Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3966–3973
Bioorganic & Medicinal Chemistry Letters

Study on supramolecular complexing ability vis-à-vis estimation of pK_a of substituted sulfonamides: Dominating role of Balaban index (J)

Alexandru T. Balaban,[a] Padmakar V. Khadikar,[b,*] Claudiu T. Supuran,[c] Abhilash Thakur[d] and Mamta Thakur[e]

[a]*Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 77551, USA*
[b]*Research Division, Laxmi Fumigation and Pest Control, Pvt. Ltd, 3, Khatipura, Indore 452007, India*
[c]*University of Florence, Dipartimento di Chimica, Laboratoria di Chimica Bioinorganica, Via della Lastruccia 3, Rm. 188, 50019-Sesto Fiorentino, Firenze, Italy*
[d]*Department of Chemistry, Rishiraj Institute of Technology, Indore, India*
[e]*Department of Chemistry, Softvision College of Biotechnology and Science, Indore, India*

Received 1 March 2005; revised 6 May 2005; accepted 14 May 2005
Available online 19 July 2005

Abstract—The supramolecular complexing ability vis-à-vis pK_a estimation of a large series of 43 sulfonamides was made using a series of molecular descriptors including topological indices. The set of topological indices chosen also contains Balaban (J) and a variety of Balaban type indices: J, J_z, J_m, J_v, J_c, and J_p. The results have shown that the most discriminating Balaban index (J) in multi-parametric regression analysis combined with indicator parameters yields excellent models and also establishes the superiority of the J index over other Balaban type indices. The statistics is improved when one of the indicator parameters is replaced by molar volume (MV). The results are discussed critically using a variety of statistics.

1. Introduction

Although the story of sulfonamides started with the discovery of their antimicrobial action, subsequent studies established their usefulness as carbonic anhydrase inhibitors, diuretics, and anti-diabetics (insulin-releasers), and more recently also as endothelia antagonists.[1–5] Studies to find correlations between physicochemical properties and biological activities of sulfonamides indicated the dominating role played by their proton-releasing ability constant, more commonly known as pK_a of the sulfonamides.[6–13]

The complexing ability to form supramolecular complexes is dependent on this pK_a parameter. It was observed that the bacteriostatic activity of the sulfonamides was due to the presence of a larger proportion of sulfonamide in an active (ionized) form. Bell and Roblin,[6] in their extensive study on the relationship between pK_a of a series of sulfonamides and their in vitro antibacterial activity against *Escherichia coli*, found that the relationship between log 1/MIC and pK_a was parabolic in nature. Further, the higher points of the curve were found to lie between pK_a 6 and 7.4; the maximal activity was thus observed in sulfonamides whose pK_a approximated the physiological pH.[7,8]

In a recent review by Hansch et al.[14] as well as in our earlier studies,[15–19] we have shown that the metal complexes of biological agents (in the present case sulfonamides) generally exhibit enhancement of the physiological activity of the ligands and that the binding affinity vis-à-vis physiological activity of such complex agents depends upon their pK_a.

The argument made above indicates that the functional relationship between the acid dissociation constant

Keywords: QSAR; pK_a; Sulfonamide; Topological index; Molar volume; Molar refraction.
* Corresponding author. Tel.: +91 731 2531906; fax: +91 731 2763618; e-mail addresses: pvkhadikar@rediffmail.com; abhilashthakur@yahoo.com

0960-894X/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.bmcl.2005.05.136

(pK_a) and the biological activity of sulfonamides is well established and could not be questioned. The pK_a is related to solubility, distribution and partition coefficients, permeability across membranes, protein binding, tubular secretion, and re-absorption in the kidneys. Consequently, modeling, monitoring, and estimation of pK_a of sulfonamides play a dominant role in explaining all these factors.

Our earlier reports[20–26] have indicated that distance-based topological indices can be used very successfully for modeling, monitoring, and estimating various physicochemical parameters as well as physiological activities of the organic compounds acting as drugs. Generally, topological indices have proved more useful in these regards. Our recent report[26] has indicated that Balaban index (J)[27] is a very useful index for this purpose. Since pK_a is also an important physicochemical parameter, we thought it worthy to investigate the usefulness of distance-based topological indices in general and the Balaban index in particular for modeling the pK_a of sulfonamides. The relevant work done related to the objective of the present study is described in our earlier publications.[20–26]

It is interesting to record that although Balaban index (J) is a highly discriminating index, comparatively very little work has been done on the use of Balaban index (J) in developing quantitative structure–property–activity–toxicity relationships (QSPR/QSAR/QSTR). The primary reason for this is that theoretical chemists have been very slow to appreciate the overriding importance of the Balaban index (J) in modeling their physicochemical and biological processes. Nevertheless, earlier, we used this index successfully in developing some QSPR/QSAR models.[26] The main objective of the present investigation is to investigate the relative potential of Balaban index (J) compared to other Balaban type indices: J_z (Balaban type index from Z-weighted distance matrix); J_m (Balaban type index from mass-weighted distance matrix); J_v (Balaban type index from van der Waals weighted distance matrix), J_e (Balaban type index from electronegativity-weighted distance matrix), and J_p (Balaban type index from polarizability-weighted distance matrix) (Table 3). The results, as discussed below, established that the Balaban index (J) is the best index for this purpose. In fulfilling our objective, we have considered a set of 43 sulfonamides (Fig. 1, Table 1) and adopted their earlier[14] reported pK_a values. The present study will be useful to medicinal chemists interested in modeling physiological activities of meta- or para-substituted benzene sulfonamides.

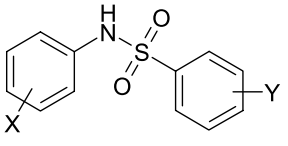

Figure 1. General structure of sulfonamides used in the present study.

Table 1. Substituents, observed pK_a, and indicator parameters[a] for sulfonamides used in the present study

Compound No.	X	Y	pK_a (Obs.)	I_X	I_Y	I_N
1	H	H	9.10	0	0	0
2	H	4-OMe	9.42	0	1	0
3	H	4-Me	9.35	0	1	0
4	H	4-F	8.90	0	0	0
5	H	4-Cl	8.47	0	0	0
6	H	4-Br	8.50	0	0	0
7	H	3-Br	8.25	0	0	0
8	H	3-NO_2	7.50	0	0	0
9	3-NO_2	H	7.93	0	0	0
10	3-NO_2	4-OMe	8.44	0	1	0
11	3-NO_2	4-Me	8.27	0	1	0
12	3-NO_2	4-F	7.85	0	0	0
13	3-NO_2	4-Cl	7.51	0	0	0
14	3-NO_2	4-Br	7.42	0	0	0
15	4-Cl	H	8.75	0	0	0
16	4-Cl	4-OMe	9.19	0	1	0
17	4-Cl	4-Me	9.02	0	1	0
18	4-Cl	4-F	8.61	0	0	0
19	4-Cl	4-Cl	8.30	0	0	0
20	4-Cl	4-Br	8.24	0	0	0
21	4-Cl	3-NO_2	7.19	0	0	0
22	4-F	H	8.85	0	0	0
23	4-F	4-OMe	9.32	0	1	0
24	4-F	4-Me	9.20	0	1	0
25	4-F	4-F	8.73	0	0	0
26	4-F	4-Cl	8.41	0	0	0
27	4-F	4-Br	8.38	0	0	0
28	4-F	3-NO_2	7.27	0	0	0
29	4-Me	4-OMe	9.80	1	1	0
30	4-Me	4-Me	9.65	1	1	0
31	4-Me	H	9.34	1	0	0
32	4-Me	4-F	9.23	1	0	0
33	4-Me	4-Cl	8.78	1	0	0
34	4-Me	3-NO_2	7.80	1	0	0
35	4-NH_2	H	10.29	0	0	1
36	4-NH_2	4-Me	10.53	0	1	1
37	4-NH_2	3-Me	10.44	0	1	1
38	4-NH_2	4-Cl	9.76	0	0	1
39	H	4-NO_2	6.66	0	0	0
40	3-NO_2	4-NO_2	5.51	0	0	0
41	4-Cl	4-NO_2	6.24	0	0	0
42	4-F	4-NO_2	6.45	0	0	0
43	4-Me	4-NO_2	6.78	1	0	0

[a] I_X = accounting for alkyl substitution at X, I_Y = accounting for alkyl substitution at Y, I_N = accounting for X = 4-NH_2 substitution.

2. Results and discussion

The set of 43 sulfonamides and their adopted pK_a values are presented in Table 1. Table 1 also includes the values of indicator parameters. The details regarding indicator parameters used are given in Sections 4.2–4.8. The calculated values of distance-based topological indices are recorded in Tables 2 and 3, whereas their interrelation and correlation with pK_a are presented in Tables 4 and 5. The details of the topological indices used are given in Sections 4. The correlation matrix (Table 5) indicated that in no case are mono-parametric regressions to be obtained for modeling pK_a of the benzene sulfonamides used. Consequently, we have to undergo multiple regression analysis. We have, therefore, adopted maximum R^2-method[28] for this purpose. The stepwise regression

Table 2. Topological indices and molecular parameters[a] related to the sulfonamides used in the present study

Compound No.	W	$^1\chi$	Sz	MV	η
1	447	7.68347	681	178.5	1.635
2	646	8.61532	976	202.5	1.610
3	538	8.07732	820	194.8	1.621
4	538	8.07732	820	182.7	1.618
5	538	8.07732	820	190.5	1.641
6	538	8.07732	820	194.7	1.657
7	528	8.07732	800	194.7	1.657
8	726	8.98800	1074	190.3	1.656
9	720	8.98800	1068	190.3	1.656
10	986	9.91985	1460	214.3	1.630
11	843	9.38185	1254	206.6	1.641
12	843	9.38185	1254	194.6	1.639
13	843	9.38185	1254	202.3	1.661
14	843	9.38185	1254	206.5	1.675
15	536	8.07732	818	190.5	1.641
16	758	9.00917	1148	214.5	1.617
17	638	8.47116	974	206.7	1.627
18	638	8.47116	974	194.7	1.625
19	638	8.47116	974	202.4	1.647
20	638	8.47116	974	206.6	1.661
21	847	9.38185	1258	202.3	1.661
22	536	8.07732	818	182.7	1.618
23	758	9.00917	1148	206.7	1.596
24	638	8.47116	974	199.0	1.606
25	638	8.47116	974	186.9	1.602
26	638	8.47116	974	194.7	1.625
27	638	8.47116	974	198.9	1.640
28	847	9.38185	1258	194.6	1.639
29	758	9.00917	1148	218.8	1.600
30	638	8.47116	974	211.1	1.609
31	536	8.07732	818	194.8	1.621
32	638	8.47116	974	199.0	1.606
33	638	8.47116	974	206.7	1.627
34	847	9.38185	1258	206.6	1.641
35	536	8.07732	818	180.8	1.669
36	638	8.47116	974	197.1	1.653
37	627	8.47116	952	197.1	1.653
38	638	8.47116	974	192.7	1.674
39	756	8.98800	1134	190.3	1.656
40	756	8.98800	1134	202.2	1.675
41	880	9.38185	1324	202.3	1.661
42	880	9.38185	1324	194.6	1.639
43	880	9.38185	1324	206.6	1.641

[a] W = Wiener index, $^1\chi$ = first-order connectivity index, Sz = Szeged index, MV = molar volume, η = index of refraction.

Table 3. Balaban index and various Balaban type indices[a] calculated for the sulfonamides used in the present study

Compound No.	J	J_z	J_m	J_v	J_e	J_p
1	1.84162	3.106	3.106	1.926	2.414	2.127
2	1.80081	3.000	3.000	1.482	2.400	1.978
3	1.82718	3.045	3.045	1.952	2.402	2.148
4	1.82718	3.080	3.080	1.892	2.423	2.043
5	1.82718	3.115	3.115	1.952	2.416	2.159
6	1.82718	3.136	3.136	1.964	2.412	2.171
7	1.86027	3.188	3.188	1.984	2.440	2.196
8	1.87293	3.092	3.092	1.883	2.487	1.988
9	1.89551	3.012	3.012	1.831	2.459	1.916
10	1.85781	2.961	2.961	1.787	2.442	1.852
11	1.88173	2.983	2.983	1.856	2.447	1.947
12	1.88173	3.008	3.008	1.815	2.462	1.882
13	1.88173	3.032	3.032	1.856	2.457	1.953
14	1.88173	3.047	3.047	1.864	2.454	1.961
15	1.83724	3.076	3.076	1.924	2.403	2.116
16	1.80106	2.990	2.990	1.847	2.392	1.983
17	1.82474	3.027	3.027	1.946	2.393	2.135
18	1.82474	3.058	3.058	1.892	2.412	2.042
19	1.82474	3.089	3.089	1.946	2.406	2.145
20	1.82474	3.107	3.107	1.957	2.402	2.156
21	1.86872	3.078	3.078	1.886	2.473	1.994
22	1.83724	3.042	3.042	1.866	2.410	2.007
23	1.80106	2.964	2.964	1.804	2.397	1.903
24	1.82474	2.997	2.997	1.893	2.399	2.035
25	1.82474	3.028	3.028	1.842	2.418	1.949
26	1.82474	3.058	3.058	1.893	2.412	2.044
27	1.82474	3.076	3.076	1.904	2.408	2.054
28	1.86872	3.054	3.054	1.845	2.479	1.921
29	1.80106	2.937	2.937	1.847	2.381	1.976
30	1.82474	2.967	2.967	1.946	2.381	2.127
31	1.83724	3.008	3.008	1.924	2.390	2.106
32	1.82474	2.998	2.998	1.892	2.399	2.034
33	1.82474	3.027	3.027	1.194	2.393	2.136
34	1.86872	3.028	3.028	1.886	2.463	1.988
35	1.83724	3.023	3.023	1.906	2.399	2.077
36	1.82474	2.980	2.980	1.930	2.389	2.100
37	1.85516	3.020	3.020	1.947	2.415	2.120
38	1.82474	3.041	3.041	1.930	2.401	2.109
39	1.79994	2.990	2.990	1.847	2.421	1.948
40	1.79994	2.962	2.962	1.797	2.456	1.845
41	1.80014	2.984	2.984	1.852	2.412	1.957
42	1.80014	2.960	2.960	1.812	2.417	1.886
43	1.80014	2.936	2.936	1.852	2.402	1.951

[a] J = Balaban index, J_z = Balaban type index from Z-weighted distance matrix, J_m = Balaban type index from mass-weighted distance matrix, J_v = Balaban type index from Van der Waals weighted distance matrix, J_e = Balaban type index from electronegativity-weighted distance matrix, and J_p = Balaban type index from polarizability-weighted distance matrix.

analysis indicated that even no bi-parametric regressions gave statistically significant models. However, suddenly out of the several tri-parametric models attempted, one containing J, $^1\chi$, and I_Y gave encouraging statistics. This model is found as below:

$$\begin{aligned} pK_a &= 8.3096(\pm 3.8828)J - 1.4330(\pm 0.1954)^1\chi \\ &\quad + 1.5314(\pm 0.22711)I_Y + 5.2562, \\ n &= 43,\ SE = 0.6571,\ R = 0.8331,\ R_A^2 = 0.6619, \\ F &= 21.554,\ Q = 1.26. \end{aligned} \quad (1)$$

Here and thereafter n is the sample size (number of compounds used), SE is the standard error of estimation, R_A^2 is the adjusted R^2, R is the multiple regression coefficient, F is the Fisher statistics, and Q[29,30] is the quality factor defined as the ratio of correlation coefficient and standard error of estimation, $Q = R/SE$.

In Eq. 1, the coefficients of J and I_Y (indicator parameter responsible for alkyl substitution at Y) terms are positive and that of $^1\chi$ is negative. Also, the coefficient of the J term is the largest. This means that compared to other two parameters J has dominating effect on the modeling of pK_a. The Balaban index J is a variant of connectivity index and represents extended connectivity. It is a good descriptor for the shape of molecules. The positive coefficient of J in Eq. 1 indicates that all these

Table 4. Correlation matrix

	pK_a	W	$^1\chi$	J	Sz	MV	η	I_X	I_Y	I_N
pK_a	1.000									
W	−0.5847	1.000								
$^1\chi$	−0.5738	0.9956	1.000							
J	−0.0765	0.1880	0.2447	1.000						
Sz	−0.5755	0.9991	0.9918	0.1553	1.000					
MV	−0.0646	0.5690	0.5603	−0.1404	0.5844	1.000				
η	−0.4305	0.1703	0.1796	0.3527	0.1545	−0.1653	1.000			
I_X	0.1232	0.0566	0.0392	−0.1468	0.0664	0.3874	−0.3497	1.000		
I_Y	0.5177	0.0823	0.0901	−0.1804	0.0927	0.5149	−0.4588	0.0065	1.000	
I_N	0.5157	−0.1976	−0.2012	0.0069	−0.1922	−0.2237	0.3622	−0.1412	0.1577	1.000

Table 5. Correlation matrix of Balaban and various Balaban type indices with pK_a

	pK_a	J	J_z	J_m	J_v	J_e	J_p
pK_a	1.00000						
J	−0.7650	1.00000					
J_z	0.02976	0.32008	1.00000				
J_m	0.02976	0.32008	1.00000	1.00000			
J_v	0.08401	0.13598	0.29452	0.29452	1.00000		
J_e	−0.60742	0.73228	0.22008	0.22008	0.02953	1.00000	
J_p	0.53777	−0.09347	0.61823	0.61823	0.26271	−0.48846	1.00000

points are favorable for the modeling of pK_a. The positive coefficient of I_Y also means that alkyl substitution at Y helps in the modeling of pK_a. We have used I_Y as an indicator parameter for presence (=1)/absence (=0) of substitution at Y. The $^1\chi$ index conveys more information about the number of atoms in the molecule, and thus its negative coefficient in Eq. 1 indicates unfavorable effect of the number of atoms on the pK_a of the sulfonamide moieties used.

Successive regressions resulted in several tetra-parametric regressions, of which one containing J, $^1\chi$, I_Y, and I_N gave better results than the tri-parametric regression discussed above. This model is found as:

$$\begin{aligned} pK_a &= 7.0112(\pm 3.3571)J - 1.2754(\pm 0.1731)^1\chi \\ &\quad + 1.3806(\pm 0.1992)I_Y + 1.1750(\pm 0.3064)I_N \\ &\quad + 6.1992, \\ n &= 43,\ SE = 0.5395,\ R = 0.8822, \\ R_A^2 &= 0.7549,\ F = 33.337,\ Q = 1.64. \end{aligned} \quad (2)$$

Like Eq. 1, here also, coefficients of J and I_Y terms are positive, while that of $^1\chi$ is negative. Therefore, the physical significances of these parameters in Eq. 2 are the same as discussed under Eq. 1. The added parameter I_N (indicator parameter responsible for *para*-amino substitution) has positive sign indicating that the presence of modeling group is favorable for the *para*-amino of pK_a. The comparison of the statistics related to Eqs. 1 and 2 indicates that with the addition of I_N to Eq. 1, the statistics is improved significantly; and this improvement is provided by I_N, i.e., due to amide substitution.

Further regression analysis has shown that when a parameter known as index of refraction η is added to the above Eq. 2, there is significant improvement in the statistics and the resulting regression expression containing J, $^1\chi$, I_Y, I_N, and η yielded the following expression:

$$\begin{aligned} pK_a &= 10.9771(\pm 2.3387)J - 0.9651(\pm 0.1253)^1\chi \\ &\quad + 0.6959(\pm 0.1678)I_Y + 2.1750(\pm 0.2535)I_N \\ &\quad - 26.8775(\pm 3.9361)\eta + 40.2672, \\ n &= 43,\ SE = 0.3775,\ R = 0.9496,\ R_A^2 = 0.8884, \\ F &= 67.802,\ Q = 2.52. \end{aligned} \quad (3)$$

The physical significance of J, $^1\chi$, I_Y, and I_N is the same as discussed under Eq. 2. In addition, the coefficient of η term in Eq. 3 is negative indicating its unfavorable contribution for the modeling of pK_a. When we were attempting regression replacing one of the indicator parameters with molecular volume (MV), we observed further improvement in the statistics according to the following expression:

$$\begin{aligned} pK_a &= 14.3950(\pm 1.9047)J - 1.3058(\pm 0.1170)^1\chi \\ &\quad + 0.0472(\pm 0.0066)\mathrm{MV} + 2.7078(\pm 0.1750)I_N \\ &\quad - 32.7892(\pm 2.5299)\eta + 37.5052, \\ n &= 43,\ SE = 0.2961,\ R = 0.9693,\ R_A^2 = 0.9313, \\ F &= 114.939,\ Q = 3.27. \end{aligned} \quad (4)$$

The comparison of Eqs. 3 and 4 indicates that like I_Y, the replaced MV term also has a positive coefficient. This means that the MV has favorable contribution for the modeling of pK_a of the benzene-sulfonamides used. It is worth to mention that no other higher parametric models yielded better results than Eq. 4. So we have to concentrate on Eq. 4 and discuss its significance in greater detail, especially in view of Hansch's result

that pharmacological efficiency is the result of electronic, steric, and hydrophobic interactions between the drug and the receptor site. We have to provide such an explanation on the basis of parameters involved in regression equations 2–4 and use the same for the following Eqs. 5–9. Fortunately, we can do so using η and MV parameters. The refractive index η is frequently employed for characterizing organic compounds. It bears a significant relationship to other properties: optical, electrical, and magnetic properties, particularly the polarizability. In the form of molar refraction (MR), it contains electric contribution. As MR, it accounts for polarizability rather than steric effect. As MR (η) along with MV accounts for dispersion forces aiding the binding of a ligand to the receptor site. In such cases, MR should have positive coefficient. It (η as MR) measures ligand ability to distort the conformation of the receptor in such a way as to preclude union with the proper substrate. Since the conformational change is detrimental, a negative coefficient should result for the MR term in this case. In our case, this is consistent with the negative coefficient of η in Eqs. 3 and 4 as well as in the following Eqs. 5–9.

To compare the relative potential of J, J_z, J_m, J_v, J_e, and J_p indices, we have used this model (Eq. 4) where J is replaced by other Balaban type indices in succession so that exact comparisons could be made. The resulting models are found as below:

$$\begin{aligned} pK_a &= 5.9591(\pm 1.8458)J_z - 0.6349(\pm 0.2032)^1\chi \\ &\quad + 0.0372(\pm 0.0090)MV + 3.0883(\pm 0.2904)I_N \\ &\quad - 36.5252(\pm 4.3535)\eta + 48.0876, \\ n &= 43,\ SE = 0.4172,\ R = 0.9380,\ R_A^2 = 0.8637, \\ F &= 54.225,\ Q = 2.25. \end{aligned} \quad (5)$$

$$\begin{aligned} pK_a &= 5.9591(\pm 1.8458)J_m - 0.6349(\pm 0.2032)^1\chi \\ &\quad + 0.0372(\pm 0.0090)MV + 3.0883(\pm 0.2904)I_N \\ &\quad - 36.5252(\pm 4.3535)\eta + 48.0876, \\ n &= 43,\ SE = 0.4172,\ R = 0.9380,\ R_A^2 = 0.8637, \\ F &= 54.225,\ Q = 2.25. \end{aligned} \quad (6)$$

$$\begin{aligned} pK_a &= 0.3768(\pm 0.5861)J_v - 1.0265(\pm 0.1809)^1\chi \\ &\quad + 0.0341(\pm 0.0101)MV + 2.5832(\pm 0.2766)I_N \\ &\quad - 28.4378(\pm 3.9560)\eta + 56.2534, \\ n &= 43,\ SE = 0.4697,\ R = 0.9208,\ R_A^2 = 0.8272, \\ F &= 41.223,\ Q = 1.960. \end{aligned} \quad (7)$$

$$\begin{aligned} pK_a &= 11.4858(\pm 4.2230)J_e - 1.4668(\pm 0.2254)^1\chi \\ &\quad + 0.0521(\pm 0.0115)MV + 2.9460(\pm 0.2859)I_N \\ &\quad - 5.0512(\pm 4.1097)\eta, \\ n &= 43,\ SE = 0.4312,\ R = 0.9337,\ R_A^2 = 0.8544, \\ F &= 50.293,\ Q = 2.165. \end{aligned} \quad (8)$$

$$\begin{aligned} pK_a &= 3.0863(\pm 1.6951)J_p - 0.4232(\pm 0.3830)^1\chi \\ &\quad + 0.0183(\pm 0.0129)MV + 2.5679(\pm 0.2666)I_N \\ &\quad - 32.1517(\pm 4.3976)\eta + 54.6784, \\ n &= 43,\ SE = 0.4527,\ R = 0.9267,\ R_A^2 = 0.8396, \\ F &= 44.979,\ Q = 2.05. \end{aligned} \quad (9)$$

The results (Eqs. 5–9), therefore, show that Balaban index (J) is superior to other Balaban type indices viz J_z, J_m, J_v, J_e, and J_p. In view of this, we have concentrated on the results given by Eq. 4. The results (Eqs. 5–9) show that among Balaban and Balaban type indices, Balaban index (J) gives the best results. In case of other Balaban type indices, we observed that similar results are obtained when J_z, J_m, and J_e are used as correlating param-

Table 6. Experimental and calculated pK_a values using the models expressed by Eq. 4

Compound No.	pK_a (Obs.)	pK_a[a]	Residual
1	9.10	8.79	0.31
2	9.42	8.94	0.48
3	9.35	9.30	0.05
4	8.90	8.83	0.07
5	8.47	8.44	0.03
6	8.50	8.11	0.38
7	8.25	8.59	−0.34
8	7.50	7.41	0.09
9	7.93	7.73	0.20
10	8.44	7.96	0.48
11	8.27	8.28	−0.01
12	7.85	7.78	0.07
13	7.51	7.42	0.09
14	7.42	7.16	0.26
15	8.75	8.58	0.16
16	9.19	8.77	0.42
17	9.02	9.11	−0.09
18	8.61	8.61	0.00
19	8.30	8.26	0.04
20	8.24	7.99	0.24
21	7.19	7.24	−0.05
22	8.85	8.97	−0.12
23	9.32	9.09	0.23
24	9.20	9.44	−0.24
25	8.73	9.00	−0.27
26	8.41	8.61	−0.20
27	8.38	8.32	0.06
28	7.27	7.59	−0.32
29	9.80	9.53	0.27
30	9.65	9.91	−0.26
31	9.34	9.44	−0.10
32	9.23	9.44	−0.21
33	8.78	9.11	−0.33
34	7.80	8.09	−0.29
35	10.29	9.92	0.37
36	10.53	10.52	0.01
37	10.44	10.95	−0.51
38	9.76	9.62	0.14
39	6.66	6.36	0.30
40	5.51	6.30	−0.79
41	6.24	6.25	−0.01
42	6.45	6.61	−0.16
43	6.78	7.11	−0.33

[a] Calculated from Eq. 4.

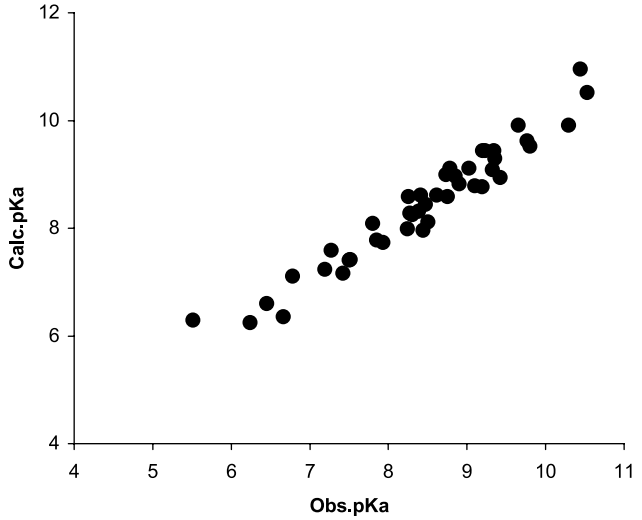

Figure 2. Showing relationship between observed pK_a and calculated pK_a values from Eq. 4.

eters. The remaining two Balaban type indices, that is, J_v and J_p also yield similar results and the former set of Balaban type indices is found better than the latter.

At this stage, it is worth commenting on adjustable-R^2, that is, R^2_A values. We observed that as we pass from the tri-parametric to tetra-parametric regression analysis there is consistent increase in R^2_A, increasing from 0.6619 to 0.9313 as we pass from Eq. 1 to Eq. 9. Such an increase in R^2_A values indicates that the added variable has a fair share in proposing the respective model. The value of R^2_A will decrease if the added variable does not reduce the unexplained variation in the dependent variables enough to off set the loss of degree of freedom.[31–34,23]

To confirm our results, we have estimated pK_a values of the sulfonamides under the present study using Eq. 4 and compared them with their observed pK_a. Such a comparison is shown in Table 6, which shows that estimated pK_a values using Eq. 4 are closest to their observed values. Further, support in our favor is obtained by calculating predictive correlation coefficients R^2_{pred} as obtained by correlating observed and calculated pK_as (Fig. 2). The $R^2_{pred} = 0.9395$ indicates that the model based on Eq. 4 is statistically excellent and also has excellent predictive power. The predictive power of the model can also be judged by estimating the quality factor Q (=R/SE); the highest value of Q obtained for Eq. 4 further supports its predictive power. It is worthy to mention that Q value goes on increasing from Eqs. 1–4 and acquires the highest value for Eq. 4.

3. Conclusions

The results and discussion above indicate that none of the molecular descriptors used singly (independently) is capable of yielding any significant mono-parametric regression model for modeling pK_a of the sulfonamides used. Also, the Balaban index (J) in combination with other indicator parameters gave a model with excellent statistics. Furthermore, replacement of one of the indicator parameters I_Y by MV was useful for modeling, monitoring, and estimating pK_a. Finally, Balaban index J is better index than the other Balaban type indices: J_z, J_m, J_v, J_e, and J_p. We note that the R^2 values are 0.97 for Hansch's correlation (Table 61 of Ref 14 in terms of two Hammett σ parameters, omiting two outliers) and 0.94 for our Eq. 4 with five parameters but without any outliers.

4. Experimental

4.1. Protonation constant (pK_a)

The proton–ligand formation constant expressed as pK_a was taken from the literature.[9,14]

4.2. Topological indices

All the topological indices used are calculated from the hydrogen suppressed molecular graphs. Though their calculations are extensively discussed in the literature,[35–39] we give the expressions used for their calculations below.

4.3. Wiener index (W)

Wiener index $W = W(G)$ of G is defined[40] as the half-sum of the elements of the distance matrix:

$$W = W(G) = \tfrac{1}{2}\sum_{i=1}\sum_{j=1}(D)_{ij}, \tag{10}$$

where $(D)_{ij}$ is the ijth element of the distance matrix, which denotes the shortest graph—theoretical distance between sites i and j of G.

4.4. The connectivity index ($^1\chi$)

The connectivity index $^1\chi = {}^1\chi(G)$ of G is defined[41] by Randic as:

$$^1\chi = {}^1\chi(G) = \sum_{\text{Bonds}}[V(i)\cdot V(j)]^{-0.5}, \tag{11}$$

where V_i is the vertex degree of vertex i.

4.5. Balaban index (J)

The Balaban index $J = J(G)$ of G is defined[27] as:

$$J = M/(\mu+1)\sum_{\text{Bonds}}(d_i\cdot d_j)^{-0.5}, \tag{12}$$

where M is the number of bonds in G, μ is the cyclomatic number of G, and d_i ($i = 1, 2, 3, \ldots N$; N is the number of vertices in G) is the distance sum. The cyclomatic number $\mu = \mu(G)$ of a cyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related acyclic graph. In the case of monocyclic graphs $\mu = 1$, it is generally calculated by means of the following expression:

$$\mu = M - N + 1. \tag{13}$$

4.6. Szeged index (Sz)

The Szeged index, $Sz = Sz(G)$, is calculated[42–44] according to the following expression:

$$Sz = Sz(G) = \sum\nolimits_{Edges} n_u \cdot n_v, \quad (14)$$

where n_u is the number of vertices lying closer to one end of the edge $e = uv$; the meaning of n_v is analogous. Edges equidistant from both the ends of an edge $e = uv$ are not taken into account.

4.7. Balaban type indices

The Balaban type indices: J_z, J_m, J_v, J_c, and J_p were calculated using Dragon Software.[46,47]

4.8. Molecular descriptors

In the present study, molar volume (MV) and index of refraction (η) (Table 2) were calculated from computer software ACD labs.[45]

4.9. Indicator parameters

In the present study, three different indicator parameters are used for understanding the significance of substituents in their activity/property.

Indicator parameter I_X accounting for alkyl substitution at X, I_Y accounting for alkyl substitution at Y position, and I_N accounting for *para*-amino substitution.

4.10. Regression analysis

All the regressions were carried out using maximum R^2 method.[28]

4.11. Software

The calculation of topological indices and regression analysis were performed using software developed by Prof. Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary. In addition, we have used ACD labs, Hyperchem, and Dragon programs for making other calculations.

Acknowledgments

The authors thank Prof. Istvan Lukovits for providing software for making calculations of some of the topological indices as well as for performing regression analyses.

References and notes

1. Balaban, A. T.; Basak, S. C.; Beteringhe, A.; Mills, D.; Supuran, C. T. *Mol. Divers.* **2004**, *8*, 401.
2. Supuran, C. T.; Scozzafava, A.; Conway, J. *Carbonic Anhydrase Inhibitors & Activators*; CRC Press: New York, USA, 2004.
3. Supuran, C. T.; Balaban, A. T. *Rev. Roum. Chem.* **1994**, *39*, 107.
4. Thakur, M.; Thakur, A.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 203.
5. Supuran, C. T.; Casini, A.; Mastrolorenzo, A.; Scozzafava, A. *Mini-Rev. Med. Chem.* **2004**, *4*, 625.
6. Bell, P. M.; Roblin, R. O. *J. Am. Chem. Soc.* **1942**, *64*, 2905.
7. Silipo, C.; Vittoria, A. *Farmaco [Sci.]* **1979**, *34*, 858.
8. Miller, G. H.; Doukas, P. M.; Seydel, J. K. *J. Med. Chem.* **1972**, *15*, 700.
9. Seydel, J. K. *J. Med. Chem.* **1971**, *14*, 724.
10. Walter, W.; Becker, R. F. *Liebigs Ann. Chem.* **1969**, *727*, 71.
11. Kakeya, N.; Aoki, M.; Kamada, A.; Yata, N. *Chem. Pharm. Bull.* **1969**, *17*, 1010, 2000.
12. Dauphin, G.; Kergowrd, A. *Bull. Soc. Chim. Fr.* **1961**, 468.
13. Yoshika, M.; Hamamoto, K.; Kubota, T. *Bull. Chem. Soc. Jpn.* **1962**, *35*, 1723.
14. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619.
15. Karmarkar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K. *Indian J. Chem. A* **2001**, *40*, 12.
16. Karmarkar, S.; Khadikar, P. V.; Agrawal, V. K.; Mathur, K. C.; Mandloi, M.; Joshi, S. *Proc. Indian Acad. Sci. (Chem. Sci.)* **2000**, *112*, 43.
17. Khadikar, P. V.; Karmarkar, S.; Joshi, S.; Gutman, I. *J. Serb. Chem. Soc.* **1996**, *61*, 89.
18. Khadikar, P. V.; Pol, B.; Ali, S. M. *Rev. Microbiol.* **1986**, *12*, 291.
19. Khadikar, P. V.; Pol, B.; Ali, S. M. *Bacteriology* **1987**, *22*, 81.
20. Khadikar, P. V.; Phadnis, A.; Shrivastava, A. *Bioorg. Med. Chem.* **2002**, *10*, 1181.
21. Agrawal, V. K.; Shrivastava, S.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2003**, *11*, 3553.
22. Khadikar, P. V.; Joshi, S.; Bajaj, A. V.; Mandloi, D. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1187.
23. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2004**, *12*, 2477.
24. Agrawal, V. K.; Bano, S.; Supuran, C. T.; Khadikar, P. V. *Eur. J. Med. Chem.* **2004**, *39*, 593.
25. Jaiswal, M.; Khadikar, P. V.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3283.
26. Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem.* **2004**, *12*, 789.
27. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
28. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley VCH: New York, 2000.
29. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
30. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
31. Agrawal, V. K.; Sohgaura, R.; Khadikar, P. V.; Phadnis, A. *Bioorg. Med. Chem.* **2001**, *9*, 3295.
32. Khadikar, P. V.; Mathuhr, K. C.; Singh, S.; Phadnis, A.; Shrivastava, A.; Mandloi, M. *Bioorg. Med. Chem.* **2002**, *10*, 1761.
33. Khadikar, P. V.; Karmarker, S.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem.* **2002**, *10*, 3163.
34. Khadikar, P. V.; Singh, S.; Mandloi, D.; Joshi, S.; Bajaj, A. V. *Bioorg. Med. Chem.* **2003**, *11*, 5045.
35. Supuran, C. T.; Scozzafava, A. *J. Enzym. Inhib.* **2000**, *15*, 597.
36. Casini, A.; Antel, J.; Abbate, F.; Scozzafava, A.; David, S.; Waldeck, H.; Schafer, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 841.
37. Clare, B. W.; Supuran, C. T. *J. Pharm. Sci.* **1994**, *83*, 768.
38. Scozzafava, A.; Supuran, C. T. *SAR. QSAR Environ. Res.* **2001**, *12*, 17.

39. Supuran, C. T.; Scozzafava, A. *Eur. J. Med. Chem.* **2000**, *35*, 867.
40. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
41. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
42. Gutman, I. *Graph Theory Notes, New York* **1994**, *27*, 9.
43. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
44. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *MATCH Commun. Math. Comput. Chem.* **2001**, *43*, 7.
45. ACD-Lab software for calculating the referred physicochemical parameters; Chemsketh 3.O. <www.acdlabs.com>.
46. Hyperchem-7 software for calculating the molecular modeling parameters. <www.hyper.com>.
47. Dragon software for calculation of Balaban type indices. <www.disat.unimib.it>.