.	-0.046	0.847	-0.893	0.546	-0.592
21.	-0.046	-	-	-	-
22.	0.000	0.482	-0.482	0.163	-0.163
23.	3.708	3.084	0.624	3.487	0.221
24.	2.740	2.941	-0.201	3.032	-0.292
25.	2.398	2.273	0.125	2.389	0.009
26.	2.230	2.213	0.017	2.311	-0.081
27.	2.204	2.002	0.202	2.199	0.005
28.	2.255	1.775	0.480	2.029	0.226
29.	2.176	2.421	-0.245	2.315	-0.139
30.	2.176	2.264	-0.088	2.102	0.074
31.	1.991	1.904	0.087	2.116	-0.125
32.	2.628	2.085	0.543	2.127	0.501
33.	2.686	2.067	0.619	2.074	0.612
34.	1.708	1.855	-0.147	1.652	0.056

(Table 12. Contd….)

Compd. No.	logKi (hCAII)				
	Equation-12		Equation-15		
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
35.	0.903	1.733	-0.83	1.728	-0.825
36.	0.301	0.333	-0.032	0.699	-0.398
37.	0.301	0.417	-0.116	0.606	-0.305
38.	0.477	0.638	-0.161	0.652	-0.175
39.	0.301	-0.044	0.345	0.270	0.031
40.	0.602	0.073	0.529	0.065	0.537
41.	1.176	1.662	-0.486	1.721	-0.545
42.	1.301	1.570	-0.269	1.556	-0.255
43.	1.301	1.364	-0.063	1.262	0.039
44.	-0.301	0.190	-0.491	-0.510	0.209
45.	-0.301	0.438	-0.739	-0.210	-0.091
46.	0.000	-0.210	0.210	-0.205	0.205
47.	2.663	2.457	0.206	2.512	0.151
48.	2.585	2.281	0.304	2.273	0.312
49.	1.380	1.650	-0.270	1.484	-0.104
50.	1.000	1.270	-0.270	1.268	-0.268
51.	1.000	1.260	-0.260	1.238	-0.238
52.	1.204	1.007	0.197	0.994	0.210
53.	1.176	1.765	-0.589	1.433	-0.257
54.	1.176	1.622	-0.446	1.229	-0.053
55.	0.954	1.172	-0.218	1.149	-0.195
56.	2.041	1.343	0.698	1.189	0.852
57.	2.097	1.335	0.762	-	-
58.	1.176	0.955	0.221	0.509	0.667
59.	0.699	0.986	-0.287	0.618	0.081
60.	-0.523	-0.783	0.260	-0.281	-0.242
61.	-0.523	-0.208	-0.315	-0.349	-0.174
62.	-0.398	0.039	-0.437	-0.100	-0.298
63.	0.000	-0.605	0.605	0.022	-0.022
64.	0.176	0.011	0.165	0.241	-0.065
65.	0.903	1.036	-0.133	1.387	-0.484
66.	0.903	0.839	0.064	1.180	-0.277
67.	1.041	0.716	0.325	0.915	0.126
68.	-0.699	-0.681	-0.018	-0.756	0.057
69.	-0.523	-0.293	-0.230	-0.391	-0.132

(Table 12. Contd….)

Compd. No.	**logKi (hCAII)**				
	Equation-12		**Equation-15**		
	Actual	***Est.***	***Res.***	***Est.***	***Res.***
70.	-0.301	-0.799	0.498	-0.386	0.085
71.	1.602	1.764	-0.162	1.574	0.028
72.	1.544	1.704	-0.160	1.552	-0.008
73.	1.544	1.626	-0.082	1.802	-0.258
74.	1.279	1.442	-0.163	1.635	-0.356
75.	1.230	1.465	-0.235	1.599	-0.369
76.	1.362	1.235	0.127	1.297	0.065
77.	1.301	1.888	-0.587	1.605	-0.304
78.	1.230	1.813	-0.583	1.392	-0.162
79.	1.176	1.378	-0.202	1.508	-0.332
80.	2.097	1.525	0.572	1.537	0.560
81.	2.193	1.517	0.676	1.500	0.693
82.	1.580	1.321	0.259	1.127	0.453
83.	1.079	1.322	-0.243	1.186	-0.107
84.	0.301	0.744	-0.443	0.568	-0.267
85.	0.176	-0.002	0.178	0.217	-0.041
86.	0.301	0.304	-0.003	0.116	0.185
87.	0.903	-	-	0.286	0.617
88.	1.255	1.332	-0.077	1.523	-0.268
89.	1.556	1.150	0.406	1.343	0.213
90.	1.431	1.008	0.423	1.056	0.375
91.	-0.301	-0.519	0.218	-0.454	0.153
92.	-0.222	-0.051	-0.171	-0.029	-0.193
93.	-0.155	-0.562	0.407	-0.269	0.114
94.	1.732	1.966	-0.234	1.807	-0.075
95.	1.699	1.943	-0.244	1.713	-0.014

$\log K_i$ (hCA-II) = -40.663 – 1.238 ($\pm$ 0.171) E_H – 31.363($\pm$ 1.823) Q_O – 0.555

($\pm$ 0.148) E_{SL} + 0.991($\pm$ 0.139) I_1 + 0.340 ($\pm$ 0.129) I_2- 0.261($\pm$ 0.125) I_3 (12)

$n = 89$, $Se = 0.421$, $R = 0.925$, $R^2_A = 0.845$, $F = 80.909$, $Q = 2.197$

The model 20 (Table**11**) is found as below:

$\log K_i$ (hCA-II) = -45.321 – 1.237 ($\pm$ 0.209) E_H – 34.644($\pm$ 2.598) Q_O – 0.533

($\pm$ 0.179) E_{SL} - 0.167 ($\pm$ 0.095) E_{SH} + 1.035 ($\pm$ 0.158) I_1 + 0.286

($\pm$ 0.157) I_2 - 0.300 ($\pm$ 0.154) I_3 (**13**)

$n = 95$, $Se = 0.525$, $R = 0.901$, $R^2_A = 0.797$, $F = 53.579$, $Q = 1.716$

This model (eq. (**13**)) appears to be better than the model (eq. (**11**)) discussed above. However, the final conclusion could only be made by making residual analysis. This is done by estimating $\log K_i$ (hCA-II) from eq. (**13**). Upon doing this it was observed that it contains compounds **1, 3, 4,**

17 and **21** as outliers. The deletion of these outliers gave the following expression:

$\log K_i$ (hCA-II) = -42.466 – 1.233 (± 0.178) E_H – 32.306 (± 2.229) Q_O – 0.493

(± 0.148) E_{SL} – 0.124 (± 0.080) E_{SH} + 0.947 (± 0.140) I_1 + 0.280

(± 0.130) I_2 -0.308 (± 0.127) I_3 (**14**)

$n = 90$, $Se = 0.432$, $R = 0.921$, $R^2_A = 0.835$, $F = 65.501$, $Q = 2.312$

We observed that in no way this model (eq. (**14**)) is statistically better than the model expressed by eq. (**12**). This clearly means that the five parametric model 19 (Table **11**) is the most appropriate for modeling $\log K_i$ (hCA-II), that too, by using quantum theoretical descriptors.

(iii) Modeling of logK$_i$ (hCA-II) Using Combinations of Topological Indices and Quantum-Theoretical Descriptors

In accordance with the objective of the present investigation we now investigate the models obtained using different combinations of topological indices and quantum-theoretical descriptors. Out of the several regressions attempted, the best model under this particular category is given in (Table **13**).

The data presented in (Table **13**) show that we have to make a choice among models 37, 38 and 39 (Table **13**) and decide the most appropriate model for modeling $\log K_i$ (hCA-II) using the combinations of topological indices and quantum-theoretical descriptors. The 9-parametric model 37 (Table **13)** is found as below:

$\log K_i$ (hCA-II) = -18.489 + 8.873 × 10^{-4} (± 3.969 x 10^{-4}) *W*- 0.824 (± 0.197) *Jhetm*

+ 2.616 (± 0.428) *Jhete*- 7.148 × 10^{-3} (± 2.865 × 10^{-3}) *BAC* – 0.670

(± 0.135) $^1\chi$+ 6.000 × 10^{-4} (± 3.046 ×10^{-4}) *Sz*-0.656 (± 0.205) E_H

– 15.700 (± 2.692) Q_O+ 0.217 (± 0.181) I_1 (**15**)

$n = 95$, $Se = 0.416$, $R = 0.940$, $R^2_A = 0.8722$, $F = 72.279$, $Q = 2.260$

The other two models 38 and 39 (Table **13**) containing 10- and 11-correlating parameters are respectively found as below:

(a) 10-Parametric Model

$\log K_i$ (hCA-II) = -20.971 + 5.617 ×10^{-4} (± 4.795 × 10^{-4}) *W*- 0.884 (± 0.204) *Jhetm*

+ 2.786 (± 0.453) *Jhete* - 6.817 × 10^{-3} (± 2.876 × 10^{-3}) *BAC* – 0.622

(± 0.142) $^1\chi$ + 7.780 × 10^{-4} (± 3.424 × 10^{-4}) *Sz*- 0.691 (± 0.207) E_H

– 16.610 (± 2.805) Q_O+ 0.318 (± 0.201) I_1 + 0.011 (± 9.652 × 10^{-3}) I_2 (**16**)

$n = 95$, $Se = 0.415$, $R = 0.941$, $R^2_A = 0.873$, $F = 63.394$, $Q = 2.268$

Table 13. Regression Parameters and Quality of Correlation for Modeling log Ki (hCAII) Using Models with Combinations Topological Indices and Quantum-Theoretical Descriptors

Model No.	*TI* and *QTD* used	*Se*	*R*	R^2A	*F*	*Q*
25.	$^1\chi$, Q_N	0.706	0.800	0.631	81.484	1.133
26.	$^1\chi$, Q_N, *W*	0.561	0.880	0.768	104.492	1.569
27.	$^1\chi$, Q_N, *W, J*	0.501	0.906	0.814	103.848	1.808
28.	$^1\chi$, Q_N, *W, J, BAC*	0.481	0.915	0.829	92.087	1.902
29.	$^1\chi$, Q_N, *W, J, BAC*, Q_O	0.473	0.919	0.835	80.114	1.943
30.	$^1\chi$, Q_N, *W, J, BAC*, Q_O, E_H	0.463	0.924	0.841	72.440	1.996
31.	$^1\chi$, Q_N, *W, J, BAC*, Q_O, E_H, E_L	0.453	0.929	0.849	66.935	2.050
32.	$^1\chi$, Q_N, *W, J, BAC*, Q_O, E_H, E_L, *J*hete	0.443	0.933	0.856	62.770	2.106
33.	$^1\chi$, Q_N, *W, J, BAC*, Q_O, E_H, E_L, Jhete	0.440	0.933	0.857	71.429	2.120
34.	$^1\chi$, Q_N, *W, BAC*, Q_O, E_H, E_L, *J*hete	0.421	0.939	0.869	70.071	2.230
35.	$^1\chi$, Q_N, *W, BAC,* Q_O, *J*hetm, E_H, E_L, *J*hete	0.419	0.940	0.870	64.030	2.244
36.	$^1\chi$, Q_N, *W, BAC*, Q_O, *J*hetm, *J*hete, *Sz*, E_H, E_L,	0.417	0.940	0.871	71.749	2.255
37.	*W*, Jhetm, Jhete, *BAC,* $^1\chi$, *Sz*, E_H, E_L, Q_O	0.416	0.940	0.873	72.279	2.260
38.	*W*, Jhetm, Jhete, *BAC,* $^1\chi$, *Sz*, E_H, Q_O, I_1	0.415	0.941	0.873	65.344	2.268
39.	*W*, *J*hetm, *J*hete, *J*hetp, *BAC,* $^1\chi$, *Sz*, ϕ_L, E_H, Q_O, I_1	0.414	0.943	0.873	59.801	2.278

(b) 11-Parametric Model

$\log K_i$ (hCA-II) = -22.599 + 2.947 × 10^{-4} (± 5.318 × 10^{-4}) *W* - 0.796 (± 0.217) *Jhetm*

+ 2.916 (± 0.466) *J*hete - 0.409 (± 0.355) *J*hetp + 9.234 × 10^{-4}

(3.643 × 10^{-3}) *BAC* – 0.590 (± 0.144) $^1\chi$ + 9.234 × 10^{-4}

(± 3.643 × 10^{-3}) *Sz* - 0.775 (± 0.219) E_H – 17.315 (± 2.866) Q_O

+ 0.413 (± 0.217) I_1 + 0.013 (± 9.799 × 10^{-3}) **(17)**

$n = 95$, $Se = 0.415$, $R = 0.942$, $R^2_A = 0.873$, $F = 59.801$, $Q = 2.270$

The observation in favor of models expressed by eqs. (**16**) and (**17**) is that they have four (**4, 11, 16, 21**) and three (**4, 16, 21**) outliers respectively. While the model expressed by Eq. (**15**), five-compounds (**4, 11, 16, 21, 57**) (see Table **14** for observed and calculated values of log Ki (hCAII) using these (esq. (**16**) and (**17**)). It means examination of residual statistics will help in deciding which model is most appropriate for modeling $\log K_i$ (hCA-II). The residual statistics, and deleting the outliers gave *R* = 0.955, 0.953 and 0.950, respectively for 9-, 10- and 11-parametric regression expressions (Table **13**) (see Figs. **1-3**).

That is, more or less these three models are similar and thus the choice among them is the model which contains fewer numbers of correlating parameters. This means that the most appropriate model for modeling $\log K_i$ (hCA-II) is the 9-parametric model, which of course uses some combination of topological and quantum-theoretical descriptors.

Comment on Adjustable R^2 (R^2_A) (see Tables (6, 10, 13))

The adjustable-R^2 (R^2_A) takes into account of adjacement of R^2 and is given by following expression:

R^2_A=1-(1-R^2) (*n*-1 / *n*-k-1) **(18)**

If a variable is added that does not contribute its fair share, then R^2_A will actually decline. This parameter R^2_A is particularly important when the number of independent variables is larger relative to the sample size. R^2 may appear artificially high if the number of variables is high compared to the sample size. In fact, R^2 will always increase when an independent variable is added, while R^2_A will decrease if the added variable does not reduce the unexplained variation enough to affect the loss of degrees of freedom.

PROBLEM OF CO- LINEARITY AND RANDIC RECOMMENDATIONS

The problem of co- linearity can be resolved in two different ways: (i) applying pure statistics and forgetting the possible physical significances of the parameter involved in the model or (ii) do not entirely depend on the statistics and use Randic recommendations.

The first approach uses the results obtained from (i) correlation matrix; (ii) Ridge statistics, (iii) λ-statistics.

(i) Correlation Matrix

In order to investigate co-linearity problem in the proposed models we have to first obtained correlation matrix for the best model in modeling log K_i (hCAII) activity (Table **15**).

Fortunately, we obtain a best model which contains *W, Jhetm, Jhete, BAC,* $^1\chi$, *Sz*, E_H, Q_O, E_{SL}, E_{SH}, I_1, I_2, and I_3 as the correlating parameters. This situation has an additional advantage that using such models containing common correlating parameters we can study relative potential of these indices in modeling the referred three activities. It is worthy to mention that the correlation matrix is very useful for determining which independent variables are likely to help explain variation in the dependent variables. Here we look the correlation close to ±1.0 since that indicates changes in the independent variables are linearly related to changes in the dependent variables. We can also use correlation matrix to determine the extent to which independent variables are correlated with one another i.e. their inter-correlated ness or auto-correlation. This can be useful in determining if certain independent variables are redundant and not needed in the model. In practice every term in the correlation matrix > 0 .4 can be taken as being suspicious due to co linearity.

A perusal of (Table **15**) shows the following:

(a) in case of esq. (1) *W,* $^1\chi$, and *Sz* are highly correlated. Similarly *J*hetm is highly correlated with *J*hete. Thus, this model expressed by eq. (**1**) suffers from co- linearity-defect;

(b) Correlation matrix involving quantum-theoretical descriptors indicates [eq. (11)] that none of the parameters used exhibit any co- linearity. That is, all the models using quantum-theoretical descriptors only will be free from co - linearity defect. Thus, the model expressed by eq. (**13**) again is free from such defect;

(c) Finally eq. (**15**) considered the combination of topological and quantum-theoretical descriptors its correlation matrix shows that topological indices *W,* $^1\chi$ and *Sz* are highly linearly correlated. Same is the case with *J*hetm and *J*hete. Thus, like eq. (**1**) this model also suffers from co- linearity defect.

In view of this we discuss below Ridge and λ-statistics to investigate further the co –linearity problem. Finally, we will use Randic recommendation for making finial conclusion.

(ii) Ridge Statistics

Application of Ridge statistics provides important statistical parameters namely variance inflation factors (*VIF*s) for each of the parameters involved in the model. The *VIF* is defined for each variable in the equation, and not for the equation as a whole, so there should be as many *VIF*s, as there are correlating parameters. The *VIF* is defined as:

VIF = 1(1-R^2_i) **(19)**

Where R_i is the multiple correlation coefficient of the i^{th} independent variable on all of the other independent variables. In the proposed models, all these *VIF*s should be less than 10 indicating that no co linearity problem exists in the model.

The *VIF*s values for the parameters involved in models 15) are given in (Table **16**). The *VIF*s for the parameters involved in this model indicate a major problem of co- linear-

Table 14. Observed and Calculated logKi (hCAII) Using Esq. (16) and (17)

Compd. No.	logKi (hCAII)				
	Equation-16			Equation-17	
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
1.	4.311	3.670	0.641	3.538	0.773
2.	4.272	3.953	0.319	3.938	0.334
3.	4.037	3.302	0.735	3.226	0.811
4.	4.170	-	-	-	-
5.	3.699	3.260	0.439	3.174	0.525
6.	2.778	2.820	-0.042	2.705	0.073
7.	2.699	3.401	-0.702	3.333	-0.634
8.	2.863	3.419	-0.556	3.313	-0.450
9.	3.017	3.339	-0.322	3.230	-0.213
10.	2.633	2.982	-0.349	2.666	-0.033
11.	1.954	-	-	2.714	-0.760
12.	2.000	1.474	0.526	1.590	0.410
13.	1.380	1.271	0.109	1.371	0.009
14.	1.114	1.158	-0.044	1.147	-0.033
15.	0.477	0.370	0.107	0.298	0.179
16.	0.699	-	-	-	-
17.	1.322	1.380	-0.058	1.360	-0.038
18.	1.362	1.273	0.089	1.215	0.147
19.	1.398	1.577	-0.179	1.653	-0.255
20.	-0.046	0.538	-0.584	0.486	-0.532
21.	-0.046	-	-	-	-
22.	0.000	0.112	-0.112	0.215	-0.215
23.	3.708	3.457	0.251	3.521	0.187
24.	2.740	3.011	-0.271	3.027	-0.287
25.	2.398	2.344	0.054	2.250	0.148
26.	2.230	2.408	-0.178	2.400	-0.170
27.	2.204	2.262	-0.058	2.249	-0.045
28.	2.255	2.055	0.200	2.166	0.089
29.	2.176	2.322	-0.146	2.276	-0.100
30.	2.176	2.125	0.051	2.047	0.129
31.	1.991	2.141	-0.150	2.149	-0.158
32.	2.628	2.181	0.447	2.176	0.452
33.	2.686	2.132	0.554	2.126	0.560
34.	1.708	1.636	0.072	1.625	0.083

(Table 14. Contd….)

Compd. No.	**logKi (hCAII)**				
	Equation-16		**Equation-17**		
	Actual	***Est.***	***Res.***	***Est.***	***Res.***
35.	0.903	1.690	-0.787	1.641	-0.738
36.	0.301	0.756	-0.455	0.773	-0.472
37.	0.301	0.670	-0.369	0.655	-0.354
38.	0.477	0.639	-0.162	0.543	-0.066
39.	0.301	0.266	0.035	0.271	0.030
40.	0.602	0.017	0.585	-0.003	0.605
41.	1.176	1.734	-0.558	1.794	-0.618
42.	1.301	1.538	-0.237	1.567	-0.266
43.	1.301	1.269	0.032	1.354	-0.053
44.	-0.301	-0.499	0.198	-0.463	0.162
45.	-0.301	-0.193	-0.108	-0.035	-0.266
46.	0.000	-0.318	0.318	-0.254	0.254
47.	2.663	2.519	0.144	2.596	0.067
48.	2.585	2.267	0.318	2.285	0.300
49.	1.380	1.460	-0.080	1.367	0.013
50.	1.000	1.284	-0.284	1.213	-0.213
51.	1.000	1.302	-0.302	1.275	-0.275
52.	1.204	1.015	0.189	1.105	0.099
53.	1.176	1.482	-0.306	1.463	-0.287
54.	1.176	1.251	-0.075	1.192	-0.016
55.	0.954	1.201	-0.247	1.199	-0.245
56.	2.041	1.278	0.763	1.299	0.742
57.	2.097	1.283	0.814	1.312	0.785
58.	1.176	0.502	0.674	0.451	0.725
59.	0.699	0.581	0.118	0.496	0.203
60.	-0.523	-0.254	-0.269	-0.297	-0.226
61.	-0.523	-0.320	-0.203	-0.368	-0.155
62.	-0.398	-0.099	-0.299	-0.178	-0.220
63.	0.000	0.081	-0.081	0.028	-0.028
64.	0.176	0.245	-0.069	0.185	-0.009
65.	0.903	1.442	-0.539	1.442	-0.539
66.	0.903	1.143	-0.240	1.077	-0.174
67.	1.041	0.963	0.078	1.001	0.040
68.	-0.699	-0.730	0.031	-0.806	0.107

(Table 14. Contd….)

Compd. No.	logKi (hCAII)				
	Equation-16		Equation-17		
	Actual	*Est.*	*Res.*	*Est.*	*Res.*
69.	-0.523	-0.368	-0.155	-0.316	-0.207
70.	-0.301	-0.461	0.160	-0.460	0.159
71.	1.602	1.619	-0.017	1.711	-0.109
72.	1.544	1.524	0.02	1.674	-0.130
73.	1.544	1.802	-0.258	1.763	-0.219
74.	1.279	1.613	-0.334	1.572	-0.293
75.	1.230	1.632	-0.402	1.643	-0.413
76.	1.362	1.312	0.05	1.412	-0.050
77.	1.301	1.627	-0.326	1.640	-0.339
78.	1.230	1.364	-0.134	1.328	-0.098
79.	1.176	1.525	-0.349	1.525	-0.349
80.	2.097	1.576	0.521	1.589	0.508
81.	2.193	1.542	0.651	1.553	0.640
82.	1.580	1.125	0.455	1.115	0.465
83.	1.079	1.158	-0.079	1.119	-0.040
84.	0.301	0.634	-0.333	0.664	-0.363
85.	0.176	0.269	-0.093	0.259	-0.083
86.	0.301	0.100	0.201	0.034	0.267
87.	0.903	0.304	0.599	0.240	0.663
88.	1.255	1.562	-0.307	1.560	-0.305
89.	1.556	1.349	0.207	1.311	0.245
90.	1.431	1.085	0.346	1.120	0.311
91.	-0.301	-0.437	0.136	-0.490	0.189
92.	-0.222	-0.041	-0.181	0.019	-0.241
93.	-0.155	-0.357	0.202	-0.367	0.212
94.	1.732	1.788	-0.056	1.892	-0.160
95.	1.699	1.697	0.002	1.861	-0.162

Table 15. Correlation Matrix for esq. (15)

	logKi(hCAII)	**W**	***J*hetm**	***J*hete**	***BAC***	$^{1}\chi$	***Sz***	E_H	Q_O	I_1
logKi (hCAII)	1.000									
W	-0.402	1.000								
Jhetm	0.451	-0.431	1.000							

(Table 15. Contd….)

	logKi(hCAII)	W	Jhetm	Jhete	*BAC*	$^1\chi$	*Sz*	E_H	Q_O	I_1
Jhete	0.577	-0.444	0.937	1.000						
BAC	-0.070	0.543	0.375	0.436	1.000					
$^1\chi$	-0.502	0.961	-0.490	-0.467	0.537	1.000				
Sz	-0.400	0.985	-0.506	-0.521	0.437	0.960	1.000			
E_H	-0.375	0.440	-0.663	-0.680	-0.206	0.440	0.496	1.000		
Q_O	-0.607	0.031	0.185	-0.016	0.170	0.038	-0.029	-0.190	1.000	
I_1	0.422	-0.530	0.380	0.258	-0.413	-0.674	-0.525	-0.173	0.048	1.000

Table 16. VIFs Values of Parameters Involved in esq. (15)

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
W	93.2853	0.9893	0.0107
Jhetm	21.0333	0.9525	0.0475
Jhete	39.2594	0.9745	0.0255
BAC	21.0189	0.9524	0.0476
$^1\chi$	49.0613	0.9796	0.0204
Sz	104.8066	0.9905	0.0095
E_H	2.2129	0.5481	0.4519
Q_O	3.0344	0.6704	0.3296
I_1	3.3820	0.7043	0.2957

ity as sime of the parameters have *VIF values* much more than 10 rangeing from 21 to 104. Following Ridge statistics as a mathematical tool and forgetting Randic recommendations; in both the cases the parameters having *VIF*>10 need to be deleted. Thus, looking to the *VIF*s requirement advised us to remove *W*, Jhetm, Jhete, BAC, $^1\chi$ and *Sz* from eq. (**15**). However, like above, we have deleted *W, Sz* and Jhete from eq. (**15**). This yielded *VIF*s values of 6.2371, 6.4168, 7.2575, 2.0922, 1.0937 and 2.1733 respectively for the parameters Jhetm, *BAC*, $^1\chi$, E_H, Q_O, I_1. Thus, in the new model all the involved parameters have *VIF*s significantly smaller than 10, thus, showing absence of co linearity defects. However, this resulting model has quite an inferior statistics compared to the original model expressed by eq. (**15**) (R = 0.8963, Se = 0.3954). At this stage it is interesting to consider the results of Shapiro and Guggenheim [58] who reported *VIF*s for the inhibition of oral bacteria by phenolic compounds using molecular connectivity as the correlating parameters. In this paper in two of their proposed models they obtained VIFs values as high as 313 / 362. Also, for one of the best model, quadratic in $^1\chi^v$ they obtained *VIF* values of 25.4 for both the terms, that is, for $^1\chi^v$ and $(^1\chi^v)^2$. Furthermore, the Ridge statistics as discussed above indicated that there is no need to remove all such parameters having *VIF*s > 10 from the model but that only removal of a couple of parameters out of several parameters having *VIF*s > 10 yielded a new model in that *VIF*s of the remaining parameters involved have values < 10. Hence, we can argue that the statistical requirement of *VIF* smaller than 10 is tentative. Therefore, in this connection we have to seriously consider Randic recommendations [59,60].

It is interesting to mention that in case of eq. (**11**), the deletion of *W* and *Sz* along as mentioned above, yield from-parametric model as below:

$$\text{logKi (hCAII)} = -5.951 - 0.901(\pm 0.238)\ \text{Jhetm} + 3.442\ (\pm 0.473)\ \text{Jhete} - 1.391\times 10^{-2}\ (\pm 3.599\times 10^{-2}) BAC + 0.224\ (\pm 0.114)\ {}^1\chi \qquad (\mathbf{20})$$

n = 95, Se = 0.580, R = 0.753, R^2A = 0.548, F = 29.429, Q = 1.300

The *VIF*s and Eigen values for the parameters involved in this model (eq. (**20**)) are given in (Tables **17** – **19**) respectively. Both the data shows that for the new model expressed by eq. (**20**) multicolinearity is not a problem.

In case eq. (**15**), we have only deleted *W* and *Sz* parameters. However, the resulting model did not show significant improved.

Table 17. Eigen Values of the Parameters Involved in esq. (15)

S. No.	Eigen value	Incremental Percent	Cumulative Percent	Condition Number
1.	4.5423	50.47	50.47	1.00
2.	2.3093	25.66	76.13	1.97
3.	1.0058	11.17	87.30	4.52
4.	0.6601	7.33	94.64	6.88
5.	0.3537	3.93	98.57	12.85
6.	0.0876	0.97	99.54	51.90
7.	0.0230	0.26	99.79	197.58
8.	0.0136	0.15	99.95	334.60
9.	0.0050	0.05	100.00	919.88

Some Condition Number greater than 100. Multicollinearity is a MILD problem.

Table 18. VIFs Values of Parameters Involved in eq. (20)

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
*J*hetm	8.6282	0.8841	0.1159
*J*hete	13.5308	0.9261	0.0739
BAC	9.3600	0.8932	0.1068
$^1\chi$	9.8729	0.8987	0.1013

Since some VIFs are greater than 10, multicollinearity is a problem.

Further, deletion of *J*hete yielded six parametric regressions expressing in that all the involved parameters have *VIF*s appreciable < 10. Thus, model is found as below:

$$\log Ki\ (hCAII) = -36.933 + 0.217(\pm 0.138)\ Jhetm + 2.294\times10^{-3}(\pm 2.030\times10^{-3})\ BAC - 7.717\times10^{-2}(\pm 6.668\times10^{-2})\ {}^{1}\chi - 0.906(\pm 0.256)E_H - 30.168(\pm 2.072)Q_O + 0.799(\pm 0.186)I_1 \quad (21)$$

$n = 95$, $Se = 0.395$, $R = 0.896$, $R^2A = 0.790$, $F = 59.890$, $Q = 2.269$

The *VIF*s and Eigen values for the parameters involved in the above model (eq. (**21**)) are shown in (Table **20** and **21)** respectively. Both these Tables exhibit that now the resulting model is free from multi-co- linearity problem.

These results further support our view that the results obtain on the basis of *VIF*s values is tentative. The simple reason being there is no need to delete all the parameters whose VIF are > 10.The Ridge tracks (Fig. (**4**) for (eq. (**20**), (**21**), (**11**) and (**13**)) shows that the models based on these equations are free from multicolinearity defect (see Ridge traces in Fig. **4,5** for details).

(iii) λ-Statistics

The Ridge regression analysis also provides -statistics helping us to resolve the problem of co linearity. The -statistics is defined as below:

Table 19. Eigen Values of the Parameters Involved in esq. (20)

S. No.	Eigen value	Incremental Percent	Cumulative Percent	Condition Number
1.	2.3567	58.91	58.91	1.00
2.	1.5331	38.33	97.24	1.54
3.	0.0727	1.82	99.06	32.46
4.	0.0377	0.94	100.00	62.55

All Condition Number less than 100. Multicollinearity is NOT a problem.

Table 20. VIFs Values of Parameters Involved in esq. (21)

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
*J*hetm	6.2371	0.8397	0.1603
BAC	6.4168	0.8442	0.1558
$^{1}\chi$	7.2525	0.8621	0.1379
E_H	2.0922	0.5220	0.4780
Q_O	1.0937	0.0856	0.9144
I_1	2.1733	0.5399	0.4601

Since some VIFs are less than 10, multicollinearity is NOT a problem.

Table 21. Eigen Values of the Parameters Involved in esq. (21)

S. No.	Eigen value	Incremental Percent	Cumulative Percent	Condition Number
1.	2.4687	41.14	41.14	1.00
2.	1.8000	30.00	71.14	1.37
3.	0.9031	15.05	86.20	2.73
4.	0.5353	8.94	95.13	4.60
5.	0.2333	3.89	99.02	10.58
6.	0.0587	0.98	100.00	42.08

All Condition Number less than 100. Multicollinearity is NOT a problem.

$$\lambda = 1/n \sum_{i=1} 1/\lambda_i \quad \textbf{(22)}$$

Where n is the number of variables in the model (regression expression) and λ_i is the eigen-values of the correlation matrix of the independent variables.

If $\lambda < 5.0$, the sub-set is considered free from co linearity problem, and the equation (model) is accepted. If λ is not < 5.0, then eigen-vector matrix is examined. The eigen-values presented in (Table **14**) directly indicates that the models expressed by eq. (**1**) and (**15**) have some condition numbers greater than 100 and, therefore, for them multi-co linearity is a mild problem. In case of eq. (**11**) and (**13**) we observed that all condition numbers are less than 100, therefore, for them multicollinearity is not a problem. We observed that the results obtained from λ-statistics are slightly different from the results obtained from λ-statistics are slightly different from the results obtained from VIFs values. Hence, it becomes absolute necessity to use Randic recommendations to resolve finally the co linearity problem. In view of this below we discuss Ridge and λ-statistics to investigate further the co-linearity problem. Finally, we will Randic recommendation for interesting such defect.

(iv) Randic Recommendations

Randic [59,60] stated that if a descriptor strongly correlates with another descriptor already used in a regression, such a descriptor in most studies should be discarded. For example $^{1}\chi$ and $^{2}\chi$, $^{1}\chi$ often strongly correlate and in many structure-property-activity studies $^{2}\chi$ has been discarded. This is not theoretically justified and despite the widespread practice should be stopped. Although two highly correlated descriptors overall depict the same features of molecular structure, it is important to recognize that even highly inter-related descriptors differ in some other structural traits. The difference between them may be relatively small but nevertheless very important for structure-property regression.

The criteria for inclusion or exclusion of descriptors should not be based on parallelism between descriptors even if overwhelming, but should be based on whether the part in which two descriptors disagree is or is not relevant for the characterization of the property considered .If the part in which the second descriptor differ from the first, regardless of how small it is, is relevant for the property under consideration, then the descriptor should be included. Randic [59,60] further stated that the selection of descriptors to be used in structure-property-activity studies should not be delegated solely to computers, although statistical criteria will continue to be useful for preliminary screening of descriptors taken from a large pool. Often in an automated selection of descriptors, a descriptor will be discarded because it is highly correlated with another descriptor already selected. But what is important is not whether two descriptors parallel one another; i. e. duplicates much of the same structural information, but whether they are complementary in those parts that are important for structure-property-activity correlations. Hence, the residual of the correlation between two descriptors should be examined and kept or discarded

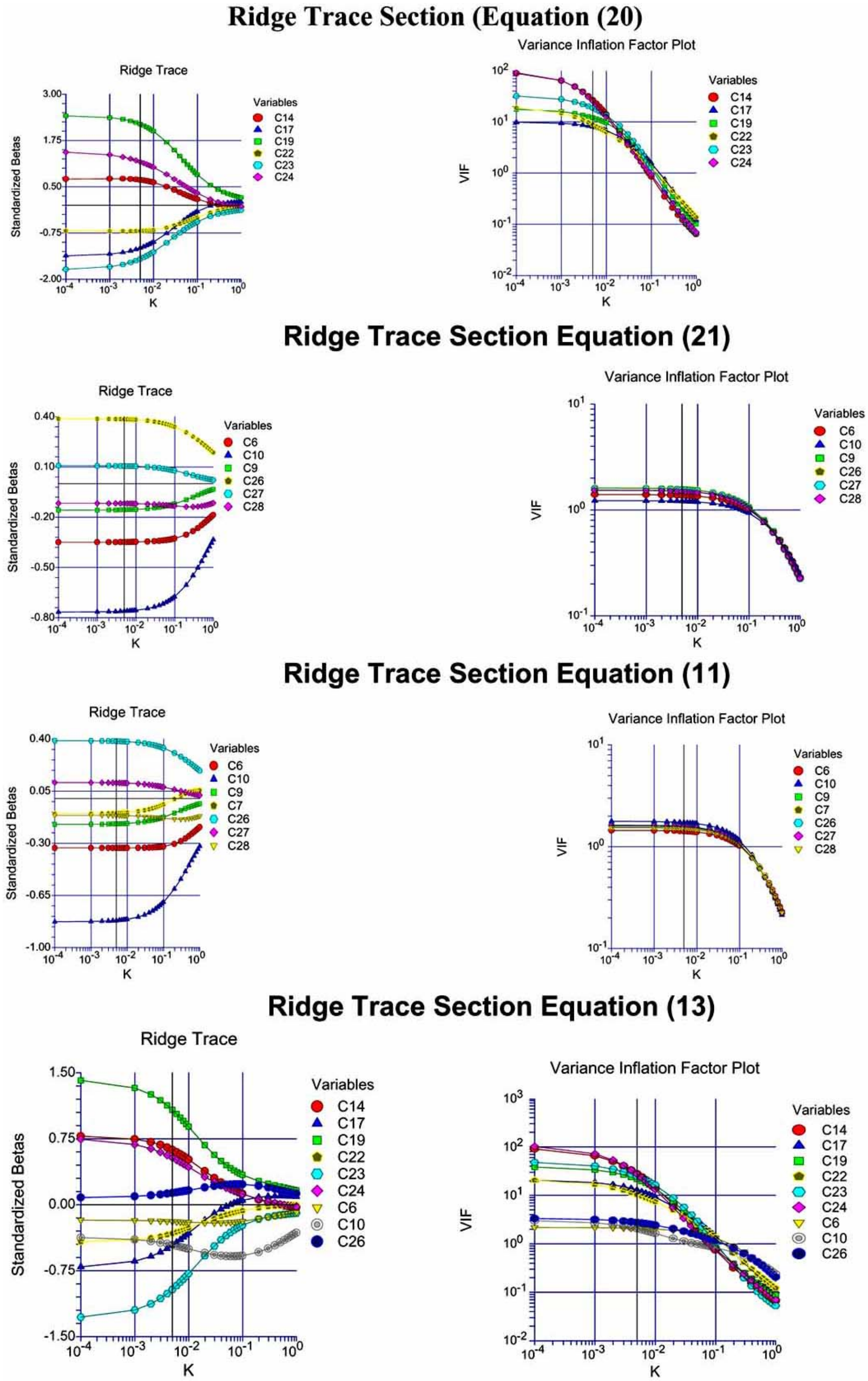


Fig. (4). Ridge traces for eqs. (20), (21), (11) and (13).

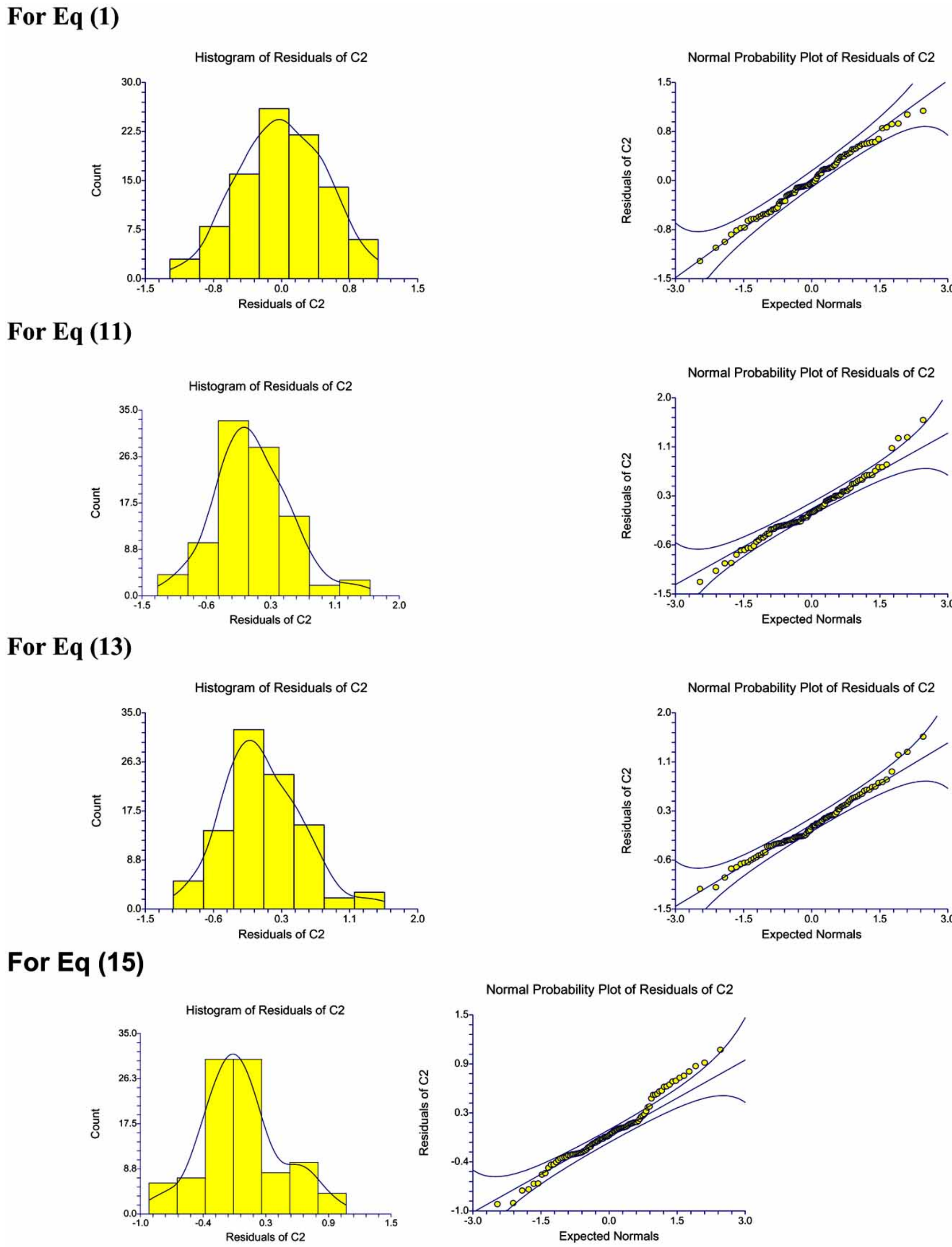


Fig. (5). Histogram and normal probability plots.

depending on how well it can improve the correlation based on already selected descriptors.

COMMENTS ON BALABAN AND BALABAN TYPE INDICES

The Balaban index (*J*) is a variant of connectivity index, represents extended connectivity and is a good descriptor for the shape of the molecule and that shape of the molecule influences heat of diffusion. It is a highly discriminating descriptor, whose values do not substantially increase with the molecular size and number of rings present in the molecule.

The Balaban-type indices *J*hetz (Balaban-type index from *Z*-weighted distance matrix i.e. Baryz-matrix), *J*hetm (Balaban-type index from mass-weighted distance matrix), *J*hetv (Balaban-type index from van der Walls-weighted distance matrix), *J*hete (Balaban-type index from electro negativity-weighted distance matrix), *J*hetp (Balaban-type index from polarizability-weighted distance matrix), and *BAC* (Balaban centric index) and the weighted *J* indices.

Apart from the fact that the Balaban index (*J*) is the highly discriminating index and that it can be weighted easily yielding different types of Balaban indices, very little work is done on the use of Balaban type indices in developing qualitative structure-property-activity-toxicity-relationships (*QSPR* / *QSAR* / *QSTR*). The primary reason for this is that theoretical chemistry has been very slow to appreciate the overriding importance of Balaban and Balaban-type indices in modifying their physicochemical and biological processes. Never-the-less, earlier [52-54], we has used this index successfully in developing some *QSPR* / *QSAR* models. Furthermore the authors, in collaboration with Balaban recently have undertaken a project for investigating the role of Balaban and Balaban-type indices is developing *QSPR* / *QSAR* / *QSTR* models [55-57]. In this sense, the present work is the extension of our earlier.

In one of our earlier report (in collaboration with Balaban), while describing super molecular complexing ability vis-à-vis estimation of pKa of substituted sulfonamides we observed that the most discriminating Balaban index (J) in multi-parametric regression analysis yielded excellent models, better than the Balaban-type indices; thus, establishing the superiority of J index over Balaban-type indices. Our recent work in collaboration with Balaban has also indicated that modeling power as well as predictive ability of the model improved highly by using Balaban and Balaban-type indices.

In the present case also we observed that statistically significant models yielded only when Balaban type indices are involved in the regression procedure. In models expressed by both the eq (**1**) and (**2**) *J*hetm and *J*hete are involved as the correlating parameters. In both these equations Jhetm has the negative coefficient while Jhete has the positive coefficient. Their coefficients are also high compared with coefficients of other parameters involved in these models. The negative coefficient of Jhetm indices that mass-weighted distances are not favorable for the exhibition of logKi(hCAII), while positive coefficient of *J*hete indicates that electro negativity-weighted distances are favorable. Same in the case with models expressed by (eq. (**15-17**)).

INTERCORELATIONS BETWEEN VARIABLES

The dendrograms are drawn using the formulae of Spearman [61], conceptually based on pooling the standardized values of the variables within each cluster to be correlated with other clusters.

ALL VARIABLES, WHETHER USED OR UNUSED IN THE EQUATIONS

As indicated by the dendrograph (Fig. (**6**)), the topological variables *J*, *J*hetv, *J*hete, *J*hetp, *J*hetm and *J*hetz are very strongly mutually correlated and also that the last two are so strongly correlated that they are effectively a single variable

CA-inhibitory activity correlates best, albeit weakly, with this group. The other topological variables ${}^1\chi$, *W* and *Sz* are also very strongly correlated with each other but negatively correlated with the first group. The atomic charges Q_O, Q_N and Q_H but not Q_C is strongly positively correlated with each other but not with any other variable, reflecting the charge on the sulfonamide group as a whole. It is well-known that ionization of the protons from this group is a necessary prelude to the CA-inhibitory activity of sulfonamides. Another cluster is formed by the orbital energies and Q_C, which are weakly positively correlated with each other, but not with any other variable.

The Indicator Variables

There is a weak correlation (Fig. (**6**)) between activity and I_1.

The Quantum-Theoretic Variables

There is only a weak positive correlation between activity and Q_C, E_L and E_{SH} and a strong negative correlation with Q_N, Q_O and Q_H .

The correlation of activity with the orbital energies does not persist when these variables are isolated.

The Topological Variables

There is a weak positive correlation of activity with the very tight cluster of *J*, *J*hete, *J*hetp, *J*hetp, *J*hetv, *J*hetm and *J*hetz. This correlation is so strong that it seems likely that only one of these variables can be included in a correlation. The correlation with *BAC,* ${}^1\chi$, *W* and *Sz* is very weak. The last three of these are very strongly intercorrelated, and are effectively a single variable. .

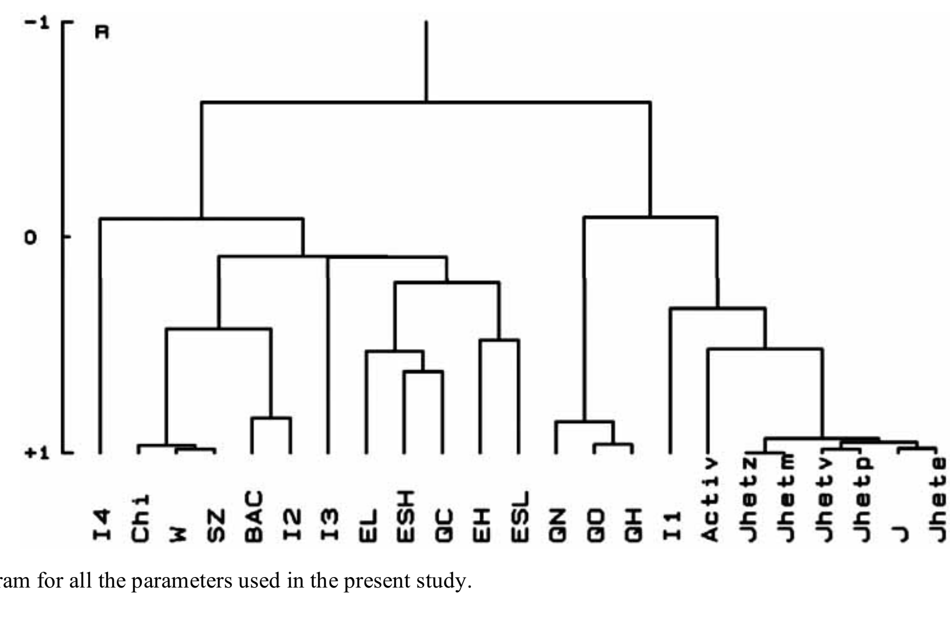


Fig. (6). Dendrogram for all the parameters used in the present study.

CONCLUSIONS

From the results and discussion made above we conclude that one can successfully use topological indices or quantum-theoretical descriptors for modeling inhibition of human carbonic anhydrase-II i.e., can estimate $\log K_i$ (hCA-II). Also, that the combinations of topological and quantum-theoretical descriptors gives still better way to model $\log K_i$ (hCA-II). The models so obtained have excellent statistics as well as excellent predictive ability (i.e. predictive power).

EXPERIMENTAL SECTION

(1) Carbonic anhydrase-II inhibitory activity: The carbonic anhydrase-II inhibitory activities in terms of $\log K_i$ (hCA-II) (nm) were adopted from our earlier work.

(2) Topological indices: All the topological indices were calculated using the DRAGAN software.

(3) Quantum-theoretical descriptors: Used from our earlier study.

(4) Regression analysis: The statistical calculations were done using MARTHA software and the final equation was calculated with the multiple linear regression facility of the statistical package NCSS. The method of maximum R^2 was used in arriving at the most appropriate model for modeling $\log K_i$ (hCA-II).

ACKNOWLEDGEMENTS

Authors are thankful to Profs A.T Balaban, M.V Diudea, I. Lukovits and others for their kind help in making calculations and for giving final shape to this paper. One of the authors, Shalini Singh, expresses her thanks to the Department of Science & Technology, Government of India, New Delhi, for awarding DST project SR/WOS-A/CS/61/2004 under Woman Scientists Scheme.

REFERENCES

[1] (a) Supuran, C.T.; Scozzafava, A.; Casini, A. *Development of sulfonamide carbonic anhydrase inhibitors (CAIs). In Carbonic anhydrase, its inhibitors and activators*, Supuran, C.T. ; Scozzafava, A.; Conway, J. Eds. CRC Press, Boca Raton (FL), **2004**, pp. 67; (b) Clare, B.W.; Supuran, C.T. *A perspective on quantitative structure-activity relationships and carbonic anhydrase inhibitors. Expert Opin. Drug Metab. Toxicol*., **2006**, *2*, 113.

[2] Mincione, F.; Menabuoni, L.; Supuran, C.T. *Clinical applications of the Carbonic anhydrase inhibitors in ophthalmology. In Carbonic anhydrase, its inhibitors and activators*, Supuran, C.T.; Scozzafava, A.; Conway, J. Eds. CRC Press, Boca Raton (FL), **2004**, pp. 243.

[3] Chegwidden, W.R.; Carter, N.D.; Edwards, Y.H. (Eds.). *The Carbonic Anhydrases*: New Horizons, Birkhauser Verlag, Basel-Boston-Berlin, **2000**.

[4] Clare, B.W.; Supuran, C.T.; *QSAR Studies of Sulfonamide Carbonic Anhydrase Inhibitors*, Supuran, C.T.; Scozzatava, A.; Conway, J. (Eds.), *Carbonic Andydrase, its Inhibitors and Activators*, CRC Press, Boca Raton: FI, USA, **2004**.

[5] Supuran, C.T. *In Carbonic Anhydrase and Modulation of Physiologic and Pathologic Process in the Organism*; Pacscas, I.; Ed.; Helicon: Timisoara (Romania), **1994**, pp 29-111.

[6] Supuran, C.T.; Scozzafava, A.; Conway, J. (Eds) *Carbonic Anhydrase–Its Inhibitors and Activators*, CRC Press, Boca Raton (USA), **2004**, pp 1.

[7] Maren, J.H. *J. Glaucoma,* **1995**, *4*, 49.

[8] Mann, T.; Keilin, D. *Nature,* **1940**, *146*, 164.

[9] Jaiswal, M.; Khadikar, P.V. *Bioorg. Med. Chem.,* **2004**, 12, 1793.

[10] Jaiswal, M.; Khadikar, P.V. *J. Indian Chem. Soc*., **2005**, *82,* 247.

[11] Jaiswal, M.; Khadikar, P.V.; Scozzafava, A.; Supuran, C.T. *Bioorg. Med. Chem. Lett.,* **2004**, *14*, 3283.

[12] Khadikar, P.V.; Diudea, M.V.; Singh, J.; Shrivastva, A.; Singh, S.; Karmarkar, S.; Lakhwani, M.; Thakur, P. *Curr. Bioac. Compd.,* **2006**, *2*, 19.

[13] Agrawal, V.K.; Bano, S.; Supuran, C.T.; Khadikar, P.V. *Eu. J. Med. Chem.*, **2004**, *39*, 593.

[14] Khadikar, P. V.; Sharma, V.; Karmarkar, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett.,* **2005**, *15,* 923.

[15] Gray, W.D.; Maren, J.H.; Sisson, G.H.; Smith, F.H., *J. Pharmacol. Exp. Ther.,* **1957**, 121, 160.

[16] Duffel, M.W.; Ing, I.,S.; Segnrra, J.H.; Dixson, J.A.; Barfkneele, C.F.; Schoenwald, R.D. *J. Med. Chem.*, **1986**, *29*, 1488.

[17] Verpoorte, J.A.; Mehta, S.; Edsall, J.T. *J. Biol. Chem.,* **1967**, *242*, 4221.

[18] Hansch, C.; Hoekman, D.; Gao, H. *Chem. Rev.*, **1996**, *96*, 1045.

[19] Hansch, C. ; Leo, A. *Exploring QSAR: Fundamental and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, **1995**.

[20] Diudea, M.V.; Florescu, M.S.; Khadikar, P.V. Molecular *Topology and Its Applications*, EFICON, Bucharest, Eficon Press, Bucuresti, **2006**, 412 pp.

[21] Diudea, M.V. (Ed) *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science, **2000**.

[22] Devillers, J.; Balaban, A.T. (Eds) *Topological Indices and Related Descriptors in QSAR and QSPR,* Gordon & Breach Science Pub., Amsterdam (The Netherlands), **1999**.

[23] Diudea, M.V. (Ed) *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science. **2000**.

[24] Jaiswal, M.; Khadikar, P.V.; Supuran, C.T. *Bioorg. Med. Chem.* **2004**, *12*, 2477.

[25] Thakur, A.; Thakur, M.; Khadikar, P.V.; Supuran, C.T. *Bioorg. Med. Chem. Lett.,* **2005**, *15,* 203.

[26] Balaban, A.T.; Khadikar, P.V.; Supuran, C.T; Thakur, A.; Thakur, M. *Bioorg. Med. Chem. Lett.,* **2005** (In Press).

[27] Agrawal, V. K; Singh, J.; Khadikar, P.V.; Supuran, C.T. *Bioorg. Med. Chem. Lett.,* **2006**, *16*, 2044.

[28] Agrawal, V. K.; Benergi, M.; Gupta, M.; Singh, J.; Khadikar, P.V.; Supuran, C.T. *Eu. J. Med. Chem*., **2005**, *40*, 1002.

[29] Mandoli, D.; Joshi, S.; Khadikar, P.V.; Khosla, N. *Bioorg. Med. Chem. Lett.,* **2005**, *15*, 405.

[30] Jaiswal, M.; Khadikar, P.V.; Supuran, C.T. *Bioorg. Med. Chem. Lett.,* **2004**, *14,* 5161.

[31] Jaiswal, M.; Khadikar, P.V.; Supuran, C.T. *Bioorg. Med. Chem.,* **2004**, *12*, 2477.

[32] Scozzafava, A.; Menabuoni, L.; Mincione, F.; Briganti, F.; Mincione, G.; and Supuran, C.T., *J. Med. Chem.,* **2000**, *43*, 4542.

[33] (a) Wiener, H. *J. Am. Chem. Soc.,* **1947**, *69*, 17, (b) Gutman, I. *A formula for the Wiener number of trees and its extension to graphs containing cycles Graph Theory Notes* New York **1994**, *27,* 9.

[34] Khadikar, P.V.; Kale, P.P.; Deshpande, N.V.; Karmarkar, S.; Agrawal, V.K. *Commun. Math. Comput. Chem., (MATCH)* **2001**, *43,* 7.

[35] Khadikar, P.V.; Deshpande, N.V.; Kale, P.P.; Dobrynin, A.; Gutman, I.; Domotor, G. The Szeged Index and an Analogy with the Wiener Index, *J. Chem. Inf. Comput. Sci*., **1995**, *35,* 545-550.

[36] Khadikar, P.V.; Karmarkar, S.; Agrawal, V.K.; Singh, J.; Shrivastava, A.; Lukovits, I.; Diudea, M.V. *Lett. Drug Des. Discov.,* **2005**, *2*, 606.

[37] Randic, M. *J. Amer. Chem. Soc.,* **1975**, *97,* 6909.

[38] Balaban, A.T. *Chem. Phys. Lett.,* **1982**, *89,* 399.

[39] Balaban, A.T. *Chemical Applications of Graph Theory*, Academic Press, London (UK), **1976**.

[40] DRAGON software for calculation of topological indices: www.disat.unimib.it

[41] MOPAC version 6, adapted for windows, http://www.ccl.net

[42] Hyperchem-7 software for calculating the molecular modeling parameters; www.hyper.com

[43] MARTHA, http://www.chem.uwa.edu.au/resarch/bclare.

[44] ORIGIN- 6 program.

[45] NCSS, http://www.ncss.com

[46] Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, John Wiley & Son, New York, 3rd Ed. **2000**.

[47] Todeschini, R. ; Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, **2000**.

[48] Hansch, C. (Ed) *Comprehensive Drug Design and Pergamon Press*: New York **1990**, pp. 19.
[49] Pogliani, L. *Amino Acids,* **1994**, *6,* 141.
[50] Pogliani, L. *J. Phys. Chem*., **1996**, *100,* 18065.
[51] Pogliani, L. *Chem. Rev.,* **2000**, *100,* 3827. .
[52] Khadikar, P.V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveshwar, H. *Bull. Soc. Chem. Belg.,* **1997**, *106*, 767.
[53] Sapre, N.S.; Sikarwar, A.; Khadikar, P.V. *Oxid. Commun.,* **2001**, *24*, 28.
[54] Thakur, A.; Thakur, M.; Khadikar, P.V.; Supuran, C.T.; Sudele, P. *Bioorg. Med. Chem.*, **2004**, *12*, 789.

Received: 14 October, 2006 Revised: 23 June, 2007 Accepted: 25 June, 2007

[55] Balaban, A.T.; Khadikar, P.V.; Supuran, C.T; Thakur, A.; Thakur, M. *Bioorg. Med. Chem. Lett.,* **2005**, *15,* 3966.
[56] Singh, J.; Lakhwani, M.; Khadikar, P.V.; Agrawal, V.K.; Balaban, A.T.; Clare, B.W.; Supuran, C.T *Rev. Roum. Chim.,* **2006**, (In Press).
[57] Khadikar,P.V.; Clare W.B.; Balaban, A.T.; Supuran, C.T.; Agarwal, V.K.; Singh, J.; Joshi,A.K.; and Lakhwani,, M. *Romania Rev.,* **2005** (In Press).
[58] Shapiro, S.; Guggenheim, B. *Quant. Struct. Act. Relat.,* **1998**, *17,* 327.
[59] Randic, M. *J. Chem. Inf. Comput. Sci.,* **1997**, *37*, 672.
[60] Randic, M. *Acta Chem. Slov.,* **1998**, *45,* 239.
[61] Spearman, C. *Br. J. Psychol.,* **1913**, *5*, 417.