QSAR study on inhibition of *E. Coli* by sulfonamides

Abhilash Thakur,[a*] Mamta Thakur,[b] and Padmakar V. Khadikar[c]

[a] *Department of Chemistry, Rishiraj Institute of Technology, Indore India*
[b] *Department of Chemistry, Kailash Chand Bansal Academy of Technology, Indore India*
[c] *Research Division, Laxmi Fumigation and Pest control, Pvt. Ltd, 3, Khatipura, Indore 452007, India*
E-mail: abhilashthakur@yahoo.com, pvkhadikar@rediffmail.com

Abstract

The paper describes a QSAR study on inhibition of *E. Coli* by sulfonamides using distance-based topological indices. The sulfonamides discussed consists of 39 derivatives with substitution at 2-, 3- and 4- positions as well as having some di-substitution. Application of multiple linear regression analysis indicated that combination of distance-based topological indices with *ad hoc* molecular descriptors and indicator parameters yielded statistically significant model for modeling inhibitory activity (log1/C) of *E. Coli*. Predictive potential of the model is obtained by cross-validation parameters as well as by using a variety of other statistics. The results have indicated that there is no involvement of a positive hydrophobic term (logP) in the inhibition process, suggesting that the binding of the sulfonamides to the active center does not depend on hydrophobic interactions. Final selection of potential sulfonamide is made by molecular modeling.

Keywords: QSAR, sulfonamides, *E. Coli* inhibition, molecular descriptors

Introduction

Bacteria have been one of the most studied test systems. The current data-base of Hansch[1] indicated that out of 1775 QSAR's on a single organism, 649 are for bacteria and that bacteria vary enormously in their response on treatment with various organic compounds acting as a drugs. It is worthy to mention the antibacterial activity of sulfa- drugs having similar structure is due to their activity as 4-aminobenzoic acid antagonist. These drugs inhibited the incorporation of [1,2] *p*-aminobenzoic acid into folic acid by folate synthetase. The folate synthetase incorporates the drug sulfamethoxazole in pterin, thus completely blocking folate synthesis[2,3]. A plethora of literature exists on the action of sulfa-drugs on all kind of bacteria. Needless to state that *p*-

aminobenzoic acid is a component of folic acid and is needed by bacteria for their survival and multiplication.

Carbonic anhydrase(CA's), class of sulfa-drugs are inhibited by sulfonamides. The ring substituted benzenesulfonamides containing $-SO_2NH_2$ groups have similar activities. In case of sulfonamides the inhibition of CA is caused mainly by the binding (coordination) of the SO_2NH^- anion to the Zn^{2+} of the enzyme, miming the bicarbonate anion in the transition state[4]. Furthermore, we have shown that inhibition of carbonic anhydrase by sulfonamides and Schiff base derived from them can be modeled excellently using distance-based topological indices. Earlier, in many cases hydrophobic parameters were used for this purpose.

Recently[1] Hansch stated that "*over the year we have been impressed by the great importance of hydrophobic effects in chemical-biological interactions as brought out by quantitative structure-activity relationships (QSAR's),…….it seems timely to examine those instances where hydrophobic terms are not significant*".

Prompted by the above, and in continuation of our earlier work[4-15], we have undertaken the present study in that we have made QSAR analysis on the inhibition of *E.Coli* by sulfonamides using a set of molecular descriptors consisting of some distance-based topological indices together with some adhoc molecular descriptors. Another objective of our study was to investigate the need (if any) of hydrophobic parameter (logP) in modeling inhibition of *E.Coli* by sulfonamides. We have used earlier data[1] for this purpose. (Figure 1, Table 1).

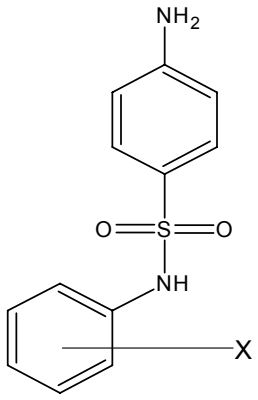

Figure 1. General structure of sulfonamides used in the present study.

Table 1. Structural details and assumed indicator parameter for the sulfonamides used in present study

Comp. No	Substituents	log1/C	I_X	I_{NO2}	I_{DS}	I_O	I_M	I_P
1	4-NMe_2	4.35	0	0	0	0	0	1
2	2-OMe	4.45	0	0	0	1	0	0
3	2-OC_2H_5	4.35	0	0	0	1	0	0
4	4-OMe	4.47	0	0	0	0	0	1
5	4-OC_2H_5	4.49	0	0	0	0	0	1
6	4-Me	4.66	0	0	0	0	0	1
7	2-Me	4.46	0	0	0	1	0	0
8	3-Me	4.60	0	0	0	0	1	0
9	H	4.80	0	0	0	0	0	0
10	3-OC_2H_5	4.80	0	0	0	0	1	0
11	3-OMe	4.80	0	0	0	0	1	0
12	4-Cl	4.89	1	0	0	0	0	1
13	4-Br	4.89	1	0	0	0	0	1
14	2-Br	4.99	1	0	0	1	0	0
15	4-I	4.95	1	0	0	0	0	1
16	3-I	5.04	1	0	0	0	1	0
17	3-Cl	5.10	1	0	0	0	1	0
18	3-Br	4.95	1	0	0	0	1	0
19	3-NO_2	5.60	0	1	0	0	1	0
20	3-CF_3	5.25	0	0	0	0	1	0
21	4-CF_3	5.40	0	0	0	0	0	1
22	4-CN	6.00	0	0	0	0	0	1
23	4-$COCH_3$	5.70	0	0	0	0	0	1
24	4-NO_2	6.00	0	1	0	0	0	1
25	4-SO_2CH_3	5.85	0	0	0	0	0	1
26	2,3-di-Me	4.32	0	0	1	1	1	0
27	2-Me,5-Cl	4.80	1	0	1	1	1	0
28	2-Me,6-Cl	4.80	1	0	1	1	0	0
29	3,4-di-Cl	5.40	1	0	1	0	1	1
30	3,5-di-Cl	5.55	1	0	1	0	1	0
31	2-Cl,4-OMe	4.77	1	0	1	1	0	1
32	2-OMe,4-Cl	5.10	1	0	1	1	0	1
33	2-Cl,4-NO_2	5.10	1	1	1	1	0	1
34	2-Me,4-NO_2	5.55	0	1	1	1	0	1
35	2-Me,4-NO_2	5.41	1	1	1	1	0	1
36	2-Br,4-NO_2	5.64	1	1	1	1	0	1
37	4-NO_2,2-CF_3	5.32	1	0	0	1	0	0
38	2-Cl	5.55	1	0	0	1	0	0
39	2-I	5.68	1	0	0	1	0	0

I_X = 1 if halogen substitution, I_{NO2} = 1 if NO_2 substitution, I_{DS} = 1 if di-substitution, I_O = 1 if substitution at ortho position, I_M = 1 if Substitution at meta position, I_P = 1 if substitution at para position.

The topological indices used for the QSAR analysis were Wiener,[16] Szeged,[17-19] first order molecular connectivity[20], and Balaban indices[21]. The adhoc molecular descriptors used were molar refractivity (MR), molar volume(MV), parachor(Pc), refractive index(η), surface tension(ST), density(d) and polarizability(α) in addition to logP and HE (hydration energy). Preliminary statistical analysis indicated the need of some indicator parameters for obtaining statistically significant models. We have, therefore, used six indicator parameters I_{P1} to I_{P6} accounting for the presence/absence of substitution at X: nitro-substitution, disubstitution, substitution at ortho-position, substitution at meta- position and substitution at para-position respectively. These indicator parameters are dummy parameters and assume only two values: one (for presence) and zero (for absence). For obtaining appropriate QSAR model we have used maximum R^2 method and followed stepwise regression analysis.[22-24] The predictive ability of the model is discussed on the basis of predictive correlation coefficient. We have separated a set of potential inhibitors of *E.Coli* and finally we have aimed at the most appropriate model using molecular modeling.

Results and Discussion

The set of 39 sulfonamides and their adopted inhibition (MIC) values of *E-Coli* expressed as log1/C are presented in Table 1 which shows that log1/C are highly influenced by the substitution on the aromatic nucleus. The log1/C value is lowest for 2,3-di-Me substitution. The data presented in Table 1 also shows the occurrence of degeneracy in the activity (log1/C) values. Consequently, we have the following sequence of activity values:

22 = 24>25>23>39>36>19>30 = 34 = 38>35>21 = 29>37>20>17 = 32 = 33>16>14>15> = 18>12 = 13>9 = 10 = 11 = 27 = 28>31>6>8>5>4>7>2>1 = 3> 26 (1)

This eq(1) shows that 3-OC_2H_5, 3-OMe, 2-Me, 5-Cl, and 2-Me, 6-Cl substitution does not change the activity (log1/C)of the parent (unsubstituted) sulfonamide. Thus, the effect due to electronic nature of these substitutents mimic that of hydrogen substitution. Similarly, the substituents 3,5- di- Cl, 2-Me, 4-NO_2, 2-Cl, on one hand and the substitutents 3-Cl, 2-OMe, 4-Cl, 2-Cl, 4-NO_2 on the other hand have an identical effect on the activity. Likewise, the three pairs (i) 4-CN; 4-NO_2 (ii) 4-CF_3, 3,5 di-Cl and (iii) 4-NH_2, 2-OC_2H_5 have independently an analogous effect on the activity.

Consequent to the occurrence of degeneracy in the activity it became essential to examine the degeneracy in the molecular descriptors also. In Table 2 the calculated values of topological indices: W, $^1\chi$, J and Sz are recorded. All these indices are important distance-based topological indices out of which W and Sz are first generation while $^1\chi$, and J are the second generation topological indices.[25-27] However, inspite of their degeneracy they can be successfully used in drug modeling[28,29]. Comparison of the observed activity and the corresponding topology of the sulfonamides used shows that the topology of the sulfonamides alone is not responsible for the

variation of the activity. The same is found to be the case with hydrophobic (logP) and other parameters (Table 3) used for modeling the activity.

The intercorrelatedness among molecular descriptors including topological indices with the activity shows (Table 4) that except J all other topological indices are highly mutually correlated, while this is not the case with the other physicochemical parameters used. Furthermore, data presented in Table 4 show that none of the molecular descriptors, including topological indices correlate well with the activity (log1/C). From this we conclude that no single variable model is capable of modeling the activity and that the refereed descriptors can be combined to obtain a statistically significant multiparametric model for modeling the activity. Also, that models containing two or more topological (except J) indices may suffer from defect due to correlation[22-24]. However, such cases are nicely dealt with Randic[30] and we will use his recommendations to analyze such cases.

Initial regression analysis indicated that out of the 12 molecular descriptors used Sz in combination of physicochemical descriptor plays a dominating role in modeling the activity. However, statistically significant models are obtained when three descriptors are used and that the quality of the model goes on improving with higher parameteric modeling (Table 5). The tri-parameteric model containing three descriptors (Sz, MV, d) is found as below:

$$\log 1/C = 7.5491 + 0.0016(\pm 3.4783 \times 10^{-4})Sz - 0.0281(\pm 0.0076)MV + 0.9825(\pm 0.3415)d \quad (2)$$

$n = 39$, $Se = 0.3550$, $R = 0.7074$, $R^2_A = 0.4576$, $F = 11.686$

Table 2. Calculated values of distance-based topological indices used in present study

Comp.No.	W	$^{1}\chi$	J	Sz
1	880	9.381	1.800	1324
2	714	9.026	1.912	1060
3	831	9.526	1.912	1209
4	758	9.009	1.801	1148
5	897	9.509	1.771	1341
6	638	8.471	1.825	974
7	616	8.488	1.890	930
8	627	8.471	1.855	952
9	536	8.077	1.837	818
10	864	9.509	1.836	1275
11	736	9.009	1.853	1140
12	638	8.471	1.825	974
13	638	8.471	1.825	974
14	616	8.488	1.890	930
15	638	8.471	1.825	974
16	627	8.471	1.855	952
17	627	8.471	1.855	952
18	627	8.471	1.855	952
19	847	9.382	1.869	1258
20	960	9.682	1.900	1414
21	1004	9.682	1.816	1502
22	758	9.009	1.801	1148
23	880	9.382	1.800	1324
24	880	9.382	1.800	1324
25	1004	9.682	1.816	1502
26	710	8.899	1.921	1068
27	712	8.882	1.916	1072
28	700	8.899	1.950	1048
29	732	8.882	1.861	1112
30	722	8.865	1.887	1092
31	847	9.420	1.875	1273
32	825	9.420	1.924	1229
33	974	9.793	1.880	1456
34	974	9.793	1.880	1456
35	974	9.793	1.880	1456
36	974	9.793	1.880	1456
37	1325	11.004	2.052	1921
38	616	8.488	1.900	930
39	616	8.488	1.900	930

W = Wienerindex, $^{1}\chi$ = Connectivity index, J= Balabanindex, Sz = Szeged index

Table 3. Hydrophobic (logP) and other physicochemical parameters used in the present study

Comp. No.	MR ^{cm3}	MV ^{cm3}	Pc ^{cm3}	η	ST $^{d/ cm}$	D $^{g/ cm3}$	Pol ^{cm3}	logP
1	80.68	218.8	617.9	1.658	63.6	1.331	31.98	1.538
2	73.91	204.8	571.9	1.641	60.7	1.358	29.30	1.172
3	78.54	221.3	612.0	1.627	58.4	1.320	31.13	1.657
4	73.91	204.8	571.9	1.641	60.7	1.358	29.30	1.178
5	78.54	221.3	612.0	1.627	58.4	1.320	31.13	1.663
6	72.17	197.1	551.5	1.653	61.3	1.330	28.61	1.789
7	72.17	197.1	551.5	1.653	61.3	1.330	28.61	1.789
8	72.17	197.1	551.5	1.653	61.3	1.330	28.61	1.789
9	67.54	180.8	513.3	1.669	64.9	1.373	26.77	1.370
10	78.54	221.3	612.0	1.627	58.4	1.32	31.13	1.663
11	73.91	204.8	571.9	1.641	60.7	1.358	29.30	1.178
12	72.37	192.7	550.4	1.674	66.4	1.466	28.69	2.017
13	75.27	197.0	564.3	1.689	67.3	1.660	29.83	2.279
14	75.27	197.0	564.3	1.689	67.3	1.660	29.83	2.662
15	80.47	202.9	587.2	1.723	70.1	1.844	31.90	2.664
16	80.47	202.9	587.2	1.723	70.1	1.844	31.90	2.664
17	72.37	192.7	550.4	1.674	66.4	1.466	28.69	2.017
18	75.27	197.0	564.3	1.689	67.3	1.660	29.83	2.279
19	73.58	192.6	570.3	1.689	76.8	1.522	29.17	1.092
20	72.53	214.3	575.3	1.592	51.8	1.475	28.75	2.378
21	72.53	214.3	575.3	1.592	51.8	1.475	28.75	2.378
22	72.11	190.7	560.9	1.680	74.7	1.430	28.58	0.728
23	76.91	212.3	596.8	1.644	62.3	1.367	30.49	0.912
24	73.58	192.6	570.3	1.689	76.8	1.522	29.17	1.096
25	80.99	222.6	631.2	1.647	64.6	1.466	32.10	0.029
26	76.79	213.3	589.8	1.639	58.3	1.295	30.44	2.208
27	76.99	209.0	588.7	1.658	62.8	1.419	30.52	2.436
28	76.99	209.0	588.7	1.658	62.8	1.419	30.52	2.842
29	77.20	204.7	587.5	1.678	67.8	1.549	30.60	2.548
30	77.20	204.7	587.5	1.678	67.8	1.549	30.60	2.664
31	78.74	216.7	609.0	1.646	62.3	1.442	31.21	2.231
32	78.74	216.7	609.0	1.646	62.3	1.442	31.21	1.819
33	78.40	204.6	607.5	1.692	77.6	1.601	31.08	2.149
34	78.20	208.9	608.6	1.671	71.9	1.470	31.00	1.515
35	86.50	214.7	644.3	1.738	81.0	1.952	34.29	2.830
36	81.30	208.8	621.4	1.706	78.3	1.782	32.23	2.388
37	78.57	226.1	632.3	1.611	61.0	1.597	31.14	2.104
38	72.37	192.7	550.4	1.674	66.4	1.466	28.69	2.423
39	80.47	202.9	587.2	1.723	70.1	1.844	31.90	3.104

MR = Molar Refractivity, MV = Molar Volume, Pc = Parachor, η = Index of Refraction, ST = Surface Tension, D = Density, Pol = Polarizability, logP = Octanol/Water partition coefficient.

Table 4. Correlation matrix

	log1/C	W	$^1\chi$	J	Sz	MR	MV	Pc	η
log1/C	1.0000								
W	.32646	1.0000							
$^1\chi$	.26855	.98683	1.0000						
J	-.06417	.23296	.30180	1.0000					
Sz	.34125	.99894	.98149	.20213	1.0000				
MR	.11600	.43270	.45818	.16635	.43726	1.0000			
MV	-.13459	.70236	.73619	.25993	.69373	.66687	1.0000		
Pc	.13511	.74297	.77441	.23999	.74315	.89905	.86563	1.0000	
η	.32063	-.36834	-.38288	-.12708	-.35253	.33595	-.47723	-.03416	1.000
ST	.53810	-.00149	-.01188	-.08591	.01580	.31581	-.39276	.11908	.8694
D	.44937	.04494	.00590	.12674	.05307	.52151	-.09413	.24501	.7516
Pol	.11613	.43253	.45804	.16625	.43711	1.0000	.66649	.89890	.3364
I_X	.14093	-.31676	-.30864	.20405	-.30913	.33288	-.17694	.05076	.6180
I_{NO2}	.43696	.60734	.60338	.24707	.61094	.29776	.05143	.39225	.2925
I_{DS}	.10381	.36906	.41359	.57485	.37239	.46304	.35814	.52011	.0846
I_O	-.09542	.15964	.24729	.71949	.14239	.33860	.25482	.34538	.0707
I_M	-.07714	-.20892	-.19358	.09581	-.21576	-.12368	-.07540	-.15665	-.0477
I_P	.26829	.50108	.44094	-.34624	.52918	.30763	.25931	.40367	.0335
logP	-.09096	-.20883	-.20018	.40665	-.21368	.27406	.00083	.04500	.3333
HE	-.51391	-.61900	-.61353	-.09123	-.62532	-.26380	-.02685	-.37890	-.2836

	ST	D	Pol	I_X	I_{NO2}	I_{DS}	I_O	I_M	I_P	logP	HE
ST	1.000										
D	.6398	1.000									
Pol	.3163	.5215	1.000								
I_X	.4209	.6387	.3328	1.000							
I_{NO2}	.6415	.3790	.2983	-.0309	1.000						
I_{DS}	.2306	.1782	.4631	.3857	.4120	1.000					
I_O	.1138	.1331	.3388	.2734	.2890	.5734	1.000				
I_M	-.1336	-.0602	-.1236	.0514	-.1674		.0370	-.3301	1.000		
I_P	.2314	.1358	.3075	-.0237	.3221	.2052	-.1256	-.5728	1.000		
logP	.0749	.5614	.2741	.7148	-.0319	.3813	.3487	.1477	-.2111	1.000	
HE	-.6699	-.3537	-.2641	.0993	-.9490	-.2877	-.1708	.2044	-.3972	.2225	1.00

The Szeged index (Sz) is the modification of Wiener index (W) for cyclic (cycle containing) compounds. Its positive coefficients indicate that the presence of an aromatic nucleus is essential

for the exhibition of the activity. The same is the case with the density parameter d. However, the negative coefficient of MV indicates that the activity goes on decreasing with the increasing value of MV. This molar volume (MV) is one of the important polarizability parameters, thus we can safely say that polarizability plays a negative role in the exhibition of the activity.

When an indicator parameter I_{NO2} is added to the above model (Eq.2) then its statistics is significantly improved. Thus, the resulted tetra-parametric model containing Sz, MV, d and I_{NO2} gives the following mode:

$\log 1/C = 8.6451 + 0.0025(\pm 5.6663 \times 10^{-4})Sz - 0.0408(\pm 0.0095)MV + 1.2998(\pm 0.3620)d - 0.5351(\pm 0.2620)I_{NO2}$ (3)

$n = 39$, $Se = 0.3400$, $R = 0.7450$, $R^2_A = 0.5027$, $F = 10.602$

The negative coefficient of the indicator parameter I_{NO2} indicates that the presence of an NO_2 group is not favorable for the exhibition of the activity. The physical significance of the remaining three parameters (Sz,MV and d) is the same as discussed for the model expressed by eq(2).

Table 5. Regression parameters and quality of correlation of the proposed models

Model	Parameters	n	Se	R	R^2_A	F
1.	Sz, ST	39	0.384	0.633	0.367	12.015
2.	Sz, MV, d	39	0.355	0.707	0.458	11.686
3.	Sz, MV, d, I_{NO2}	39	0.340	0.745	0.503	10.602
4.	Sz, MV, d, $I_{NO2,}$ ST	39	0.312	0.798	0.582	11.581
5.	Sz, MV, d, $I_{NO2,}$ ST, I_P	39	0.292	0.825	0.620	11.345
6.	Sz, MV, $I_{NO2,}$ ST, I_P	35	0.213	0.911	0.800	28.288

The successive regression analysis indicates that a penta-parametric model containing Sz,MV,d, $I_{NO2,}$ and ST yielded still better results:

$\log 1/C = 5.6033 + 0.0029(\pm 5.3481 \times 10^{-4})Sz - 0.0359(\pm 0.0089)MV + 0.6914(\pm 0.3991)d - 1.0246(\pm 0.2997)I_{NO2} + 0.0404(\pm 0.0148)ST$ (4)

$n = 39$, $Se = 0.3117$, $R = 0.7981$, $R^2_A = 0.5820$, $F = 11.581$

The physical significance attached to Sz, MV, d, and I_{NO2} are the same as discussed for the model expressed by equation 3. The positive coefficient of ST indicates that the activity goes on increasing with the increase in the magnitude of ST. In our recent publication[31] we have stated that surface tension (ST) can be considered as an inverse steric parameter, which thus is responsible for the improved statistics of above model.

The quality of above model is significantly improved by the addition of an indicator parameter I_P. This hexa-parametric model containing Sz, MV, d, I_{NO2}, ST and I_P is found as below:

$\log 1/C = 0.2971 + 0.0035(\pm 6.1001 \times 10^{-4})Sz - 0.0394(\pm 0.0087)MV + 0.6326(\pm 0.3821)d - 1.2293(\pm 0.3147)I_{NO2} + 0.0533(\pm 0.0154)ST - 0.2669(\pm 0.1283)I_P$ (5)

$n = 39$, $Se = 0.2971$, $R = 0.8248$, $R^2_A = 0.62003$, $F = 11.345$

The physical significances attached with Sz, MV, d, I_{NO2}, and St are the same as discussed above. The added indicator parameter I_P has a negative coefficient. This parameter is responsible for the presence of substituents at the para-position. Thus, its negative coefficient indicates that substitution at the para-position is not favorable for the exhibition of the activity.

Further, analysis of equation 5 indicates that compounds **1,33,38** and **39** gave high residues i.e., the difference between observed and calculated log1/C. Thus, they can be considered as outliers and can be removed from the regression procedure. When we did so a tremendous improvement in the statistics was observed so that the correlation coefficient increases from 0.8248 to 0.9110. Furthermore, this improved model requires less correlating parameters. This is a penta-parametric model containing Sz, MV, I_{NO2}, ST and I_P and is found as below:

$\log 1/C = 4.0682 + 0.0037(\pm 4.3912 \times 10^{-4})Sz - 0.0355(\pm 0.0062)MV + 0.0663(\pm 0.0096)ST - 1.3108(\pm 0.2282)I_{NO2} - 0.1817(\pm 0.0936)I_P$ (6)

$n = 35$, $Se = 0.2127$, $R = 0.9110$, $R^2_A = 0.8005$, $F = 28.288$

The physical significance of parameters contained in the model is the same as discussed above. And there is no significant improvement in the statistics when the parameter d is added to the above equation. Instead, the resulting six parametric model suffers from the defect in that coefficient of the d (0.0219) was smaller than its standard deviation (0.3051). Such models are not allowed statistically.[22,29-33]

In order to confirm our findings we have compared the calculated log1/C values from equations 5 and 6 with the observed values of log1/C. Such a comparison is shown in Table 6 and demonstrated in Figures 2 and 3. The residual, that is the difference between observed and calculated log1/C indicates that the model expressed by equation 6 is the best for modeling log1/C. Further, evidence in our favour are obtained by calculating predictive correlation coefficients (R_{pred}) using Figures 2 and 3, which comes out to be 0.6802 and 0.8298 respectively, supporting that the model expressed by eq(6) is the best model.

Table 6. Comparison of observed and calculated antibacterial activity against *E-Coli* using different models

S. No	log1/C	Calc log1/C	Residual	Calc log1/C	Residual
1.	4.35	4.92	-0.57	5.23*	-0.88
2.	4.45	4.68	-0.23	4.74	-0.29
3.	4.35	4.41	-0.06	4.56	-0.21
4.	4.47	4.72	-0.25	4.89	-0.42
5.	4.49	4.60	-0.11	4.86	-0.37
6.	4.66	4.43	0.23	4.56	0.10
7.	4.46	4.54	-0.08	4.58	-0.12
8.	4.60	4.62	-0.02	4.66	-0.06
9.	4.80	5.01	-0.21	4.98	-0.18
10.	4.80	4.64	0.16	4.80	-0.00
11.	4.80	4.83	-0.03	4.91	-0.11
12.	4.89	4.96	-0.07	5.05	-0.16
13.	4.89	4.96	-0.07	4.96	-0.07
14.	4.99	5.08	-0.09	4.98	0.01
15.	4.95	5.00	-0.05	4.93	0.01
16.	5.04	5.19	-0.15	5.03	0.01
17.	5.10	5.15	-0.05	5.15	-0.05
18.	4.95	5.15	-0.20	5.06	-0.11
19.	5.60	5.52	0.08	5.67	-0.07
20.	5.25	5.14	0.10	5.13	0.12
21.	5.40	5.19	0.21	5.27	0.13
22.	6.00	6.07	-0.07	6.32	-0.32
23.	5.70	5.13	0.57	5.38	0.32
24.	6.00	5.48	0.52	5.73	0.27
25.	5.85	5.54	0.31	5.82	0.02
26.	4.32	4.21	0.11	4.31	0.01
27.	4.80	4.71	0.09	4.78	0.02
28.	4.80	4.62	0.18	4.69	0.11
29.	5.40	5.10	0.30	5.23	0.17
30.	5.55	5.30	0.25	5.34	0.21
31.	4.77	4.83	-0.06	5.03	-0.26
32.	5.10	4.67	0.42	4.87	0.23
33.	5.10	5.56	-0.46	5.84*	-0.74
34.	5.55	5.01	0.54	5.31	0.24
35.	5.41	5.57	-0.16	5.71	-0.30
36.	5.64	5.55	0.09	5.74	-0.10
37.	5.32	5.46	-0.14	5.70	-0.38
38.	5.55	5.07	0.47	5.07*	0.48
39.	5.68	5.11	0.57	4.95*	0.73

*Data point not incorporated in calculations.

[a]Calculated from eq. 5; [b]Calculated from eq. 6

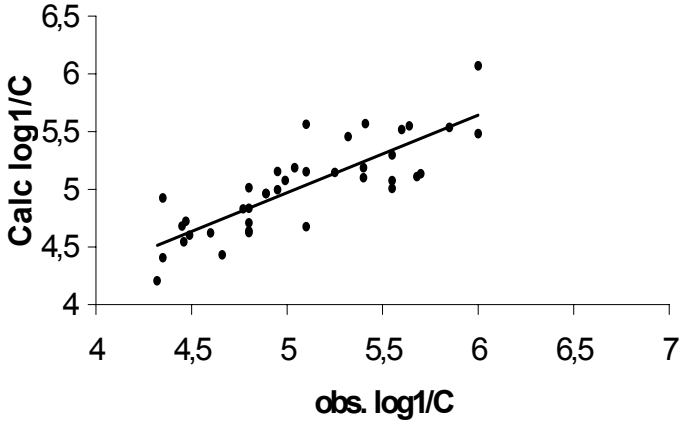

Figure 2. Graph obtained between observed and calculated log1/C from (eq.5).

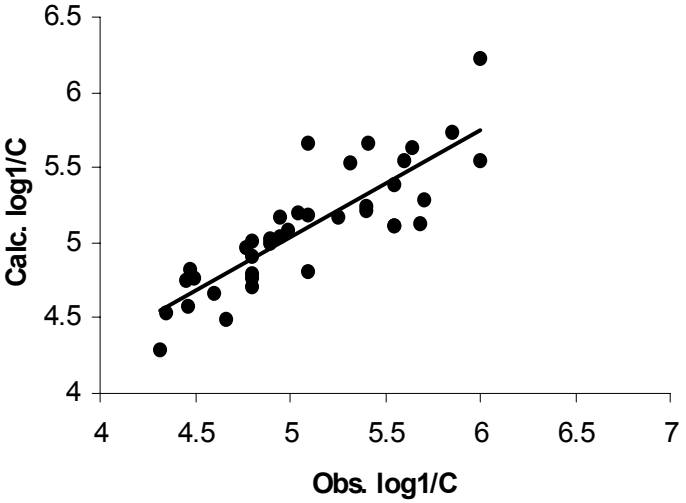

Figure 3. Graph obtained between observed and calculated log1/C from (eq.6).

In order to investigate the role of hydrophobic parameter (logP) for modeling the antibacterial activity of the sulfonamides against *E.Coli* we have used logP as one of the correlating parameters. In majority cases we obtained model in that the coefficients of logP term was considerably smaller than its standard deviation. Such models are not allowed statistically.[22,29-33]

Conclusions

From the results and discussion made above we conclude that the distance-based topological indices can be used successfully for modeling inhibition of *E.Coli* by sulfonamide and that for the present set of sulfonamides Szeged index (Sz) is found to be prominent. Also, that this index (Sz) yields statistically significant models upon combination with other molecular descriptors. Like earlier cases[4-15] here also a model containing five descriptors yielded excellent results. The consistent increase in R^2_A value as we pass from bi- to penta-parametric models supports this conclusion.

Experimental Section

Antibacterial activity against E-Coli. The set of 39 sulfonamides used in the present study is given in Table1. The MIC values expressed as log1/C for the inhibition of *E-Coli* by these sulfonamides were taken from the literature[1].

Topological indices. All the used topological indices were calculated using all hydrogen suppressed graph by deleting all the carbon hydrogen as well as hateroatomic hydrogen bonds from the structure of the sulfonamides. The calculations of these indices are well documented in the literature and therefore, their detailed calculations are not given here. However, below we have given the final expression for the calculation of these indices.

Wiener index (W)

Wiener index W = W(G) of G is defined[16] as the half-sum of the elements of the distance matrix:

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

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

Szeged index (Sz)

The Szeged index, Sz = Sz(G), is calculated[17-19] according to the following expression:

$$Sz = Sz(G) = \sum n_u \cdot n_v$$

edges 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 equidistance from both the ends of an edges, e = uv are not taken into account.

The connectivity index($^1\chi$)

The connectivity index $^1\chi = {}^1\chi(G)$ of G is defined[20] by Randic as:

$$^1\chi = {}^1\chi(G) = \sum_{i,j} [\, d(i) \cdot d(j)]^{-0.5}$$

Balaban index (J)

The Balaban index J = J(G) of G is defined[21] as:

$J = M/\mu+1 \sum(d_i.d_j)^{-0.5}$ Bonds

where M is the number of bonds in G, μ is the cyclomatic number of G, and di (i = 1,2,3,….N; N is the number of vertices in G) is the distance sum. The cyclomatic number μ = μ(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 case of monocyclic graph μ = 1 otherwise it is calculated by means of the following expression

M = M-N+1

Physicochemical parameters. various physicochemical parameters (descriptors) Viz, MR (molar refraction), MV (molar volume), Pc (parachor), η(refractive index), Surface tension (ST), density (d), α (polarizability) and logP (logarithm of octanol/water partition coefficient) were calculated using ACD Labs software.[35] The expressions for the calculation of these parameters are available in the literature.

Regression analysis. We have used maximum R^2 method[22-24] and adopted step-wise regression for obtaining statistically significant models.

Reference and Notes

1. Hansch,C;Kurup,A;Garg,R;Gao,H; Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms; Chem.Rev **2001**, *101*, 619-672.
2. Balaban, A.T.; Basak, S.C.; Beteringh, A.; Mills, D.; Supuran, C.T., QSAR Study using Topological Indices for Inhibition of Carbonic Anhydrase II by Sulfanilides and Schiff's Bases, Molecular Diversity, **2004**, *8*, 401-412.
3. Kumar, K.; Bindal, M.C.; Singh, P.; Gupta, S.P., Effect of Molecular Size on Carbonic Anhydrase Inhibition by Sulfonamides, Int. J. Quant. Chem., **1981**, *20*, 123-129.
4. Agrawal, V.K.; Bano, S.; Supuran, C.T.; Khadikar, P.V., Eur. J. Med. Chem., **2004**, *39*, 593-598.
5. Jaiswal, M.; Khadikar, P.V.; Scozzafavo, A.; Supuran, C.T., Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides Bioorg. Med. Chem. Lett., **2004**, *14*, 3283-3290.
6. Jaiswal, M.; Khadikar, P.V.; Supuran, C.T., Bioorg. Med. Chem. Lett., **2004**, *14*, 2477.
7. Thakur, M.; Thakur, A.; Khadikar, P.V.; Supuran, C.T.; Sudele, P., QSAR study on Benzene Sulphonamide Carbonic Anhydrase Inhibitors:Topological Approach Using Balaban Index. Bioorg. Med. Chem. **2004**, *12*, 4, 789-793.

8. Agrawal, V.K.; Shriwastava, S.; Khadikar, P.V.; Supuran, C.T., Quantitative structure–Activity relationship study on sulfanilamide schiff's bases: carbonic anhydrase (CA) inhibitors Bioorg.Med. Chem, **2003**,*11*, 5353-5362
9. Saxena, A.; Agrawal, V.K.; Khadikar, P.V., Oxid. Commun., **2003**, *26*, 1.
10. Agrawal, V.K.; Khadikar, P.V., Modelling of carbonic anhydrase inhibitory activity of sulfonamides using molecular negentropy Bioorg. Med. Chem Lett, **2003**,*13*, *447-453*
11. Agrawal, V.K.; Sharma, R.; Khadikar, P.V, Quantitative structure–Activity relationship studies on 5-phenyl-3-ureido-1,5-benzodiazepine as cholecystokinin-A receptor antagonists Bioorg. Med. Chem., **2002**, *10*, 3571-3581.
12. Agrawal, V.K.; Shriwastava, R.; Khadikar, P.V., QSAR Studies on some antimalarial sulfonamides Bioorg. Med. Chem., **2001**, *9*, 3287-3293.
13. Jaiswal, M.; Khadikar, P.V.; Supuran, C.T., Topological modeling of lipophilicity, diuretic activity, and carbonic inhibition activity of benzene sulfonamides: a molecular connectivity approach Bioorg. Med. Chem. Lett., **2004**. *14*, 5661-5666.
14. Thakur,M; Agrawal,A; Thakur,A; Khadikar,PV, QSAR study on phenolic activity: need of positive hydrophobic term (log*P*) in QSAR *Bioorg. and Med. Chem.*, **2004**, *12*, 2287-2293.
15. Mandloi. D.; Joshi, S.; Khadikar, P.V.; Khosla, N., QSAR study on the antibacterial activity of some sulfa drugs: building blockers of Mannich bases.,Bioorg. Med. Chem. Lett., **2005**, *15*, 405-411.
16. Wiener,H. Structural Determination of Paraffin Boiling Points, J. Am.Chem. Soc., **1947,** *69*, 17.
17. Khadikar, P.V.; Deshpande, N.V.; Kale, P.P.; Dobrynin, A.; Gutman, I.; Domotor,G. The Szeged Index and An Analogy With Wiener Index, J. Chem. Inf. Compt. Sci, **1995**, *35*, 547-550.
18. Khadikar,P.V.; Kale,P.P.; Deshpande,N.V.; Karmarkar,S.; Agrawal,V.K.; Szeged Index of Hexagonal Chains*,* Commun. Math. Comput. Chem. (MATCH) **2001**, *43*, 7-15.
19. Gutman, I.; Formula for Wiener Number of Trees and Its Extensions to Graphs Containing Cycles, *Graph Theory Notes, New York*, **1994**, *27*, 9-15.
20. Randic,M.; On Characterization of Molecular Branches; *J. Am. Chem. Soc*., **1975**, *97*, 6609-6615.
21. Balaban,A.T.; Highly Discriminating Distance Based Topological Indices; *Chem. Phys. Lett.* **1982**, *89*,399-404.
22. Chaterjee,S.; Hadi,A.S.; Price,B., *Regression Analysis by Examples*, 3rd ed. Wiley VCH: New york, **2000**.
23. Waterbeemd, H., Vande (Ed) Chemometric Method in Molecular Design, in Methods and Principles in Medicinal Chemistry, Vol2, VCH, Weinleium (Ger), **1995**.
24. Devillers, J.; Karcher, W., (Eds) Applied Multivariate Analysis in SAR and Environmental Statistics, Vol.2, in Chemical and environmental Sciences, Klawer Academic Pub. Dorderch (Netherland), **1991**.

25. Balaban, A.T.; Using Real Numbers as Vertex Invariant for Third Generation Topological Indices; *J. Chem. Inf. Comput. Sci.,* **1992**, *32*, 23-28.
26. Balaban, A.T.; Balaban, T.S.; New Vertex Invariant and Topological Indices of Chemical Graphs Based on Information on Distance; *Math. Chem*., **1991**, *8*, 383-397**.**
27. Khadikar, P.V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovit, I.; Kaveeshwar, M.,A QSAR Study of The Effect of Benzohydroxamic Acids on DNA Synthesis, *Bull. Soc. Chim. Belg.*, **1997**, *106*, 767-772.
28. Randic, M.; Comparative Regression Analysis- Regressions Based on Single Descriptor, *Croat. Chem. Acta*. **1993**, *66*, 289-312.
29. Thakur, M.; Thakur, A.; Khadikar, P.V,; Supuran, C.T., QSAR Study on pKa Vis-a- Vis Physiological Activity of Sulfonamides: A Domination Role of Surface Tension, *Bioorg. Med. Chem. Lett*., **2005**, *15*, 203-209.
30. Khadikar, P.V.; Singh, S.; Jaiswal, M.; Mandloi, D.; Topological estimation of electronic absorption bands of arene absorption spectra as a tool for modeling their toxicity and environmental pollution**,** *Bioorg. Med. Chem. Lett*., **2004**, *14*, 4795-4801.
31. Khadikar, P.V.; Singh, S.; Mandloi, D.; Joshi, S.; Bajaj, A.V., QSAR study on bioconcentration factor (BCF) of polyhalogented biphenyls using the PI index, *Bioorg. Med. Chem*., **2003**, *11*, 5045-5050.
32. Thakur, A.; Thakur, M.; Khadikar, P.V., Topological modeling of benzodiazepine receptor binding**,** *Bioorg. Med. Chem.*, **2003**, *11*, 5203-5207.
33. Khadikar, P.V.; Lukovit, I.; Agrawal, V.K.; Shriwastava, S.; Jaiswal, M.; Gutman, I.; Karmarkar, S.; Shriwastava, A., Equilized Electro-Negativity and Topological Indices: Application for Modeling Toxicity of Nitrobenzen Derivatives, *Indian J. Chem*. **2003**, *42 A*, 1436-1441.
34. HyperChem 7 Demoversion from *WWW.Hyperchem.com*.
35. ACD-Lab software for calculating the referred physicochemical parameters; Chemsketh S.O, *WWW. Acdlabs.com*.