# Comparative QSAR Study on Para-substituted Aromatic Sulphonamides as CAII Inhibitors: Information versus Topological (Distance-Based and Connectivity) Indices

**Article** *in* Chemical Biology & Drug Design · April 2008

DOI: 10.1111/j.1747-0285.2007.00625.x · Source: PubMed



**7 authors**, including:

Shaik Bashirulla
National Institute of Technical Teachers' Tr…
**33** PUBLICATIONS **63** CITATIONS
SEE PROFILE

Shalini Singh
Bareilly College, Bareilly
**44** PUBLICATIONS **320** CITATIONS
SEE PROFILE

Omar Deeb
Al-Quds University
**38** PUBLICATIONS **375** CITATIONS
SEE PROFILE

*Chem Biol Drug Des 2008;* **71:** 244–259

Research Article



doi: 10.1111/j.1747-0285.2007.00625.x

# Comparative QSAR Study on Para-substituted Aromatic Sulphonamides as CAII Inhibitors: Information versus Topological (Distance-Based and Connectivity) Indices

**Jyoti Singh[1], Basheerulla Shaik[1], Shalini Singh[2], Vijay K. Agrawal[1], Padmakar V. Khadikar[3,*], Omar Deeb[4] and Claudiu T. Supuran[5]**

*[1]QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa-486 003, India*
*[2]Department of Chemistry, Bareilly College, Bareilly 243001, UP, India*
*[3]Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd., 3, Khatipura, Indore 452 007, India*
*[4]Faculty of Pharmacy, Al-Quds University, PO Box 20002, Jerusalem, Palestine*
*[5]Laboratorio di Chimica Bioinorganica, Dipartimento di Chimica, University of Florence, via della Lastruccia, 3, RM-188, Polo Scientifico, 50019 Sesto Fioventino, Fireze, Italy*
**Corresponding author: Padmakar V Khadikar, pvkhadikar@rediffmail.com*

**Comparative quantitative structure–activity relationship studies on para-substituted aromatic sulphonamides carbonic anhydrase II (CAII) inhibitors are reported in this paper. The study is made utilizing (i) information indices along; (ii) distance-based and connectivity indices and (iii) combination of information, distance-based and connectivity type topological indices. The study has shown that distance-based and connectivity type indices are superior for modelling, monitoring and estimating CAII inhibition. The results are critically discussed using a variety of statistical parameters. Our results show that starting from the mono-parametric regression itself, our results are superior: Furthermore, our methodology allowed carrying out much higher-parametric regressions, yielding a nine-parametric model with $R^2$ as high as 0.8375. The eight-parametric regression, gave $R^2$ = 0.8343. As there is not much difference, we have considered the eight-parametric regression the best.**

**Key words:** aromatic sulphonamides, carbonic anhydrase II, information indices, QSAR, regression analysis, topological indices

Received 24 June 2007, revised and accepted for publication 19 December 2007

## Introduction

Quantitative structure–activity relationships (QSAR) are widely used in modelling a variety of physico-chemical parameters as well as biological activity of chemically active compounds (1). In most cases, topological indices such as Wiener (W) (2), Randic connectivity (3,4), Kier–Hall valence connectivity (5,6), Szeged (Sz) (7–10) and Padmakar-Ivan (PI) (11–18) indices are used. Recently, we advocated the use of Balaban and Balaban type indices for modelling carbonic anhydrase inhibitors (19–21).

QSAR study on carbonic anhydrase inhibitors were studied earlier by us (22–28) and also by many other authors (29–31) through QSAR. Needless to say, carbonic anhydrases (CAs, EC 4.2.1.1) are the metallo-enzymes and were extensively studied in the last decade. Also, that sulphonamides represent an important class of biological active compounds and lead to different classes of pharmacological agents such as antibacterial sulphonamides, sulphonamides that inhibits the zinc-enzyme carbonic anhydrase, which are then used in the treatment of some of diabetes, antithyroid drugs and others (1).

Recently, one of the authors (Supuran) has reported QSAR study on para-substituted aromatic sulphonamides as CAII inhibitors using information indices (30) in which a set of 47 compounds were initially modelled using 29 topological indices, which ultimately resulted into four models with excellent statistics. Of these four models, a model containing ${}^1\chi_{inf}$, ${}^0\chi^v_{inf}$, ${}^1\chi^v_{inf}$ along with N-rings (tetra-parametric model in the following table) as the correlating parameters is found the best. In arriving at these excellent models, successive regression analyses up to five-parametric regressions were performed by Supuran and the results are summarized below:

| Model | $R^2$ | RMS | $F$ |
|---|---|---|---|
| (i) Bi-parametric | 0.6642 | 0.2926 | 43.52 |
| (ii) Tri-parametric | 0.6984 | 0.2773 | 33.18 |
| (iii) Tetra-parametric | 0.7283 | 0.2632 | 28.14 |
| (iv) Penta-parametric | 0.7296 | 0.2625 | 22.12 |

Considering the similar statistics of tetra- and penta-parametric regression, no higher parametric regressions were performed by Supuran and also depending upon lesser number of correlating

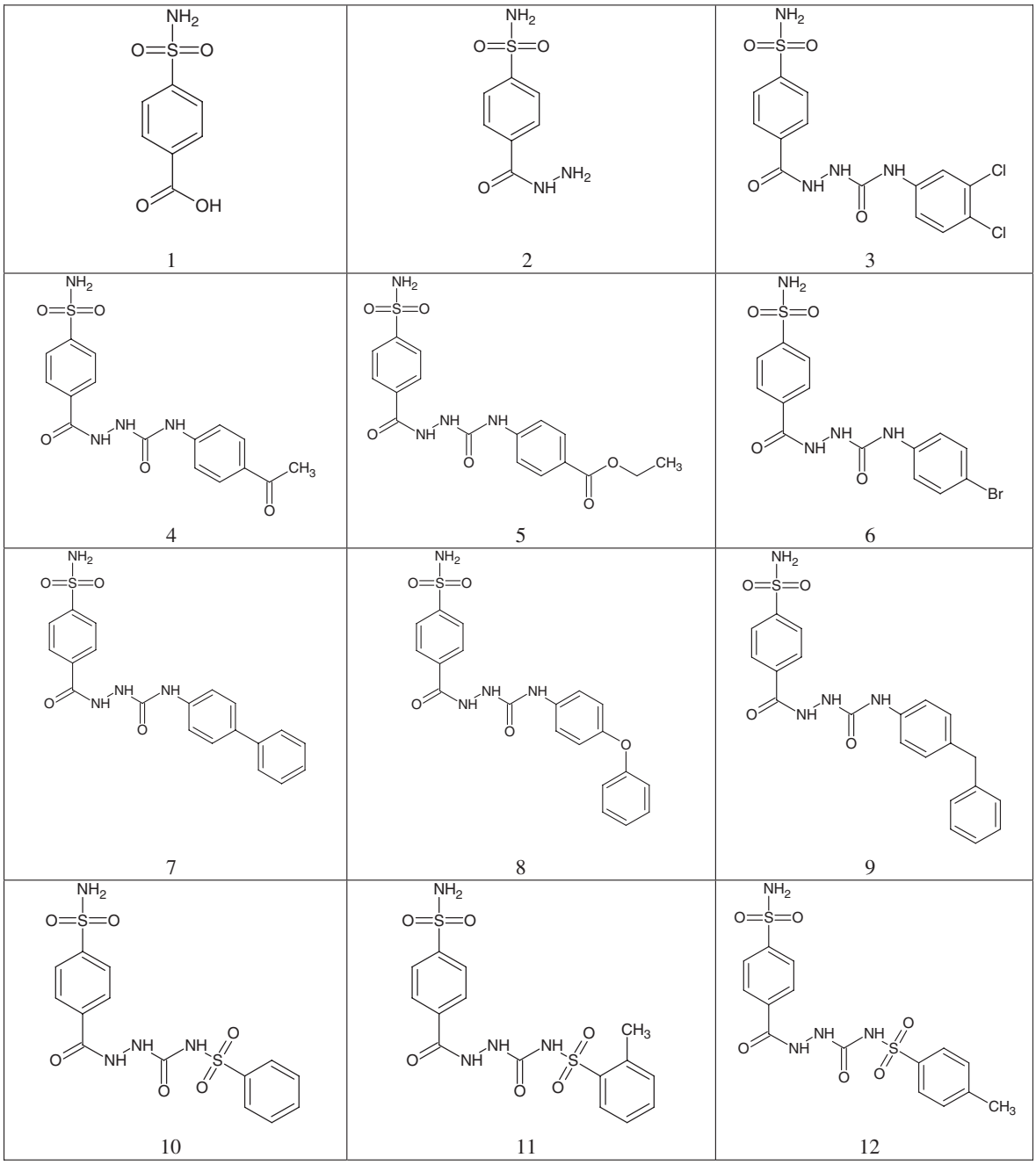


**Figure 1:** Structural details of para-substituted aromatic sulphonamides used in the present study.

parameters, the four-parametric regression was considered the best. In an attempt to obtain still better results, we undertook this study which, as discussed below, establishes that the methodology used by us in the present study is far superior to the earlier method (30). As will be discussed, our results show that starting from the mono-parametric regression itself, our results are superior. Furthermore, our methodology allowed carrying out much higher-parametric regressions, yielding a nine-parametric model with $R^2$ as high as 0.8375. The eight-parametric regression, gave $R^2 = 0.8343$. As there is not much difference, we have considered the eight-parametric regression the best. The results are now taken up for discussion.

## Results and Discussion

The structural details of the sulphonamides used are given in Figure 1. They incorporate hydrazine moieties, urea, sulfaureas or

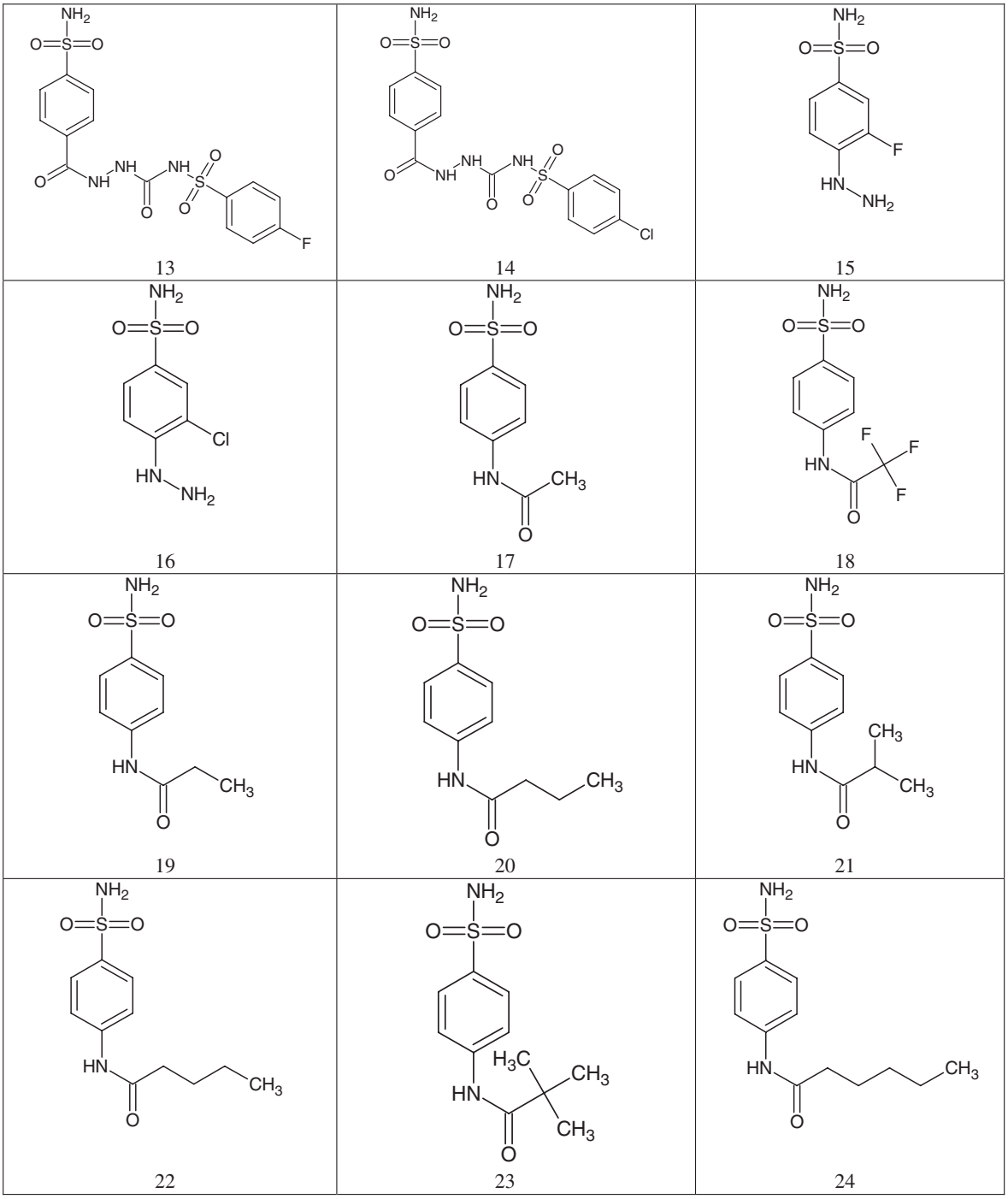


**Figure 1:** (Continued)

a simple aliphatic tail. The experimental activity (30) and the calculated values of topological indices are summarized in Tables 1–3. These indices are 2D descriptors accounting for internal atom arrangement of the compound and encode in numerical form, information about molecular size, shape, branching, presence of heteroatom and multiple bonds. To decide which topological indices are useful for proposing appropriate model for modelling the activity, we performed variable selection in multiple regression analysis. This helped us to set up the best combination of descriptors and thus propose the best model. This procedure adopted by us also helps us to arrive at the optimal model complex in predicting a response variable by a reduced set of descriptors out of the larger pool, which are not highly intercorrelated. Even if by chance, one or more proposed models contain

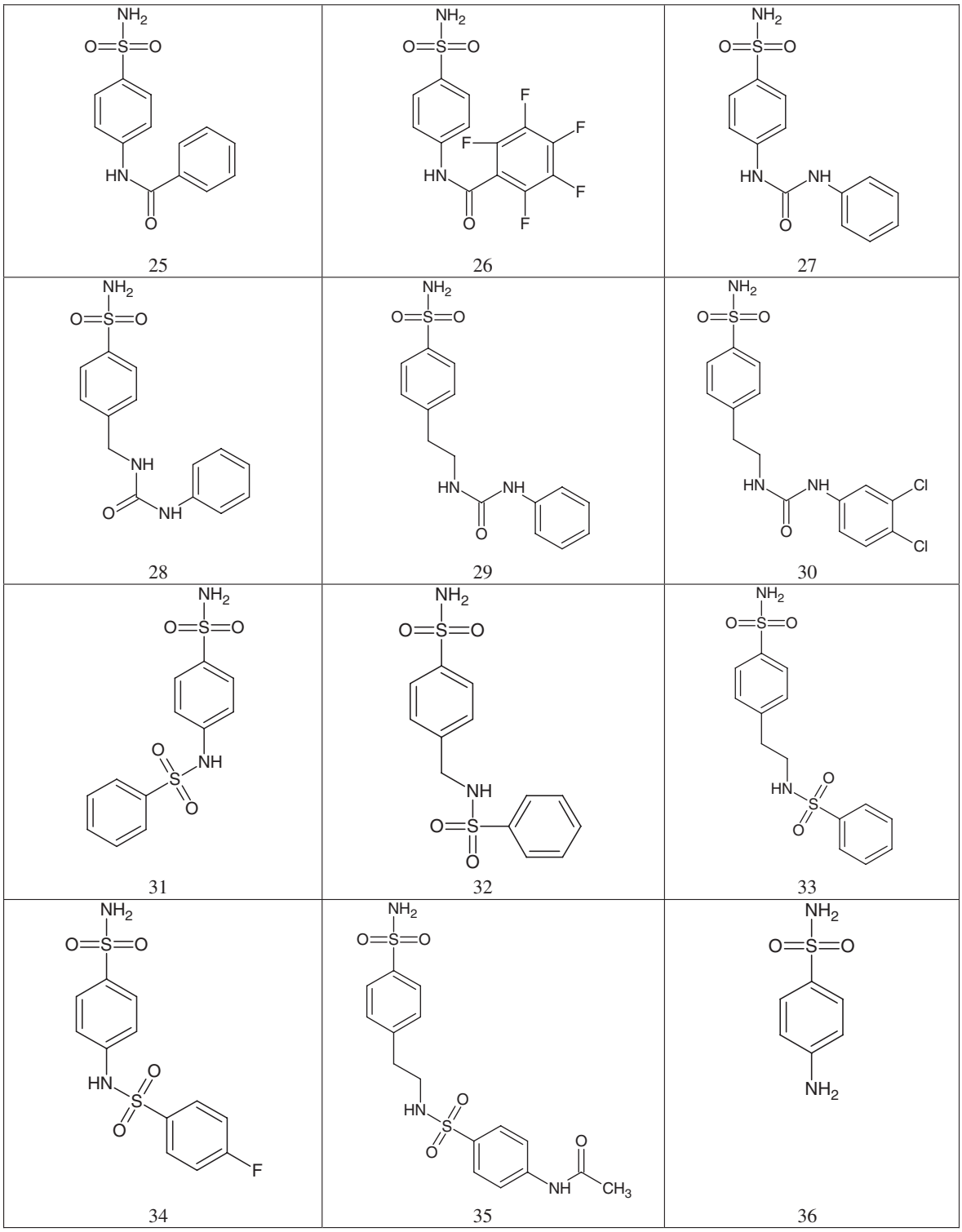


**Figure 1:** (Continued)

highly correlated parameters, the model will suffer from the defect because of colinearity. Such a defect, if exists, will be discussed as per the recommendations given by Randic (32,33). The successive regression analysis using method of maximum $R^2$, yielded nine models. These models are given in Table 4. We observed that these models are all statistically significant and go from one to nine-variable models. We discuss these models in detail:

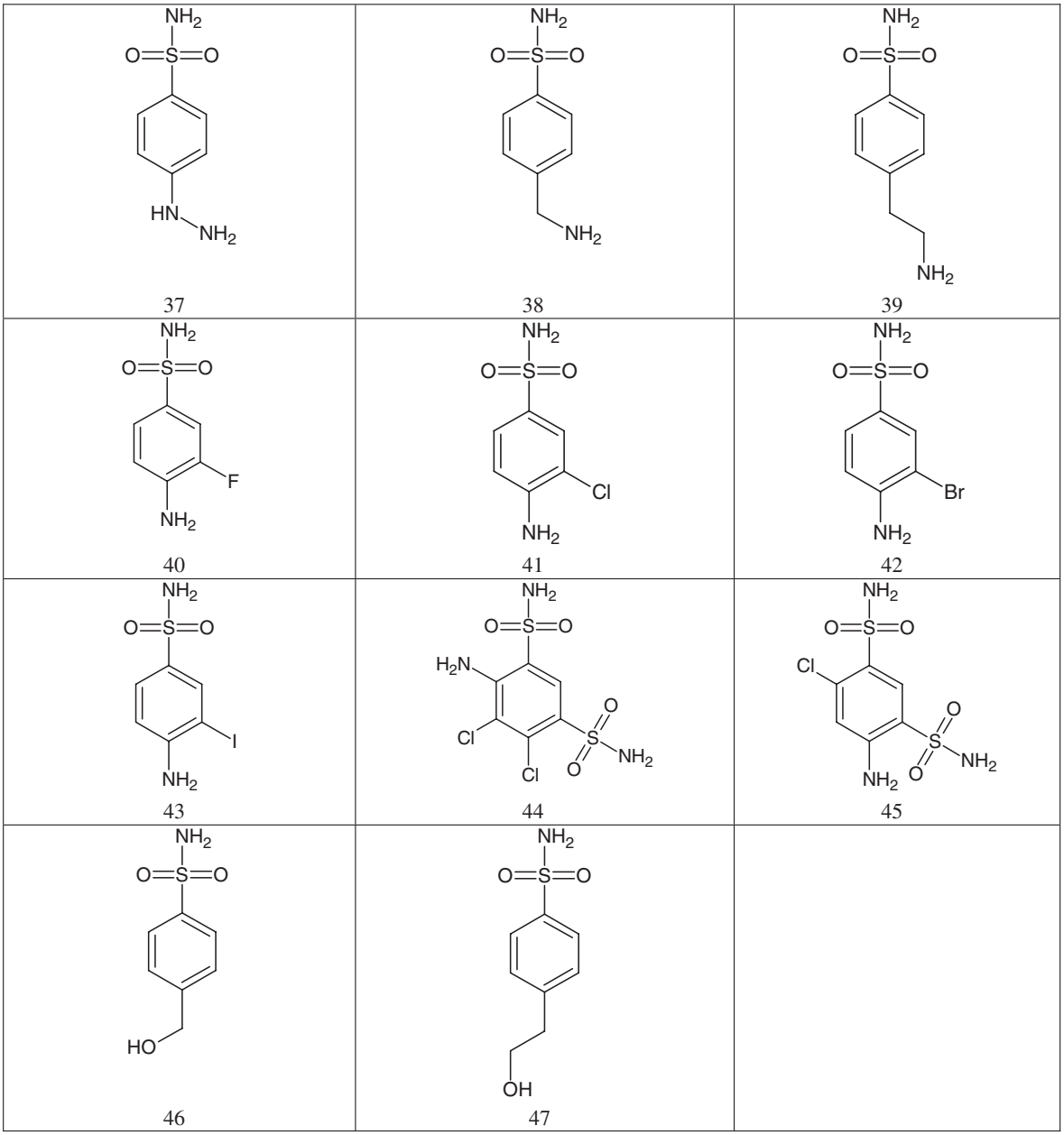


**Figure 1:** (Continued)

### *The best one-variable model*

The best one-variable model contains ${}^{0}\chi^{v}$ as the correlating parameter. This model is shown below:

$$\log K_i = 3.1259 - 0.1342(\pm 0.0156)\,{}^{0}\chi^{v} \qquad (1)$$

$$n = 47, R^2 = 0.6233, R^2A = 0.6150,\ \mathrm{SE} = 0.1811, F = 74.472$$

Here and hereafter, $n$ is the number of compounds used, $R^2$ is the coefficient of variance, $R^2A$ is the adjustable $R^2$. SE is the standard error of estimation and $F$ is the Fisher's statistics.

This eqn 1 shows that even a single parameter, ${}^{0}\chi^{v}$, explains 62% of variation in the activity ($\log K_i$). That is, a decrease in the magnitude of ${}^{0}\chi^{v}$ is favourable for the exhibition of $\log K_i$. In other words, a decrease in the number of heteroatom increases $\log K_i$.

### *The best two-variable model*

The successive regression analysis indicated that by adding of Fractional partial positive surface area (FPSA1) to the above eqn 1, there is an appreciable improvement in the statistics which is demonstrated by the following model:

**Table 1:** Calculated values of topological indices

| Compound no. | log $K_i$ | $W$ | $^0\chi$ | $^1\chi$ | $^2\chi$ | $^3\chi$ | $^0\chi^v$ | $^1\chi^v$ | $^2\chi^v$ | $^3\chi^v$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.4116 | 252 | 10.0605 | 5.9097 | 6.2218 | 3.9394 | 6.8050 | 4.3184 | 3.3250 | 1.7841 |
| 2 | 2.0934 | 316 | 10.7676 | 6.4477 | 6.3633 | 4.4453 | 7.2911 | 4.5379 | 3.4428 | 1.8848 |
| 3 | 1.1139 | 1810 | 18.6126 | 11.6637 | 11.5069 | 8.3586 | 13.9144 | 8.0467 | 6.0446 | 3.6030 |
| 4 | 1.1761 | 2069 | 19.3197 | 12.1637 | 11.8980 | 8.4860 | 13.0549 | 7.6169 | 5.4899 | 3.0666 |
| 5 | 0.9542 | 2590 | 20.7339 | 13.2017 | 12.4200 | 9.0920 | 13.9630 | 8.0251 | 5.6523 | 3.1874 |
| 6 | 0.8633 | 1629 | 17.7423 | 11.2530 | 10.9991 | 7.7873 | 13.6106 | 7.8948 | 5.8927 | 3.2056 |
| 7 | 1.0414 | 2838 | 20.8552 | 13.8255 | 13.0521 | 9.8598 | 14.6466 | 8.6628 | 6.1607 | 3.5896 |
| 8 | 1.2553 | 3186 | 21.5623 | 14.3087 | 13.4998 | 10.0096 | 15.0549 | 8.8210 | 6.1710 | 3.4978 |
| 9 | 1.1761 | 3186 | 21.5623 | 14.3087 | 13.4998 | 10.0096 | 15.1466 | 8.9128 | 6.2857 | 3.5896 |
| 10 | 1.8261 | 1967 | 19.3720 | 12.0931 | 12.1652 | 8.5010 | 13.6879 | 8.8493 | 6.8868 | 4.0701 |
| 11 | 1.7324 | 2156 | 20.2423 | 12.5039 | 12.6985 | 8.9060 | 14.1879 | 9.0993 | 7.1368 | 4.3483 |
| 12 | 0.9912 | 2196 | 20.2423 | 12.4870 | 12.7870 | 8.9117 | 14.1879 | 9.0993 | 7.1368 | 4.1951 |
| 13 | 0.9777 | 2196 | 20.2423 | 12.4870 | 12.7870 | 8.9117 | 14.0658 | 9.0383 | 7.0758 | 4.1647 |
| 14 | 0.9590 | 2196 | 20.2423 | 12.4870 | 12.7870 | 8.9117 | 14.8218 | 9.4162 | 7.4537 | 4.3537 |
| 15 | 1.7076 | 242 | 10.0605 | 5.9477 | 6.0216 | 4.0960 | 6.7609 | 4.2727 | 3.300 | 1.8036 |
| 16 | 1.8808 | 242 | 10.0605 | 5.9477 | 6.0216 | 4.0960 | 7.5168 | 4.6507 | 3.6781 | 2.0770 |
| 17 | 2.3909 | 325 | 10.7676 | 6.3929 | 6.7133 | 3.8763 | 7.3440 | 4.5615 | 3.4281 | 1.7801 |
| 18 | 2.1239 | 562 | 13.2676 | 7.6042 | 8.5865 | 5.2368 | 8.4779 | 5.1284 | 3.9259 | 2.0227 |
| 19 | 2.3655 | 402 | 11.4747 | 6.9309 | 6.8328 | 4.5041 | 7.8440 | 4.8115 | 3.5531 | 1.8871 |
| 20 | 2.3560 | 494 | 12.1818 | 7.4309 | 7.2133 | 4.5887 | 8.3440 | 5.0615 | 3.6781 | 1.9497 |
| 21 | 2.4116 | 481 | 12.3450 | 7.3036 | 7.5407 | 4.9147 | 8.3440 | 5.0615 | 3.8031 | 1.9940 |
| 22 | 2.3304 | 602 | 12.8890 | 7.9309 | 7.5668 | 4.8577 | 8.8440 | 5.3115 | 3.8031 | 2.0121 |
| 23 | 2.3617 | 562 | 13.2676 | 7.6042 | 8.5865 | 5.2368 | 8.8440 | 5.3115 | 4.1781 | 2.1009 |
| 24 | 1.7993 | 727 | 13.5960 | 8.4309 | 7.9204 | 5.1077 | 9.3440 | 5.5615 | 3.9281 | 2.0747 |
| 25 | 1.5682 | 784 | 13.8805 | 8.9654 | 8.6948 | 6.2312 | 9.8440 | 6.0615 | 4.4281 | 2.4940 |
| 26 | 1.2304 | 1394 | 18.2317 | 11.0356 | 11.1696 | 8.9654 | 11.7338 | 7.0064 | 5.3730 | 3.2039 |
| 27 | 2.3802 | 940 | 14.5876 | 9.4485 | 9.1643 | 6.2677 | 10.2911 | 6.2587 | 4.4909 | 2.4686 |
| 28 | 2.0212 | 1120 | 15.2947 | 9.9485 | 9.5060 | 6.5997 | 10.7911 | 6.5087 | 4.6291 | 2.5378 |
| 29 | 1.8751 | 1321 | 16.0018 | 10.4485 | 9.8595 | 6.8413 | 11.2911 | 6.7587 | 4.7541 | 2.6068 |
| 30 | 1.1139 | 1678 | 17.7423 | 11.2530 | 10.9890 | 7.8233 | 13.5589 | 7.8926 | 5.8880 | 3.4951 |
| 31 | 1.6902 | 868 | 14.8031 | 9.2887 | 9.6113 | 6.5967 | 10.9769 | 7.5438 | 6.0726 | 3.6563 |
| 32 | 1.6021 | 1042 | 15.5103 | 9.7887 | 9.9459 | 6.9788 | 11.4769 | 7.7938 | 6.2108 | 3.6444 |
| 33 | 1.4472 | 1237 | 16.2174 | 10.2887 | 10.2995 | 7.2154 | 11.9769 | 8.0438 | 6.3358 | 3.7135 |
| 34 | 0.9542 | 1004 | 15.6734 | 9.6825 | 10.2331 | 7.0073 | 11.3549 | 7.7328 | 6.2616 | 3.7508 |
| 35 | 1.8751 | 2040 | 19.3720 | 12.0764 | 12.3117 | 8.2617 | 13.8324 | 8.9451 | 6.9764 | 4.0386 |
| 36 | 2.4771 | 152 | 8.4831 | 4.9990 | 5.3229 | 3.2406 | 5.9357 | 3.8837 | 3.0111 | 1.5668 |
| 37 | 2.5051 | 201 | 9.1903 | 5.5370 | 5.4912 | 3.6489 | 6.3829 | 4.0837 | 3.1111 | 1.6668 |
| 38 | 2.2304 | 201 | 9.1903 | 5.5370 | 5.4912 | 3.6489 | 6.4357 | 4.1337 | 3.1494 | 1.6918 |
| 39 | 2.2041 | 262 | 9.8974 | 6.0370 | 5.8724 | 3.7685 | 6.9357 | 4.3837 | 3.2744 | 1.7609 |
| 40 | 1.7782 | 189 | 9.3534 | 5.4097 | 5.8306 | 3.8230 | 6.3137 | 4.0727 | 3.2001 | 1.7036 |
| 41 | 2.0414 | 189 | 9.3534 | 5.4097 | 5.8306 | 3.8230 | 7.0696 | 4.4507 | 3.5781 | 1.9770 |
| 42 | 1.6021 | 189 | 9.3534 | 5.4097 | 5.8306 | 3.8230 | 7.8997 | 4.8657 | 3.9931 | 2.2774 |
| 43 | 1.8451 | 189 | 9.3534 | 5.4097 | 5.8306 | 3.8230 | 8.4711 | 5.1515 | 4.2789 | 2.4841 |
| 44 | 1.4472 | 458 | 13.5939 | 7.4593 | 8.7580 | 5.6607 | 10.6919 | 7.1777 | 6.1827 | 4.1461 |
| 45 | 1.8751 | 399 | 12.7237 | 7.0317 | 8.3710 | 4.8500 | 9.5580 | 6.6108 | 5.6157 | 3.4145 |
| 46 | 2.0969 | 201 | 9.1903 | 5.5370 | 5.4912 | 3.6489 | 6.3967 | 4.1142 | 3.1397 | 1.6821 |
| 47 | 2.0414 | 262 | 9.8974 | 6.0370 | 5.8724 | 3.7685 | 6.8967 | 4.3642 | 3.2647 | 1.7560 |

$$\log K_i = 2.7906 - 0.1537(\pm 0.0151)^0\chi^v + 1.2202(\pm 0.3582)\text{FPSA1} \quad (2)$$

$$n = 47, R^2 = 0.7020, R^2A = 0.6884, \text{SE} = 0.1629, F = 51.814$$

The positive coefficient of the added parameter, namely FPSA1, makes a favourable contribution to the exhibition of log $K_i$.

### *The best three-variable model*

Further step-wise regression indicated the occurrence of a best three-variable model containing FNSA1 as the additional correlating parameters. Only a slight improvement in statistics was observed accordingly for the following regression expression:

$$\log K_i = -584399.4505 - 0.1588(\pm 0.0148)^0\chi^v + 584403.5083(\pm 281917.5869)\text{FPSA1} + 584402.2745(\pm 281917.6030)\text{FNSA1} \quad (3)$$

$$n = 47, R^2 = 0.7290, R^2A = 0.7101, \text{SE} = 0.1571, F = 38.563$$

It is interesting to mention that the parameters FPSA1 and FNSA1 are highly correlated and that only a slight improvement in

**Table 2:** Calculated values of topological indices

| Compound no. | $^{1}\chi^{shape}$ | $^{2}\chi^{shape}$ | $^{3}\chi^{shape}$ | TMSA | PPSA1 | PPSA2 | PPSA3 | PNSA1 | PNSA2 | PNSA3 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9.7160 | 3.1716 | 2.3059 | 370.925 | 125.423 | 42.0484 | 3.6441 | 245.502 | −82.305 | −14.983 |
| 2 | 10.6694 | 3.7611 | 2.5369 | 373.042 | 104.239 | 30.9548 | 2.9986 | 268.803 | −79.823 | −9.984 |
| 3 | 19.3461 | 7.9333 | 5.5230 | 577.465 | 182.246 | 115.8730 | 5.7310 | 395.219 | −251.280 | −24.368 |
| 4 | 19.4349 | 7.9929 | 5.5579 | 662.537 | 314.148 | 181.9850 | 7.4817 | 348.390 | −201.820 | −20.949 |
| 5 | 21.3710 | 9.3224 | 6.3321 | 738.585 | 379.415 | 266.2600 | 11.3560 | 359.169 | −252.050 | −21.105 |
| 6 | 18.2614 | 7.6599 | 5.4939 | 634.346 | 246.586 | 129.6760 | 5.6400 | 387.760 | −203.920 | −20.931 |
| 7 | 20.3957 | 8.7764 | 5.5745 | 732.054 | 420.840 | 195.0960 | 4.4133 | 311.214 | −144.270 | −16.114 |
| 8 | 21.3318 | 9.4118 | 6.3031 | 744.746 | 414.138 | 259.8510 | 7.6103 | 330.608 | −207.440 | −20.084 |
| 9 | 21.3708 | 9.4386 | 6.3238 | 741.640 | 436.513 | 202.3600 | 4.4135 | 305.127 | −141.450 | −16.105 |
| 10 | 19.5830 | 7.6524 | 5.3790 | 578.303 | 277.103 | 156.3630 | 5.5525 | 301.200 | −169.960 | −14.857 |
| 11 | 20.5704 | 7.8736 | 5.3668 | 596.109 | 310.126 | 175.1100 | 5.4590 | 285.983 | −161.480 | −13.407 |
| 12 | 20.5704 | 7.8736 | 5.6291 | 616.614 | 310.603 | 175.2750 | 5.3407 | 306.011 | −172.680 | −14.935 |
| 13 | 20.5013 | 7.8291 | 5.5935 | 597.094 | 236.891 | 169.0550 | 8.1489 | 360.203 | −257.060 | −23.114 |
| 14 | 20.8570 | 8.0581 | 5.7779 | 607.285 | 231.529 | 152.1320 | 6.8771 | 375.756 | −246.900 | −21.556 |
| 15 | 9.9344 | 3.3039 | 2.1568 | 374.844 | 101.559 | 36.2327 | 4.2006 | 273.286 | −97.499 | −12.176 |
| 16 | 10.2919 | 3.5240 | 2.3280 | 382.795 | 94.8565 | 28.9304 | 3.4585 | 287.938 | −87.819 | −11.580 |
| 17 | 10.7092 | 3.7864 | 3.1352 | 399.742 | 178.816 | 58.4715 | 5.5483 | 220.926 | −72.241 | −12.352 |
| 18 | 13.4849 | 4.3549 | 3.7509 | 421.361 | 120.324 | 66.2009 | 5.5590 | 301.037 | −165.630 | −17.247 |
| 19 | 11.7034 | 4.4333 | 3.3783 | 423.504 | 208.468 | 68.6793 | 4.60639 | 215.035 | −70.843 | −11.652 |
| 20 | 12.6984 | 5.1086 | 3.9404 | 456.553 | 238.409 | 78.6294 | 4.0666 | 218.144 | −71.946 | −12.085 |
| 21 | 12.6984 | 4.6471 | 3.6226 | 431.435 | 214.043 | 70.8453 | 4.0854 | 217.392 | −71.954 | −11.838 |
| 22 | 13.6940 | 5.8093 | 4.5817 | 466.184 | 257.889 | 85.0679 | 3.6863 | 208.294 | −68.708 | −11.384 |
| 23 | 13.6940 | 4.4803 | 3.8679 | 435.980 | 226.082 | 75.0684 | 4.0455 | 209.898 | −69.695 | −11.237 |
| 24 | 14.6901 | 6.5325 | 5.2002 | 483.797 | 272.296 | 89.8223 | 3.3318 | 211.501 | −69.768 | −11.406 |
| 25 | 13.2856 | 5.2307 | 3.4278 | 425.229 | 214.153 | 70.9917 | 3.0490 | 211.076 | −69.972 | −11.808 |
| 26 | 17.8477 | 6.1854 | 3.3980 | 546.635 | 152.757 | 147.3900 | 8.9726 | 393.877 | −380.040 | −33.547 |
| 27 | 14.2240 | 5.8458 | 4.1658 | 458.320 | 240.961 | 89.7621 | 3.3388 | 217.359 | −80.970 | −10.606 |
| 28 | 15.2036 | 6.5073 | 4.6897 | 546.538 | 322.339 | 119.6250 | 5.0416 | 224.198 | −83.204 | −10.851 |
| 29 | 16.1852 | 7.1884 | 5.2695 | 532.723 | 308.622 | 114.5960 | 4.0674 | 224.100 | −83.212 | −11.077 |
| 30 | 18.7247 | 7.9765 | 5.7525 | 600.122 | 251.774 | 136.9280 | 6.1635 | 348.349 | −189.450 | −21.843 |
| 31 | 14.4100 | 5.1748 | 3.5124 | 448.559 | 264.776 | 96.3634 | 4.7469 | 183.783 | −66.887 | −7.563 |
| 32 | 15.3900 | 5.7800 | 4.0000 | 509.824 | 326.336 | 118.2220 | 6.0530 | 183.489 | −66.473 | −7.986 |
| 33 | 16.3719 | 6.4059 | 4.4825 | 546.344 | 351.729 | 127.4830 | 5.8737 | 194.615 | −70.538 | −8.698 |
| 34 | 15.3213 | 5.3719 | 3.7291 | 461.260 | 227.245 | 116.7150 | 7.2785 | 234.016 | −120.190 | −15.648 |
| 35 | 19.9482 | 7.8908 | 5.8343 | 618.191 | 355.213 | 189.1120 | 9.3786 | 262.978 | −140.010 | −14.305 |
| 36 | 8.0604 | 2.5367 | 1.9704 | 334.601 | 125.423 | 27.1029 | 3.0468 | 209.178 | −45.202 | −9.0625 |
| 37 | 9.0115 | 3.1143 | 2.2011 | 364.718 | 116.303 | 26.2103 | 3.0799 | 248.415 | −55.983 | −7.859 |
| 38 | 9.0512 | 3.1391 | 2.2216 | 364.718 | 162.805 | 34.9203 | 4.1116 | 201.913 | −43.309 | −8.510 |
| 39 | 10.0436 | 3.7799 | 2.7278 | 395.123 | 184.440 | 39.5970 | 3.9375 | 210.683 | −45.231 | −9.062 |
| 40 | 8.9818 | 2.7400 | 1.9237 | 343.040 | 97.538 | 34.0996 | 3.9894 | 245.502 | −85.829 | −14.096 |
| 41 | 9.3389 | 2.9474 | 2.0947 | 352.601 | 92.176 | 27.3982 | 3.2470 | 260.425 | −77.409 | −13.044 |
| 42 | 9.5274 | 3.0588 | 2.1870 | 336.557 | 80.111 | 21.5307 | 2.6138 | 256.445 | −68.922 | −11.502 |
| 43 | 9.7756 | 3.2075 | 2.3109 | 342.351 | 148.358 | 32.0924 | 2.3285 | 193.993 | −41.964 | −8.219 |
| 44 | 14.3813 | 4.2090 | 2.7845 | 441.420 | 62.637 | 31.7503 | 4.0201 | 378.783 | −192.000 | −15.754 |
| 45 | 13.0965 | 3.8021 | 2.7956 | 426.908 | 83.820 | 36.7799 | 4.3847 | 343.088 | −150.550 | −13.628 |
| 46 | 9.0512 | 3.1391 | 2.2216 | 372.274 | 166.826 | 44.4740 | 5.9081 | 205.447 | −54.770 | −12.966 |
| 47 | 10.0436 | 3.7799 | 2.7278 | 408.628 | 195.164 | 52.0203 | 5.6787 | 213.465 | −56.898 | −13.823 |

statistics has occurred because of the addition of FNSA1. This indicates that eqn 3 could be transferred into eqn 2. Comments on the occurrence of both FPSA1 and FNSA1 in the same model are made in the following section. At this stage, it is worth mentioning that when there is a complete absence of linear relationship among the predictor variables, they are said to be orthogonal. In most regression applications, such as in the present case, the predictor variables are not orthogonal. In such case, one can use orthogonally and reduce the number descriptors. But this will be a different story resulting in yet another problem and thus another paper. What we can say is that many times, the lack orthogonality is not serious enough to affect the analysis. However, sometimes the predictor variables are so strongly intercorrelated that the regression results are ambiguous.

### *The best four-variable model*

The best four-variable model is found to contain $^{3}\chi^{shape}$ index in addition to other three parameters ($^{0}\chi^{v}$, FPSA1 and FNSA1). The improvement in the statistics is considerably high:

**Table 3:** Calculated values of topological indices

| Compound no. | DPSA1 | DPSA2 | DPSA3 | FPSA1 | FPSA2 | FPSA3 | FNSA1 | FNSA2 | FNSA3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | −120.080 | 124.3540 | 18.6267 | 0.3381 | 0.1134 | 0.0098 | 0.6618 | −0.2219 | −0.0404 |
| 2 | −164.560 | 110.7780 | 12.9824 | 0.2795 | 0.0830 | 0.0080 | 0.7205 | −0.2140 | −0.0268 |
| 3 | −212.970 | 367.1540 | 30.0995 | 0.3156 | 0.2006 | 0.0099 | 0.684 | −0.4351 | −0.0422 |
| 4 | −34.242 | 383.8070 | 28.4308 | 0.4741 | 0.2747 | 0.0112 | 0.5258 | −0.3046 | −0.0316 |
| 5 | 20.246 | 518.3120 | 32.4606 | 0.5137 | 0.3605 | 0.0154 | 0.4862 | −0.3413 | −0.0286 |
| 6 | −141.170 | 333.5920 | 26.5706 | 0.3888 | 0.2045 | 0.0089 | 0.6112 | −0.3215 | −0.0330 |
| 7 | 109.626 | 339.3700 | 20.5271 | 0.5749 | 0.2665 | 0.0060 | 0.4251 | −0.1971 | −0.0220 |
| 8 | 83.530 | 467.2910 | 27.6942 | 0.5560 | 0.3489 | 0.0102 | 0.4439 | −0.2785 | −0.0270 |
| 9 | 131.385 | 343.8120 | 20.5188 | 0.5886 | 0.2729 | 0.0060 | 0.4114 | −0.1907 | −0.0217 |
| 10 | −24.097 | 326.3240 | 20.4090 | 0.4791 | 0.2703 | 0.0096 | 0.5208 | −0.2939 | −0.0257 |
| 11 | 24.143 | 336.5880 | 18.8664 | 0.5202 | 0.2938 | 0.0091 | 0.4797 | −0.2709 | −0.0225 |
| 12 | 4.591 | 347.9600 | 20.2757 | 0.5038 | 0.2843 | 0.0087 | 0.4962 | −0.2801 | −0.0242 |
| 13 | −123.310 | 426.1110 | 31.2625 | 0.3968 | 0.2831 | 0.0137 | 0.6032 | −0.4305 | −0.0387 |
| 14 | −144.230 | 399.0320 | 28.4331 | 0.3812 | 0.2505 | 0.0113 | 0.6187 | −0.4066 | −0.0355 |
| 15 | −171.730 | 133.7320 | 16.3766 | 0.2709 | 0.0967 | 0.0112 | 0.7290 | −0.2601 | −0.0325 |
| 16 | −193.080 | 116.7490 | 15.0380 | 0.2478 | 0.0756 | 0.0090 | 0.7522 | −0.2294 | −0.0303 |
| 17 | −42.110 | 130.7130 | 17.8999 | 0.4474 | 0.1463 | 0.0139 | 0.5526 | −0.1807 | −0.0309 |
| 18 | −180.710 | 231.8270 | 22.8058 | 0.2856 | 0.1571 | 0.0131 | 0.7144 | −0.3931 | −0.0409 |
| 19 | −6.567 | 139.5220 | 16.2585 | 0.4923 | 0.1621 | 0.0108 | 0.5077 | −0.1673 | −0.0275 |
| 20 | 20.265 | 150.5750 | 16.1519 | 0.5221 | 0.1723 | 0.0089 | 0.4778 | −0.1576 | −0.0265 |
| 21 | −3.348 | 142.7990 | 15.9238 | 0.4961 | 0.1642 | 0.0095 | 0.5038 | −0.1668 | −0.0274 |
| 22 | 49.595 | 153.7760 | 15.0706 | 0.5531 | 0.1825 | 0.0079 | 0.4468 | −0.1474 | −0.0244 |
| 23 | 16.184 | 144.7630 | 15.2822 | 0.5186 | 0.1721 | 0.0093 | 0.4814 | −0.1599 | −0.0258 |
| 24 | 60.795 | 159.5900 | 14.7384 | 0.5629 | 0.1857 | 0.0069 | 0.4371 | −0.1442 | −0.0236 |
| 25 | 3.0774 | 140.9630 | 14.8574 | 0.5037 | 0.1670 | 0.0071 | 0.4963 | −0.1646 | −0.0278 |
| 26 | −241.120 | 527.4260 | 42.5192 | 0.2795 | 0.2698 | 0.0164 | 0.7205 | −0.6952 | −0.0614 |
| 27 | 23.602 | 170.7320 | 13.9450 | 0.5258 | 0.1959 | 0.0073 | 0.4742 | −0.1767 | −0.0231 |
| 28 | 98.140 | 202.8290 | 15.8927 | 0.5898 | 0.2189 | 0.0093 | 0.4102 | −0.1522 | −0.0199 |
| 29 | 84.521 | 197.8080 | 15.1447 | 0.5794 | 0.2151 | 0.0077 | 0.4206 | −0.1562 | −0.0208 |
| 30 | −96.575 | 326.3780 | 28.0063 | 0.4196 | 0.2281 | 0.0102 | 0.5804 | −0.3157 | −0.0364 |
| 31 | 80.993 | 163.2500 | 12.3099 | 0.5902 | 0.2149 | 0.0105 | 0.4097 | −0.1491 | −0.0169 |
| 32 | 142.847 | 184.6950 | 14.0391 | 0.6400 | 0.2319 | 0.0118 | 0.3599 | −0.1304 | −0.0157 |
| 33 | 157.114 | 198.0210 | 14.5710 | 0.6438 | 0.2334 | 0.0107 | 0.3562 | −0.1291 | −0.0159 |
| 34 | −6.771 | 236.9090 | 22.9268 | 0.4927 | 0.2530 | 0.0157 | 0.5073 | −0.2606 | −0.0339 |
| 35 | 92.236 | 329.1180 | 23.6838 | 0.5746 | 0.3059 | 0.0151 | 0.4254 | −0.2265 | −0.0231 |
| 36 | −83.755 | 72.3047 | 12.1093 | 0.3749 | 0.0810 | 0.0091 | 0.6251 | −0.1351 | −0.0271 |
| 37 | −132.110 | 82.1935 | 10.9387 | 0.3189 | 0.0718 | 0.0084 | 0.6811 | −0.1535 | −0.0215 |
| 38 | −39.108 | 78.2289 | 12.6221 | 0.4464 | 0.0958 | 0.0112 | 0.5536 | −0.1187 | −0.0233 |
| 39 | −26.243 | 84.8280 | 13.0000 | 0.4668 | 0.1002 | 0.0099 | 0.5332 | −0.1145 | −0.0229 |
| 40 | −147.970 | 119.9290 | 18.0853 | 0.2844 | 0.0994 | 0.0116 | 0.7156 | −0.2502 | −0.0411 |
| 41 | −168.250 | 104.8070 | 16.2910 | 0.2614 | 0.0777 | 0.0092 | 0.7385 | −0.2195 | −0.0370 |
| 42 | −176.330 | 90.4524 | 14.1153 | 0.2380 | 0.0640 | 0.0077 | 0.7619 | −0.2048 | −0.0342 |
| 43 | −45.635 | 74.0565 | 10.5483 | 0.4334 | 0.0938 | 0.0068 | 0.5666 | −0.1226 | −0.0240 |
| 44 | −316.150 | 223.7530 | 19.7742 | 0.1419 | 0.0719 | 0.0091 | 0.8581 | −0.4350 | −0.0357 |
| 45 | −259.268 | 187.3254 | 18.0124 | 0.1964 | 0.0861 | 0.0102 | 0.8036 | −0.3526 | −0.0319 |
| 46 | −38.621 | 99.2439 | 18.8740 | 0.4481 | 0.1195 | 0.0158 | 0.5518 | −0.1471 | −0.0348 |
| 47 | −18.301 | 108.9190 | 19.5020 | 0.4780 | 0.1270 | 0.0140 | 0.5220 | −0.1390 | −18.3010 |

$$\log K_i = -591050.4814 - 0.2167(\pm 0.0343)^{0}\chi^{v} + 591054.3330(\pm 274230.9659)\text{FPSA1} + 591053.4925(\pm 274230.9623)\text{FNSA1} + 0.1465(\pm 0.0785)^{3}\chi^{\text{shape}} \quad (4)$$

$n = 47, R^2 = 0.7496, R^2A = 0.7258, \text{SE} = 0.1528, F = 31.439$

Here, in addition to FPSA1 and FNSA1, $^{3}\chi^{\text{shape}}$ also has positive effect on the exhibition of log $K_i$.

### *The best five-variable model*

The best five-variable model exhibited significant improvement in the statistics. This model is found below:

$$\log K_i = -440911.9505 - 0.3566(\pm 0.0711)^{0}\chi^{v} + 440915.5590(\pm 258901.1026)\text{FPSA1} + 440915.1279(\pm 258901.0510)\text{FNSA1} + 0.1762(\pm 0.0557)^{3}\chi^{\text{shape}} - 0.0279(\pm 0.0099)\text{DPSA3} \quad (5)$$

**Table 4:** Summary of results obtained from variable selection in multi-variable modelling

| Model no. | Parameters used | $R^2$ | $R^2A$ | CV | $F$ |
|---|---|---|---|---|---|
| 1 | $^0\chi^v$ | 0.6233 | 0.6150 | 0.1811 | 74.472 |
| 2 | $^0\chi^v$, FPSA1 | 0.7020 | 0.6684 | 0.1629 | 68.563 |
| 3 | $^0\chi^v$, FPSA1, FNSA1 | 0.7290 | 0.7101 | 0.1571 | 38.563 |
| 4 | $^0\chi^v$, FPSA1, FNSA1, $^3\chi^{shape}$ | 0.7496 | 0.7258 | 0.1528 | 31.439 |
| 5. | $^0\chi^v$, FPSA1, FNSA1, $^3\chi^{shape}$, DPSA3 | 0.7890 | 0.7633 | 0.1420 | 30.669 |
| 6. | $^0\chi^v$, FPSA1, FNSA1, $^3\chi^{shape}$, FNSA2, DPSA3 | 0.7952 | 0.7656 | 0.1413 | 26.039 |
| 7. | $^0\chi^v$, FPSA1, FNSA1, $^2\chi$, $^3\chi^{shape}$, FNSA2, DPSA3 | 0.8062 | 0.7714 | 0.1395 | 23.172 |
| 8. | $^2\chi$, $^3\chi$, $^2\chi^v$, $^3\chi^{shape}$, PNSA2, DPSA3, FPSA1, FNSA1 | 0.8343 | 0.7995 | 0.1307 | 23.920 |
| 9. | $^2\chi$, $^3\chi$, $^2\chi^v$, $^3\chi^{shape}$, FPSA1, FNSA1, PNSA1, PNSA2, DPSA3 | 0.8375 | 0.7980 | 0.1312 | 21.191 |

$$n = 47, R^2 = 0.7890, R^2A = 0.7633, \text{SE} = 0.1420, F = 30.669$$

Such an improvement in the quality of the model is because of the added parameter DPSA3, the negative coefficient of which indicates that a decrease in the magnitude of DPSA3 increases log $K_i$. This parameter, *i.e.* DPSA3 is defined as the difference between total charge weighted partial positively charged molecular surface area and total charge weighted partial negatively charged molecular surface area i.e. DPSA3 = PPSA3–PNSA3.

### *Further higher parametric models*

Only a slight improvement in statistics occurs when we go in for six- and seven-parametric regression analysis. However, for an eight-parametric regression, there is considerable improvement in the statistics, such that $R^2 = 0.7890$ for five-parametric models increases to $R^2 = 0.8343$ and that for nine-parametric model, the $R^2$ is found to be 0.8375. Looking to the smaller number of descriptors, we considered that the eight-parametric model is the most appropriate model for modelling log $K_i$. Thus eight-parametric model is as found below:

$$\begin{aligned}\log K_i = {} & 534095.2166 + 0.5421(\pm 0.1363)^2\chi \\ & - 0.6371(\pm 0.1267)^3\chi - 0.3663(\pm 0.0783)^2\chi^v \\ & - 0.1761(\pm 0.0777)^3\chi^{shape} - 0.0060(\pm 0.0025)\text{PNSA} \\ & - 2 - 0.0535(\pm 0.0185)\text{DPSA} - 3 \\ & + 534099.9358(\pm 239377.0497)\text{FPSA} - 1 \\ & + 534099.9094(\pm 239377.0614)\text{FNSA} - 1 \qquad (6)\end{aligned}$$

$$n = 47, R^2 = 0.8343, R^2A = 0.7995, \text{SE} = 0.1307, F = 23.920$$

### *Optimum number of descriptors*

To know the optimum number of descriptors to be used for modelling log $K_i$, we plotted a graph between the number of descriptors against corresponding values of $R^2$ and $R^2A$ on the same graph paper (Figure 2). These curves became parallel to x-axis (number of descriptors used), when the number of descriptors eight indicates that the maximum of eight descriptors can be used for modelling log $K_i$.

### *Comments on the occurrence of both the parameters FPSA1 and FNSA1 in the same model*

In all the models discussed above include the topological descriptors FPSA1 and FNSA1, which are perfectly collinear ($R^2 = 1$). We

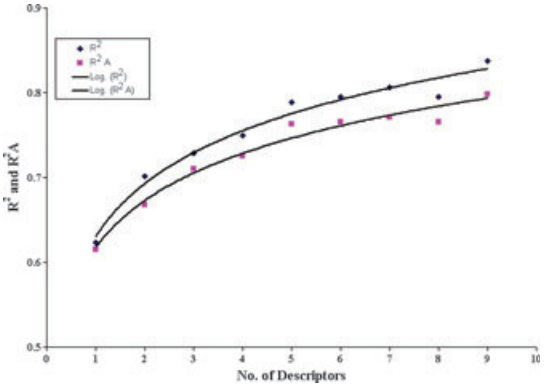


**Figure 2:** Correlation of number of descriptors with $R^2$ and $R^2A$.

observed that in going from eqn 2 to eqn 2, and so on, $R^2$ value increases by about 0.0270 units in each step. The difference in $R^2$ from the first model to the last model is about 0.11. So, we had to provide a stronger support for the inclusion of these two collinear descriptors in the models proposed by us, else we have to read of these descriptors. The best way to deal with such a problem is to calculate variance inflation factor (VIF). This is otherwise called the model inflation factor (MIF). We then discuss our results accordingly. This problem because of collinearigty is discussed in the following section.

### *On the occurrence of colinearity*

At this stage, it is worth examining the occurrence or otherwise of colinearity in the proposed models. The best candidates for this purpose would obviously be the eight- and nine-parametric models. This we can do in two different ways: (i) by examining the correlation matrices for the eight- and nine-parameteric models and/or (ii) by calculating VIF for each of the parameters in the model (35). The correlation matrices for the eight- and nine-parameteric models are given in Table 5 showing that both these models suffer from defect because of colinearity. To confirm this finding, we calculated VIF, which is a measure of multicollinearity, for each of the parameters involved in both these models. The VIF is defined as $1/(1-R_i^2)$, where $R_i$ is the multiple correlation coefficient of the ith independent variable on all of the other independent variables. A VIF 10 or more (no upper limit is defined) for large data sets indicates a collinearity problem. For small data sets, even VIFs of five or more

**Table 5:** Correlation matrix for models 8 and 9 (see Table 4)

| | log $K_i$ | $^2\chi$ | $^3\chi$ | $^2\chi^v$ | $^3\chi^{shape}$ | PNSA2 | DPSA3 | FPSA1 | FNSA1 | PNSA1 |
|---|---|---|---|---|---|---|---|---|---|---|
| log $K_i$ | 1.0000 | | | | | | | | | |
| $^2\chi$ | 0.7467 | 1.0000 | | | | | | | | |
| $^3\chi$ | 0.7765 | 0.9888 | 1.0000 | | | | | | | |
| $^2\chi^v$ | 0.7609 | 0.9142 | 0.8857 | 1.0000 | | | | | | |
| $^3\chi^{shape}$ | 0.6098 | 0.9116 | 0.8894 | 0.7572 | 1.0000 | | | | | |
| PNSA2 | 0.7186 | 0.7240 | 0.7425 | 0.6508 | 0.5689 | 1.0000 | | | | |
| DPSA3 | 0.6776 | 0.6505 | 0.6756 | 0.5275 | 0.5434 | −0.9389 | 1.0000 | | | |
| FPSA1 | 0.0400 | 0.3984 | 0.3894 | 0.2754 | 0.5480 | 0.2126 | −0.1277 | 1.0000 | | |
| FNSA1 | 0.0400 | 0.3984 | 0.3894 | 0.2754 | 0.5480 | −0.2126 | 0.1277 | 1.0000 | 1.0000 | |
| PNSA1 | 0.6878 | 0.5921 | 0.6033 | 0.5491 | 0.4614 | 0.9091 | 0.8120 | 0.4409 | 0.4409 | 1.0000 |

**Table 6:** VIF values for the eight-parametric model (Table 4); eq.(6)

| Independent variable | Variance inflation factor (VIF) | Tolerance |
|---|---|---|
| $^3\chi^{shape}$ | 11.0716 | 0.0903 |
| PNSA2 | 29.7714 | 0.0336 |
| DPSA3 | 13.3582 | 0.0749 |
| FPSA1 | $8.110^{10}$ | 0.0000 |
| FNSA1 | $8.110^{10}$ | 0.0000 |
| $^2\chi$ | 121.3337 | 0.0082 |
| $^3\chi$ | 65.9756 | 0.0152 |
| $^2\chi^v$ | 10.7513 | 0.0930 |

**Table 7:** VIF values for the nine-parametric model (Table 4)

| Independent variable | Variance inflation (VIF) | Tolerance |
|---|---|---|
| $^3\chi^{shape}$ | 20.2160 | 0.0495 |
| PNSA1 | 30.0504 | 0.0333 |
| PNSA2 | 30.7217 | 0.0306 |
| DPSA3 | 13.5779 | 0.0736 |
| FPSA1 | $8.180^{10}$ | 0.0000 |
| FNSA 1 | $8.180^{10}$ | 0.0000 |
| $^2\chi$ | 125.7524 | 0.0080 |
| $^3\chi$ | 76.9505 | 0.0130 |
| $^2\chi^v$ | 10.7519 | 0.0930 |

(here also no upper limit is defined) can signify collinearity. The variables with a high VIF are candidates for exclusion from the model. The VIF values for eight- and nine-parametric models are presented in Tables 6 and 7, respectively. These tables also record the values of yet another parameter called tolerance. This is also a parameter used for investigating collinearity problem. It is just the denominator of VIF. As can be seen from these tables, there are some parameters whose VIF values are much larger than 10. Statistically, therefore, multicollinearity is a problem with these models. The Ridge regression data presented in Figures 3 and 4 further support the occurrence of multicollinearity in these eight- and nine-parametric models. Thus, to get rid of such abuse, we need to delete the parameters having largest VIF values in succession so that ultimately we obtain a model free from collinearity defect. The deletion of the parameters having larger values of VIF yielded a tetra-parameters model containing $^2\chi^v$, $^3\chi^{shape}$, PNSA1, DPSA3 as the correlating parameters and all these involved parameters now have VIF values smaller than 10:

| Parameter | VIF |
|---|---|
| $^2\chi^v$ | 2.6769 |
| $^3\chi^{shape}$ | 2.5835 |
| PNSA1 | 3.2080 |
| DPSA3 | 3.3055 |

Now, all the correlating parameters have VIF <10 and thus there is no colinearity problem. Multiple regressions performed using these four parameters yielded the following model.

$$\log K_i = 3.458 - 0.1944(\pm 0.0500)^2\chi^v + 0.0079(\pm 0.0480)^3\chi^{shape} - 0.0017(\pm 0.0012)\text{PNSA1} - 0.0178(\pm 0.0118)\text{DPSA3} \quad (7)$$

$$n = 47, R^2 = 0.7003, R^2A = 0.6717, \text{SE} = 0.1672, F = 24.531$$

In this model, the coefficient of $^3\chi^{shape}$ is much smaller than its standard deviation. Such models are not allowed statistical. The trial and error procedure adopted by us indicated that $^3\chi^{shape}$ can be replaced by $^0\chi^v$ yielding a statistically allowed four-parametric model as below:

$$\log K_i = 3.4617 - 0.1036(\pm 0.0858)^2\chi^v - 0.0478(\pm 0.0435)^0\chi^v - 0.0017(\pm 0.0012)\text{PNSA1} - 0.0133(\pm 0.0116)\text{DPSA3} \quad (8)$$

$$n = 47, R^2 = 0.7085, R^2A = 0.6807, \text{SE} = 0.1649, F = 25.515$$

When we calculated VIF values for each of these four-parameters (eqn 8), we obtained the following results:

| Parameter | VIF |
|---|---|
| $^2\chi^v$ | 9.4406 |
| $^0\chi^v$ | 8.1065 |
| PNSA1 | 3.1255 |
| DPSA3 | 3.3289 |

All the VIF values are <10 and this new model (eqn 8) is also free from defect caused by colinearity. We observe that using VIF vis-a-

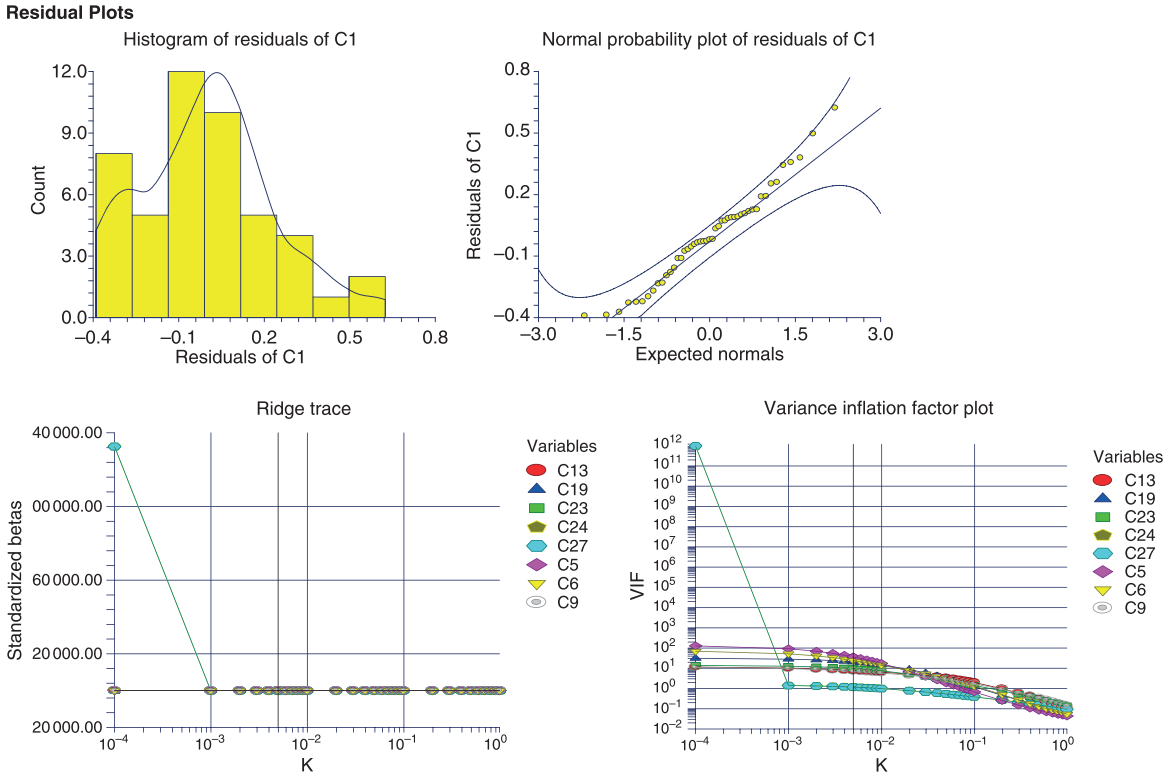


**Figure 3:** Ridge statistics for the 8-parametric model.

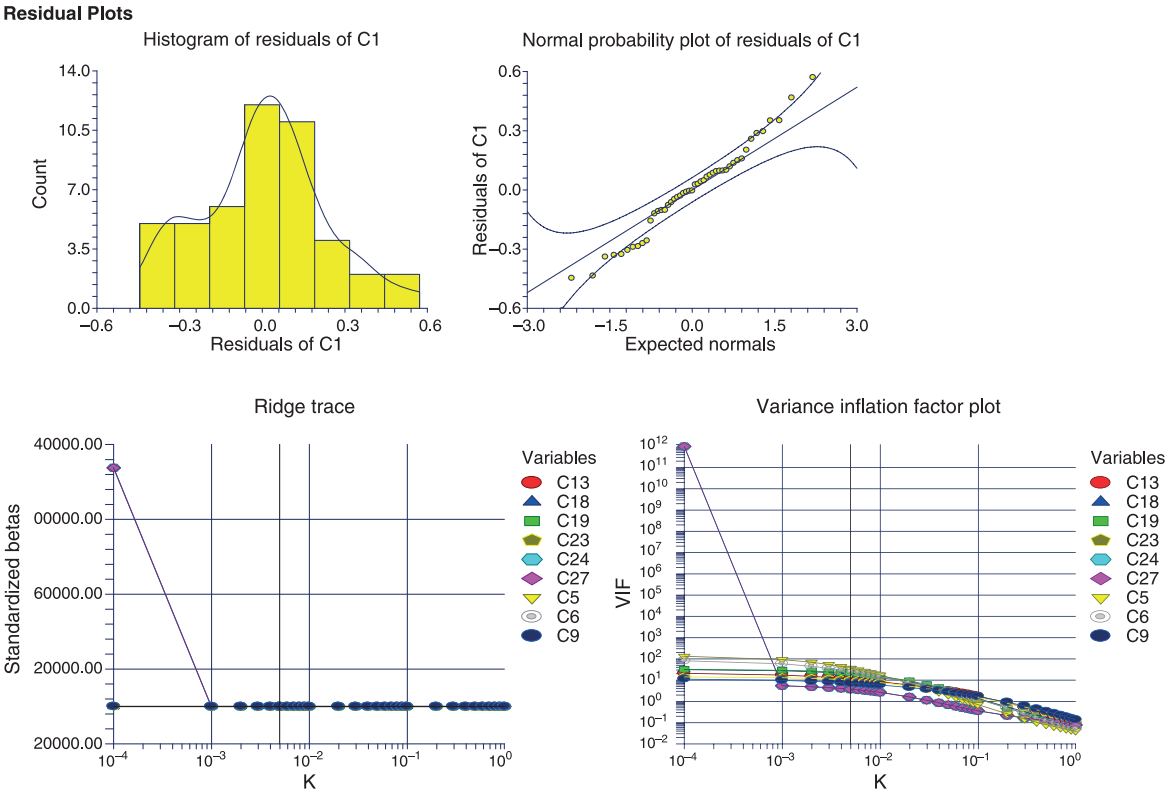


**Figure 4:** Ridge statistics for the 9-parametric model.

vis MIF statistics, the resulting models do not contain collinear parameters, namely FPSA1 and FNSA1 and are completely free from the defect caused by collinearity.

### *Comparison of our results with those of earlier reported model by Melagraki and other (30)*

It is not possible to make an exact comparison of our results with those of the earlier worker (30). The earlier study has indicated that the tetra-parametric model contained $^{1}\chi_{inf}$, $^{0}\chi^{v}_{inf}$, $^{1}\chi^{v}_{inf}$ and N-rings as the correlating parameters is the best. However, earlier authors did not calculate VIF values of these parameters. Now, we report these earlier undetermined VIF values below. As the earlier tetra parametric model was found the best, we first compared our results of tetra-parametric models with it. Such a comparison is shown in the following table:

| Model | Parameters | VIF | $R^2$ | $R^2A$ | SE | $F$ |
|---|---|---|---|---|---|---|
| Old model | $^{1}\chi_{inf}$ | 1.1057 | 0.7283 | 0.7024 | 0.2632 | 28.140 |
| | $^{0}\chi^{v}_{inf}$ | 2.5420 | | | | |
| | $^{1}\chi^{v}_{inf}$ | 1.7498 | | | | |
| | N-rings | 1.7181 | | | | |
| Our model | $^{0}\chi^{v}$ | 9.4406 | 0.7085 | 0.6807 | 0.1649 | 25.515 |
| | $^{2}\chi^{v}$ | 8.1065 | | | | |
| | PNSA1 | 3.1215 | | | | |
| | DPSA3 | 3.3289 | | | | |
| Mixed model | $^{0}\chi^{v}$ | 1.7761 | 0.7521 | 0.7285 | 0.1521 | 31.853 |
| | DPSA3 | 1.9827 | | | | |
| | $^{1}\chi_{inf}$ | 1.1600 | | | | |
| | $^{1}\chi^{v}_{inf}$ | 1.0743 | | | | |

This comparison shows that all the three tetra-parametric models mentioned in the above table have VIF values smaller than 10 and thus they are all free from the defect because of colinearity. A close examination of the above table shows that the choice of the descriptors used by us yields a model inferior to the earlier model (30) using topological information indices. However, when we mixed the descriptors used by us with those used earlier (30), we obtained an excellent tetra-parametric model superior to the old. It seems that N-ring and $^{0}\chi^{v}_{inf}$ parameters are not that good for modelling log $K_i$. Instead, $^{0}\chi^{v}$ and DPSA-3 are more suitable for this purpose. Thus, we conclude that Kier and hall topological as well as Kier and Hall information topological indices play a dominating role for modelling log $K_i$. Hence, our attempt at mix-modelling is far superior to the earlier attempt (30). This comparison is not enough as the earlier study provided results mono- to penta-parametric models. Therefore, we compared statistics of mono- to penta-variable models in both the cases. Such a comparison is demonstrated in the following table in which the earlier reported values are shown in the parenthesis, while our results are given in bold.

| Model | $R^2$ | SE | $F$ |
|---|---|---|---|
| *Monoparametric* | **0.6233** | **0.1811** | **74.472** |
| *Biparametric* | **0.7020** | **0.1629** | **51.814** |
| | (0.4909-0.6642) | (0.3603-0.2926) | (43.52) |
| *Triparametric* | **0.7290** | **0.1571** | **36.563** |
| | (0.5275-0.6984) | (0.3470-0.2773) | (33.18) |
| *Tetraparametric* | **0.7496** | **0.1528** | **31.439** |
| | (0.6055-0.7283) | (0.3171-0.2632) | (28.14) |
| *Pentaparametric* | **0.7890** | **0.1420**. | **30.669** |
| | (0.7296) | (0.2625) | (22.12) |

The aforementioned comparison (**values in bold**) finally establishes that in total, the methodology used by us is much superior to the earlier used methodology.

### *Correlation between observed and calculated log* $K_i$

The aforementioned results and discussion indicated that the eight-parametric model is the best for modelling log $K_i$ and that the nine-parametric model exhibited only a slight improvement in the statistics. To confirm this, we calculated log $K_i$ using these two models and compared them with the observed (experimental) log $K_i$. Such a comparison is demonstrated in Table 5 and demonstrated in Figures 5 and 6, respectively. We also observed here that only a small improvement occurred in $R^2$, when we passed from eight- to nine-parametric model thus yielding $R^2 = 0.8343$ and 0.8376, respectively (see Figures 5 and 6). These results are, therefore, in favour of eight-parametric model, proposed by us. It is clear that the models 8 and 9 (Table 8) are the best models and that in both the cases, the residuals are closer to zero (Table 5). Normally, the residuals greater than twice the standard

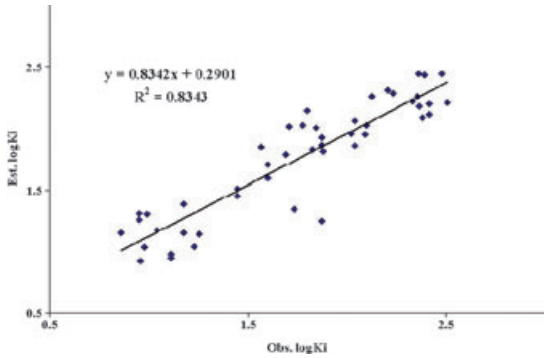


**Figure 5:** Correlation of observed and calculated (estimated) activity (log $K_i$) using model 8.

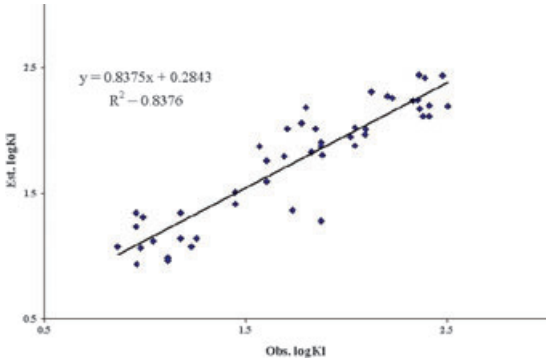


**Figure 6:** Correlation of observed and calculated (estimated) activity (log $K_i$) using model 9 (see Table 4).

**Table 8:** Comparison of observed and calculated activity (log $K_i$) using models 8 and 9 (from Table 4)

| | | Model (8) | | Model (9) | |
|---|---|---|---|---|---|
| Compound no. | Obs. | Cal. | Res. | Cal. | Res. |
| 1 | 2.412 | 2.11 | 0.302 | 2.11 | 0.302 |
| 2 | 2.093 | 1.948 | 0.145 | 1.969 | 0.124 |
| 3 | 1.114 | 0.950 | 0.164 | 0.966 | 0.148 |
| 4 | 1.176 | 1.390 | −0.214 | 1.344 | −0.168 |
| 5 | 0.954 | 1.258 | −0.303 | 1.236 | −0.282 |
| 6 | 0.863 | 1.151 | −0.288 | 1.075 | −0.211 |
| 7 | 1.041 | 1.171 | −0.129 | 1.117 | −0.075 |
| 8 | 1.255 | 1.144 | 0.111 | 1.140 | 0.115 |
| 9 | 1.176 | 1.151 | 0.025 | 1.139 | 0.037 |
| 10 | 1.826 | 1.826 | 0.000 | 1.826 | 0.000 |
| 11 | 1.732 | 1.348 | 0.384 | 1.365 | 0.368 |
| 12 | 0.991 | 1.305 | −0.313 | 1.310 | −0.319 |
| 13 | 0.978 | 1.037 | −0.059 | 1.065 | −0.088 |
| 14 | 0.959 | 0.925 | 0.034 | 0.932 | 0.027 |
| 15 | 1.708 | 2.013 | −0.305 | 2.013 | −0.306 |
| 16 | 1.881 | 1.811 | 0.070 | 1.802 | 0.079 |
| 17 | 2.391 | 2.432 | −0.041 | 2.417 | −0.026 |
| 18 | 2.124 | 2.260 | −0.136 | 2.307 | −0.183 |
| 19 | 2.366 | 2.178 | 0.187 | 2.170 | 0.195 |
| 20 | 2.356 | 2.258 | 0.098 | 2.243 | 0.113 |
| 21 | 2.412 | 2.197 | 0.214 | 2.197 | 0.215 |
| 22 | 2.330 | 2.221 | 0.110 | 2.239 | 0.091 |
| 23 | 2.362 | 2.444 | −0.082 | 2.443 | −0.082 |
| 24 | 1.799 | 2.142 | −0.342 | 2.183 | −0.384 |
| 25 | 1.568 | 1.849 | −0.281 | 1.874 | −0.306 |
| 26 | 1.230 | 1.039 | 0.191 | 1.073 | 0.157 |
| 27 | 2.380 | 2.087 | 0.294 | 2.114 | 0.267 |
| 28 | 2.021 | 1.956 | 0.065 | 1.948 | 0.073 |
| 29 | 1.875 | 1.864 | 0.011 | 1.903 | −0.028 |
| 30 | 1.114 | 0.978 | 0.136 | 0.982 | 0.132 |
| 31 | 1.690 | 1.788 | −0.098 | 1.792 | −0.101 |
| 32 | 1.602 | 1.595 | 0.007 | 1.593 | 0.009 |
| 33 | 1.447 | 1.508 | −0.061 | 1.510 | −0.063 |
| 34 | 0.954 | 1.311 | −0.356 | 1.345 | −0.391 |
| 35 | 1.875 | 1.249 | 0.627 | 1.279 | 0.596 |
| 36 | 2.477 | 2.442 | 0.035 | 2.437 | 0.041 |
| 37 | 2.505 | 2.210 | 0.295 | 2.193 | 0.312 |
| 38 | 2.230 | 2.285 | −0.054 | 2.257 | −0.027 |
| 39 | 2.204 | 2.312 | −0.108 | 2.273 | −0.069 |
| 40 | 1.778 | 2.026 | −0.248 | 2.059 | −0.281 |
| 41 | 2.041 | 1.857 | 0.185 | 1.879 | 0.163 |
| 42 | 1.602 | 1.706 | −0.104 | 1.759 | −0.157 |
| 43 | 1.845 | 2.004 | −0.159 | 2.010 | −0.165 |
| 44 | 1.447 | 1.456 | −0.009 | 1.415 | 0.032 |
| 45 | 1.875 | 1.924 | −0.049 | 1.871 | 0.004 |
| 46 | 2.097 | 2.026 | 0.070 | 2.006 | 0.091 |
| 47 | 2.041 | 2.060 | −0.019 | 2.021 | 0.020 |

division are considered as outliers. Therefore, it seems likely that the compound 35 lies like an outlier. However, deletion of this compound from the process of regression did not show any significant improvement in the statistics.

From Table 8, we observed that $R^2A$ goes on increasing as we pass from one- to nine-variable models. This means that in each case, the added parameters in progressive regression have enough contribution towards the activity. This, therefore, justifies our attempt to investigate multi-parametric regression up to nine-parametric model. Generally, $R^2$ increases with an increase in the number correlating parameters; however, this is not the case with $R^2A$. In the case of $R^2A$, with added descriptors, $R^2A$ will decline if the new descriptors do not have enough contribution towards the model (34,35,40).

### *Model based on the combination of parameters used by us with the parameters used earlier*

With a hope to obtain better model, we mixed the earlier parameters with the parameters used in this study. The step-wise regression analysis using the method of maximum–$R^2$ indicated that here also four parameters are needed to model log $K_i$, excellently, the parameters used being: ${}^0\chi^v$, ${}^1\chi_{inf}$, ${}^1\chi^v_{inf}$, DPSA-3. In this model, ${}^0\chi^v$ and DPSA-3 are the parameters which we used in this study and the remaining two parameters are from an earlier study (30). This model is found to be the most appropriate for modelling log $K_i$:

$$\log K_i = 2.1604 - 0.1080(\pm 0.0174)\,{}^0\chi^v - 0.0227(\pm 0.0083)\mathrm{DPSA3} - 0.3716(\pm 0.1016)\,{}^1\chi_{inf} + 0.7092(\pm 0.2458)\,{}^1\chi^v_{inf} \quad (9)$$

$$n = 47, R^2 = 0.7521, R^2A = 0.7285, \mathrm{SE} = 0.1521, F = 31.853$$

To decide which methodology yields a better model, it is necessary to compare all the four-parametric models discussed above.

### *Randic recommendations*

Randic (32,33) stated that 'the selection of descriptors to be used in structure-property-activity studies should not be delegated solely to the computers, although the statistical criteria will continue to be useful for preliminary screening of descriptors taken from a large pool. Often in an automated selection of descriptors, a descriptor will be discarded because it is highly correlated with another descriptor already selected. But what is important is not whether the two descriptors parallel each other, *i.e.* duplicate much of the same structural information, but whether they are in those parts that are important for structure–property–activity correlations. If they differ in the domain which is important for the property/activity considered, both descriptors should be retained; if they differ in parts that are not relevant for the correlation of the considered property/activity, one of them can be discarded. Hence, the residual of the correlation between two descriptors should be examined and kept or discarded depending on how well it can improve the correlation based on already selected descriptors'.

Randic (32,33) further 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 interrelated descriptors differ in some other structural traits. The difference between them may be relatively small, but nevertheless very important for structure–property regression'. Randic further argued that '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'.

Randic (32) gave an example of a model where two variables are highly correlated ($R > 0.98$). When single descriptor models were built ($R$s<0.117), they were not able to predict the molar refraction. However, there is an improvement on the prediction when the two descriptors are combined in a two variable model ($R = 0.971$). In this case, it is clear that the collinearity is not factor in the construction of the model and that the R values increase when collinearity is allowed, *i.e.* this example is statistically justified.

In another paper, Peterangelo and Seybold (Int. J. Quantum Chem. 2004, 96, 1–9) showed four examples where Randic's suggestion was explored. In all of them, the single variable models present $R$'s < 0.3633, and the two variable models present $R$'s > 0.8471. In these examples, it is also clear that the collinearity effect is statistically allowed.

Perfect multicollinearity occurs when one of the independent variables in a regression equation is perfectly correlated with another variable (i.e. $R = \pm 1.000$). One of the problems with multicollinearity is that, in this situation, it is impossible to calculate the least-squares. Another problem with multicollinearty is that it increases the standard errors associated with the individual regression coefficients.

As mentioned earlier, in the present case, the topological descriptors FPSA1 and FNSA1 are perfectly collinear ($R^2 = 1$), and our data show that in going from eqn 2 to eqn 3, and so on, the $R^2$ value increases by about 0.0270 units in each step. The difference in $R^2$ from the first model to the last model is about 0.11 (very far from the increases on $R^2$ shown in either Randic's or Peterangelo's examples).This, therefore, very much justifies the application of VIF/MIF statistics and the consequent development of models free from colinearity defect. All this happened by considering very high values of collinear parameters: FPSA1 and FNSA1.

### *Cross-validation*

It is worth mentioning that models with excellent statistics need not necessarily mean that they possess excellent prediction capacity also. To be an excellent model, it should have excellent prediction capacity also. The prediction capacity can be judged in two different ways: (i) by estimating Pogliani's quality factor, Q (ii) by evaluating cross-validated parameters. The use of Q factor is questioned and, therefore, we used cross-validated parameters for estimating prediction capacity of the proposed models.

In principle, 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 data are 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:

$$\text{PRESS} = \sum_{y} (Y_{est} - Y_{obs})^2 \tag{10}$$

$$\text{SSY} = \sum_{y} (Y_{obs} - Y_{mean})^2 \tag{11}$$

$$r^2_{cv} = q^2 = 1.0 - \frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{\sum_{i=1}^{n} (Y_{obs} - Y_{mean})^2} \tag{12}$$

$$S_{PRESS} = S_{cv} = \sqrt{\frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{N - M - 1}} \tag{13}$$

$$\text{PSE} = S_{pred} = \sqrt{\frac{\sum_{i=1}^{n} (Y_{obs} - Y_{est})^2}{N}} \tag{14}$$

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 ≥0.9. The actual predictive ability (predictive power) of the model is validated using an external prediction set (41). 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 show that except for the one-variable model, all other models are reliable models. The $S_{PRESS}$ as well as PSE are good parameters to be used for discussing the uncertainty in prediction. The lower the value of

**Table 9:** Cross-validated parameters for the proposed models

| Model | PRESS | SSY | PRESS/SSY | $r^2_{CV}$ | $S_{PRESS}$ | PSE |
|---|---|---|---|---|---|---|
| 1 | 4.5128 | 7.4684 | 0.6042 | 0.3957 | 0.3166 | 0.3098 |
| 2 | 3.5709 | 8.4103 | 0.4246 | 0.5754 | 0.2848 | 0.2756 |
| 3 | 3.2465 | 8.7347 | 0.3716 | 0.6283 | 0.2747 | 0.2628 |
| 4 | 2.9997 | 8.8915 | 0.3339 | 0.6660 | 0.2672 | 0.2526 |
| 5 | 2.5276 | 9.4536 | 0.2673 | 0.7326 | 0.2482 | 0.2319 |
| 6 | 1.9848 | 9.9964 | 0.1985 | 0.8014 | 0.2285 | 0.2055 |

these parameters, the better will be the predictive ability of the model. A perusal of Table 9 shows that both these parameters go on decreasing as we pass from one- to eight-variable models and that it is the lowest for the model 8. Hence, once again, we find that the most appropriate model for modeling log $K_i$ (hCA-II) is this eight-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 three-parametric model, all other higher parametric models have PRESS/SSY <0.4 thereby indicating them to be reasonable models. This ratio for the eight-parametric model is more or less nearer 0.1 indicating it to have the best prediction capacity.

### *Experimental*

- *Inhibitory activity*. The inhibitory activity (log $K_i$) for the set of 47 compounds was adopted from the earlier work of the authors (Supuran) (31).

- *Topological indices*. Twenty-nine topological indices used in the present study were either calculated using DRAGON software[a] or KARELSON and CHEMAXON software[b]. The structure optimization was made using HYPERCHEM[c] and ACD LABS[d] softwares.

- *Regression analysis*. The step-wise regression analysis based on the method of maximum-$R^2$ (35–37) was proposed using NCSS[e] software. The meaning of the topological indices are given in Appendix A.

## Acknowledgements

One of the authors, Shalini Singh expresses her thanks to the Department of Science & Technology, Government of India, New Delhi, for awarding D ST project SR/WOS-A/CS/61/2004 under Woman Scientists scheme and to Principal for his interest and for providing facility to carry out this work.

## References

1. Supuran C.T., Scozzafava A., Casini A. (2004) Development of sulfonamide carbonic anhydrase inhibitors (CAIs). In: Supuran C.T., Scozzafava A., Conway J., editors. Carbonic Anhydrase, Its Inhibitors and Activators. Boca Raton (FL): CRC Press;p. 67.
2. Clare B.W., Supuran C.T. (2006) A perspective on quantitative structure-activity relationships and carbonic anhydrase inhibitors. Expert Opin Drug Metab Toxicol;2:113–137.
3. Wiener H. (1947) Structural determination of paraffins boiling points. J Am Chem Soc;69:17–20.
4. Randic M. (1975) On characterization of molecular branching. J Amer Chem Soc;97:6609–6615.
5. Randic M. (2001) The coonectivity index 25 years The connectivity index 25 years after. J Mole Graph Model;20:19–35.
6. Kier L.B., Hall L.H., Murray W.J., Randic M. (1975) Molecular connectivity. I: Relation to non-specific anesthesia. J Pharm Sci;64:1971–1974.
7. Kier L.B., Hall L.H. (1976) Molecular Connectivity in Chemistry and Drug Research. New York: Academic Press.
8. Khadikar P.V., Deshpande N.V., Kale P.P., Dobrynin A., Gutman I., Domotor G. (1995) The szeged index and an analogy with the wiener index. J Chem Inf Comput Sci;35:547–550.
9. Gutman I., Khadikar P.V., Rajput P.V., Karmarkar S. (1995) The szeged index of polyacenes. J Serb Chem Soc;60:759–764.
10. Khadikar P.V., Karmarkar S., Agrawal V.K., Singh J., Shrivastava A., Lukovits I., Diudea M.V. (2005) Szeged index – applications for drug modeling. Lett Drug Des Discov;2:606–624.
11. Gutman I., Klavzar S. (1995) An algorithm for the calculation of the szeged index of benzenoid hydrocarbons. J Chem Inf Comput Sci;35:1011–1014.
12. Khadikar P.V. (2000) On a novel structural descriptor PI. Nat Acad Sci Lett;23:113–118.
13. Khadikar P.V., Karmarkar S., Agrawal V.K. (2001) A Novel PI index and its applications to QSPR/QSAR studies. J Chem Inf Comput Sci;41:934–949.
14. Khadikar P.V., Kale P.P., Deshpande N.V., Karmarkar S., Agrawal V.K. (2001) Novel PI indices of hexagonal chains. J Math Chem;29:143–150.
15. Khadikar P.V., Karmarkar S., Varma R.G. (2002) On the estimation of PI index of polyacenes. Acta Chim Slov;49:755–771.
16. Ashrafi A.R., Loghman A. (2006) PI index of some benzenoid graph. J Chilean Chem Soc;51:968–970.
17. Manoochehrian B., Yousefi-Azari H., Asharfi A.R. PI polynomial of some benzenoid graphs. Commu Math Comput Che (MATCH) In Press.
18. John P.E., Khadikar P.V., Singh J. (2007) A method for computing the pi index of benzenoid hydrocarbons using orthogonal cuts. J Math Chem;42:37–47.
19. Khadikar P.V., Diudea M., Singh J., John P., Shrivastava A., Singh S., Karmarkar S., Lakhwani M., Thakur P. (2006) Use of PI index in computer-aided designing of bioactive compounds. Curr Bioact Comp;2:19–56.
20. Singh J., Lakhwani M., Khadikar P.V., Agrawal V.K., Balaban A.T., Clare B.W., Supuran C.T. (2006) QSAR study on the inhibition of the human cytosolic isozyme VII. Rev Roum Chim;51:691–701.
21. Khadikar P.V., Clare B.W., Balaban A.T., Supuran C.T., Agrawal V.K., Singh J., Joshi A.K., Lakhwani M. (2006) QSAR predection of CAI, CAII, CAIV inhibitory activities: relative potential of balaban and balaban type indices. Rev Roum Chem;51:703–717.
22. Balaban A.T., Khadikar P.V., Supuran C.T., Thakur A., Thakur M. (2005) Study on supramolecular complexing ability vis-à-vis estimation of pKa of substituted sulfonamides: dominating role of Balaban index (J). Bioorg Med Chem Lett;15:3966–3973.

23. Singh J., Lakhwani M., Khadikar P.V., Agrawal V.K., Supuran C.T. (2006) QSAR study on murine recombinant isozyme mCAXIII: topological vs structural descriptors. Arkivoc;14:103–118.
24. Singh S., Singh J., Ingle M., Mishra R., Khadikar P.V. (2006) A QSAR study on carbonic anhydrase inhibition: predicting log-Ki(hCAI) using of NMR chemical shift of ($SO_2NH_2$ as a molecular descriptor. Arkivok;16:1–15.
25. Khadikar P.V., Deeb O., Jaber A., Singh J., Lakhwani M. (2006) Development of quantitative structure–activity relationship for a set of carbonic anhydrase inhibitors: use of quantum and chemical descriptors. Lett Drug Des Discovery;3:622–635.
26. Agrawal V.K., Singh J., Khadikar P.V., Supuran C.T. (2006) QSAR study on topically acting sulfonamides incorporating GABA moieties: a molecular connectivity approach. Bioorg Med Chem Lett;16:2044–2051.
27. Thakur A., Thakur M., Khadikar P.V., Supuran C.T. (2005) QSAR study on pKa vis-à-vis physiological activity of sulfonamides: a dominating role of surface tension (inverse steric parameter). Bioorg Med Chem Lett;15:203–209.
28. Khadikar P.V., Sharma V., Karmarkar S., Supuran C.T. (2005) Novel use of chemical shift in NMR as molecular descriptor: a first report on modeling carbonic anhydrase inhibitory activity and related parameters. Bioorg Med Chem Lett;15:931–936.
29. Khadikar P.V., Sharma V., Karmarkar S., Supuran C.T. (2005) QSAR studies on benzene sulfonamide carbonic anhydrase inhibitors: need of hydrophobic parameter for topological modeling of binding constants of sulfonamides to human CA-II. Bioorg Med Chem Lett;15:923–930.
30. Jantschi L., Bolboaca S.D. (2006) Modeling the inhibitory activity of carbonic anhydrase IV of substituted thindizole and thiadiazoline disulphonamides: integrretion of structural information. Rev Electron Biomed/Electron J Biomed;2:22–33.
31. Melagraki G., Afantitis A., Sarimveis H., Igglessi-Marapouln O., Supuran C.T. (2006) QSAR study on para-substituted aromatic sulphonamides as carbonic anhydrase II inhibition using topological information indices. Bioorg Med Chem;14:1108–1114.
32. Balaban A.T., Basak S.C., Beteringhe A., Mills D., Supuran C.T. (2004) QSAR study using topological indices for inhibition of anhydrase II by sulphonamides and Schiff's base. Mol divers;8:401–412.
33. Randic M. (1997) On characterization of chemical structure. J Chem Inf Comput Sci;37:672.
34. Randic M. (1998) On characterization of molecular attributs. Acta Chim Slov;45:239.
35. Khadikar P.V., Mandloi D. (2004) QSAR study on binding constants of benzenesulfonamaides to human CA II: need or otherwise of hydrophobic parameter. Bioinformatics, India;2:86–91.
36. Chaterjee S., Hadi A.S., Price B. (2000) Regression Analysis by Examples, 3rd edn. New York: Wiley.
37. Diudea M.V., Florescu M.S., Khadikar P.V. (2006) Molecular Topology and Its Applications. Bucharest: EFICON.


## Notes

[a]DRAGON software for calculation of topological indices: http://www.disat.unimib.it

[b]Karelson, M. Molecular Descriptors in QSAR/QSPR, Wiely-Interscience, 2000 and CHEMAXON (http://www.chemaxon.com) softwares for the calculation of topological indices.

[c]HYPERCHEM-7 software for calculating the molecular modeling parameters; http://www.hyper.com

[d]ACD-LAB software for calculating the referred physicochemical parameters; CHEM SKETCH 3.0, http://www.acdlabs.com

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

## Appendix A

| Wiener index | *W* |
|---|---|
| Randic index (order 0) | $^0\chi$ |
| Randic index (order 1) | $^1\chi$ |
| Randic index (order 2) | $^2\chi$ |
| Randic index (order 3) | $^3\chi$ |
| Kier&Hall index (order 0) | $^0\chi^v$ |
| Kier&Hall index (order 1) | $^1\chi^v$ |
| Kier&Hall index (order 2) | $^2\chi^v$ |
| Kier&Hall index (order 3) | $^3\chi^v$ |
| Kier shape index (order 1) | $^1\chi^{shape}$ |
| Kier shape index (order 2) | $^2\chi^{shape}$ |
| Kier shape index (order 3) | $^3\chi^{shape}$ |
| Total molecular surface area [Empirical PC] | TMSA |
| Partial positive surface area [Empirical PC] | PPSA1 |
| Total charge weighted PPSA [Empirical PC] | PPSA2 |
| Atomic charge weighted PPSA [Empirical PC] | PPSA3 |
| Partial negative surface area [Empirical PC] | PNSA1 |
| Total charge weighted PNSA [Empirical PC] | PNSA2 |
| Atomic charge weighted PNSA [Empirical PC] | PNSA3 |
| Difference in CPSAs (PPSA1–PNSA1) [Empirical PC] | DPSA1 |
| Difference in CPSAs (PPSA2–PNSA2) [Empirical PC] | DPSA2 |
| Difference in CPSAs (PPSA3–PNSA3) [Empirical PC] | DPSA3 |
| Fractional PPSA (PPSA-1/TMSA) [Empirical PC] | FPSA1 |
| Fractional PPSA (PPSA-2/TMSA) [Empirical PC] | FPSA2 |
| Fractional PPSA (PPSA-3/TMSA) [Empirical PC] | FPSA3 |
| Fractional PNSA (PNSA-1/TMSA) [Empirical PC] | FNSA1 |
| Fractional PNSA (PNSA-2/TMSA) [Empirical PC] | FNSA2 |
| Fractional PNSA (PNSA-3/TMSA) [Empirical PC] | FNSA3 |