Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 3499–3507

BIOORGANIC & MEDICINAL CHEMISTRY

Prediction of Lipophilicity of Polyacenes Using Quantitative Structure–Activity Relationships

Padmakar V. Khadikar,[a,*] Vijay K. Agrawal[b] and Sneha Karmarkar[a]

[a]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd., 3 Khatipura, Indore-452 007, India*
[b]*QSAR and Computer Chemical Lab, A.P.S. University, Rewa-486 003, India*

Received 13 May 2002; accepted 13 June 2002

Abstract—Predictive models for the lipophilicity (logP) of first 25 derivatives of polyacenes are reported. The models are derived from distance-based numerical descriptors which encode information about topology of each compounds in the data set. A new PI-type index called Sadhna index and abbreviated as Sd is introduced for the first time, and its relative correlation potential is established using the results obtained from Wiener (W), Szeged (Sz), first-order Randic connectivity (χ), and Padmakar–Ivan indices. The data show that lipophilicity (logP) is best modelled in bi-parametric model containing PI and Sd indices. The effect due to size, shape, branching, steric and polarity effects on the exhibition of lipophilicity is critically discussed. The predictive ability of the models is discussed on the basis of cross-validation parameters.

Introduction

The basic assumption of Quantitative Structure–Activity/Property Relationships (QSARs/QSPRs) is that there are some quantitative relationships between molecular structure and activities (properties) of a molecule.[1,2] The term structure does not necessarily mean the spatial arrangement of atoms in a molecule itself, but physico-chemical properties inherent in that arrangement.[3]

The polyacenes, L_h, with h-hexagones, are linearly para-annealated rather than the chemical and benzenoids which posses translational symmetry (Fig. 1).

The chemistry of polyacenes is very much of interest to synthetic chemists, environmental chemists, cancer research chemists, structural chemists, etc.[4–8] These facts motivate our concern for QSPR in that we investigate now topological prediction of lipophilicity (logP) of polyacenes. Such a study as discussed below, has not been attempted so far.

The lipophilicity expressed by the logarithmic partition coefficient (logP) is a very important physico-chemical parameter which describes a partitioning equilibrium of solute molecules between water and an immisible organic solvent[9–11] It is of particular importance in drug design not only because it encodes a wealth of structural information in spite of its success in describing hydration–dehydration effect, logP has the bias of all empirical parameter. It can be determined either from the costly and time consuming experiments or from the approximate empirical formula with limited reliability.[12] However, the objective of the present study is not to introduce another method for the determination of logP, but to use topological indices for predicting lipophilicity (logP) of a series of polyacenes. Therefore, it is quite meaningful to seek the theoretical parameters (topological indices in our case) which may model lipophilicity, that is, logP of polyacenes under present study.

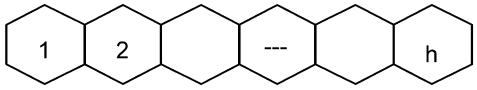


Figure 1. The structure of polyacenes, hence, L_1 = benzene, L_2 = naphthaline, L_3 = anthracene, and so on.

In view of the above and to fulfil our objective we introduce here a new topological index for predicting lipophilicity (logP) of the polyacenes used in the present investigation. This index is named as Sadhna index and abbreviated as Sd. This index was conceived through prolonged experiments on the newly introduced Padmakar–Ivan (PI) index.[13–15] While attempting an elementary cut

*Corresponding author. Tel.: +91-731-531906; e-mail: vijay-agrawal@lycos.com

0968-0896/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0968-0896(02)00226-2

Table 1. The newly introduced Sadhna index (Sd) and other related indices used for comparison

No.	Name	Abbreviation	References
1	Sadhna index	Sd	Newly introduced
2	Wiener index	W	18
3	Szeged index	Sz	16,17
4	Padmakar–Ivan index	PI	13–15
5	First-order Randic Connectivity index	χ	19

method for the estimation of PI index (see Experimental for details) we have conceived the Sd index. This index, like the Szeged index,[16,17] is applicable only to cyclic graphs. In the case of polyacenes, this new index, Sd, is defined as below:

$$\mathrm{Sd} = 2h(5h + 1) \quad (1)$$

where, h is the number of hexagons involved in the polyacene molecule (for detail see Experimental). The estimation of Sd for other polycyclic systems is underway and the results will be published elsewhere.

Another objective of the present investigation is to discuss relative potential of Sd index compared to other related topological indices in modelling lipophilicity (logP) of polyacenes. Such indices used in the present investigation are given in Table 1. All these indices, like the Sd index, are distance-based topological indices. The results as given below show that the Sd index can be successfully used for predicting lipophilicity (logP) of the polyacenes used. Furthermore, as shown below, the Sd index is found to be better than the Wiener[18] (W), Szeged[16,17] (Sz), and PI[13–15] indices, but is bettered by the first-order-Randic connectivity index (χ)[19] (Table 1).

Results and Discussion

The topological indices, namely W, Sz, PI, χ and Sd, for the first 25 members of the polyacene series are summarised in Table 2. Table 2 also records the lipophilicity (logP) of the polyacenes used. The correlation of these five descriptors along with their correlations with lipophilicity (logP) are shown in Table 3. It shows that all the five topological indices are highly correlated. The correlation ranges between 0.9273 and 1.0000. The same is

Table 2. The polyacenes, their lipophilicity (logP), the newly introduced Sadhna Index (Sd) and other related indices used for comparison

Polyacene	LogP	MV	α	W	Sz	PI	χ	Sd
L_1	2.202	26.3	10.40	27	54	24	3	12
L_2	3.396	129.5	17.48	109	243	96	4.967	44
L_3	4.59	157.6	24.55	279	640	216	6.933	96
L_4	5.784	191.7	31.62	519	1381	384	8.899	168
L_5	6.978	225.8	38.70	1011	2506	600	10.866	260
L_6	8.172	260.5	45.77	1037	4119	864	12.832	372
L_7	9.366	294.0	52.84	2479	6308	1176	14.798	504
L_8	10.56	382.2	59.92	3569	9161	1536	16.765	656
L_9	11.754	362.2	66.99	4939	12,766	1944	18.731	828
L_{10}	12.948	396.4	74.06	6621	17,211	2400	20.697	1020
L_{11}	14.142	430.5	81.14	8647	22,584	2904	21.663	1232
L_{12}	15.336	464.7	88.21	11049	28,933	3456	24.63	1464
L_{13}	16.53	498.8	95.28	13,859	36,466	4056	26.596	1716
L_{14}	17.724	532.9	102.36	17,109	45,151	4704	28.562	1988
L_{15}	18.918	567.0	109.43	20831	55,116	5400	30.529	2280
L_{16}	20.112	601.0	116.50	25,057	66,449	6144	32.495	2592
L_{17}	21.306	635.2	123.58	29,819	79,230	6936	34.461	2924
L_{18}	22.506	669.4	130.65	35,149	93,571	7776	36.428	3276
L_{19}	23.614	703.5	137.72	41,079	109,536	8664	38.394	3648
L_{20}	24.88	737.6	144.80	4764	127,221	9600	40.36	4040
L_{21}	26.082	771.7	151.87	54,867	146,714	10,584	42.327	4452
L_{22}	27.276	805.8	158.94	62,789	168,103	11,616	44.293	4884
L_{23}	28.47	839.9	166.02	71,439	191,476	12,696	46.259	5336
L_{24}	29.664	874.0	173.09	80,849	212,505	13,824	48.226	5808
L_{25}	30.858	908.2	180.16	91,051	244,526	15,000	50.192	6300

Table 3. Correlation table for the parameters used in Table 2

	LogP	MV	W	Sz	PI	χ	Sd
LogP	1.0000						
MV	0.9978	1.0000					
W	0.9265	0.9173	1.0000				
Sz	0.9258	0.9166	0.9999	1.0000			
PI	0.9707	0.9629	0.9887	0.9885	1.0000		
χ	0.9999	0.9977	0.9273	0.9267	0.9713	1.0000	
Sd	0.9712	0.9634	0.9885	0.9882	1.0000	0.9718	1.0000

Table 4. Proposed regression models for modelling lipophilicity (logP) of polyacenes.

Mono-parametric models
(1) $\log P = 9.1705 + 2.9079\times10^{-4}(2.4628\times10^{-5})W$
(2) $\log P = 9.\ 2100 + 1.0875\times10^{-4}(9.2576\times10^{-6})Sz$
(3) $\log P = 6.9662 + 0.0018(9.2970\times10^{-5})PI$
(4) $\log P = 0.4221 + 0.6064(0.0018)\chi$
(5) $\log P = 6.9192 + 0.0043(2.1992)Sd$

Bi-parametric models
(6) $\log P = 4.3032 + 0.0045\ (2.4127\times10^{-4})PI + 4.6500\times10^{-4}(4.0783\times10^{-5})W$
(7) $\log P = 4.2782 + 0.0045(2.3645\times10^{-4})PI - 1.7236(1.4958\times10^{-5})Sz$
(8) $\log P = 0.3720 - 1.603810^{-5}\ (2.3006\times10^{-5})PI + 0.6115(0.0075)\ \chi$
(9) $\log P = 1.0123 - 0.2486\ (4.0042\times10^{-4})PI + 0.5965(9.5410\times10^{-4})Sd$
(10) $\log P = 4.2390 + 0.0107(5.5858\times10^{-4})Sd - 4.5750\times10^{-4}\ (3.9626\times10^{-5})W$
(11) $\log P = 4.2149 + 0.0106(5.4752\times10^{-4})Sd - 1.6959\times10^{-4}\ (1.4537\times10^{-5})Sz$
(12) $\log P = 0.3721 - 3.8381\times10^{-5}\ (5.5224\times10^{-5})Sd + 0.6116(0.0070)\chi$

Monoparametric models based on MV and α
(13) $\log P = -0.6480 + 0.0341(4.7353\times10^{-4})MV$
(14) $\log P = 0.4462 + 0.1695(7.0309\times10^{-6})\ \alpha$

found to be the case when these topological indices are correlated with lipophilicity (logP). The correlation of Sd with W, Sz, PI and $^1\chi$ follows the following sequence:

$$PI > W > Sz > \chi \quad (2)$$

The data presented in Table 3 shows that:

(i) χ is the excellent topological index for predicting lipophilicity (logP);
(ii) Sd is the next topological index which gives slightly better results than the PI, W and Sz indices for predicting logP;
(iii) W and Sz have more or less similar predicting potential.

The observed high collinearity among W, Sz, PI, χ, and Sd index means that they are not independent variables for the other one(s). Also that, in general, two or more variables which are highly inter-correlated are not used simultaneously in the multiple regression analysis as they may suffer from the defect due to collinearity.[20,21] However, the results as discussed below show that bi-parametric model involving such combinations do not suffer from the defect due to collinearity. This is in accordance with the results reported by Randic[22] and is well explained below.

Simple regression resulted in five mono-parametric models for modelling the lipophilicity (logP). These models are presented in Table 4. The quality of regression expression (Table 5) show that the Sd index is a better index than the W, Sz and PI indices for modelling lipophilicity (logP) and is worse than the χ index.

The mono-parametric regression expressions based on Sd and χ indices are found as:

$$\log P = 6.9192 + 0.0043(\pm2.1992\times10^{-4})\ Sd \quad (3)$$

$n = 25$, SE = 2.1393, $R = 0.9712$, $F = 381.711$, $Q = 0.4540$

$$\log P = 0.4221 + 0.6064(\pm0.0018)\chi \quad (4)$$

$n = 25$, SE = 0.1254, $R = 0.9999$, $F = 117{,}845.94$, $Q = 7.9737$

These regression expressions correspond to models (5) and (4), respectively, in Table 4.

In the above equations, the parameter Q is called the quality factor and is defined[23] in the literature as the ratio of correlation coefficient (R) to the standard error of estimation (SE), that is, $Q = R/SE$, meaning thereby the higher the value of R, the lower the SE, the higher will be Q, and the higher will be the quality of model.

In spite of the fact that we have obtained statistically excellent mono-parametric models, we attempted bi-parametric models also for modelling lipophilicity (logP) of the polyacenes used. The results are shown in Table 4. The quality of the corresponding regression, as shown in Table 5, indicate that the bi-parametric model involving PI and Sd is the most appropriate for modelling lipophilicity (logP). This regression expression (5), which corresponds to model (9) in Table 4 is found as:

Table 5. The result of regression analysis: regression parameters and quality of correlations for modelling lipophilicity (logP) of polyacenes

Model no.	Parameters	SE	R	R^2_A	F	Q
1	W	3.3769	0.9265	—	139.413	0.2744
2	Sz	3.3918	0.9258	—	137.997	0.2730
3	PI	2.1548	0.9707	—	375.903	0.4505
4	χ	0.1254	0.9999	—	117,845.014	7.9737
5	Sd	0.1393	0.9712	—	381.711	0.4540
6	PI, W	0.8382	0.9958	0.9909	1307.141	1.1880
7	PI, Sz	0.8306	0.9959	0.9911	1331.266	1.1990
8	PI, χ	0.1268	0.9999	0.9998	57,605.837	7.8857
9	PI, Sd	0.0165	1.0000	1.0000	3.3901×10^6	60.6060
10	Sd, W	0.8233	0.9960	0.9912	1355.328	1.2098
11	Sd, Sz	0.8159	0.9960	0.9914	1380.048	1.2207
12	Sd, χ	0.1268	0.9999	0.9998	57,598.583	7.8857
13	MV	0.5392	0.9986	—	5297.698	1.8520
14	α	0.0020	1.0000	—	5.7639×10^8	500.000

$$\log P = 1.0123 - 0.2486(\pm 4.0042 \times 10^{-4})PI + 0.5965(\pm 9.5410 \times 10^{-4})Sd \quad (5)$$

$n = 25$, SE = 0.0165, $R = 1.0000$, $F = 7.9737 \times 10^6$, $Q = 60.6060$

The data presented in Tables 4 and 5 show that the models based on binary combinations of: (i) PI, χ and (ii) Sd, χ have similar statistics and that these models are slightly worsened than the model discussed above.

It is worth mentioning that the sign of the regression coefficient is important because it can contribute to the understanding of the drug-action mechanism and/or give us the useful information about the drug design. Table 4 shows that the signs of the PI, W indices in model (6) are positive while in other models the terms involved have positive and negative coefficients. The positive coefficients of PI and W indicate that these quantities should be large for the exhibition of lipophilicity (logP). In model (7), the Sz index has a negative coefficient, which means that large values of PI and lower Sz favour lipophilicity (logP). In models (8) and (9), the sign of PI is changed to negative. This may now be attributed to very high collinearity between PI and χ as well as Sd. The same is found to be the case with other models.

At this stage, it is worth mentioning that the positive coefficients encode contributions due to size, shape, branching, polarity and steric effect. Out of these, the contributions due to size, shape and branching is taken care of by the distance-based topological indices used in the present study. In order to account for steric and polarity effect, we have used molecular volume (MV) and polarizability (α) for the purpose. MV is related to the geometric locus around molecules where repulsive and attractive interactions with approaching atoms balance each other. The correlation in Table 3 shows that logP, MV and χ are highly linearly correlated; the correlation potential of PI and Sd indices in the exhibition of both logP and MV are similar and, comparatively, W and Sz are less significant for this purpose.

The data presented in Tables 6 and 7 show that here also the same three types of bi-parametric models are excellent for modelling MV. Table 7 shows that the quality of these three excellent bi-parametric models involving: (i) PI, χ and (ii) PI, Sd and (iii) Sd, χ are similar. This has prompted us to re-examine the correlation between logP, Sd, $^1\chi$, MV, and α and then to investigate regression models so as to offer contribution(s) of size, shape, branching, steric, and polar effect in exhibiting lipophilicity (logP) of the polyacenes (Table 8). Regression analysis shows that contribution of MV and towards lipophilicity (logP) is found as below:

$$\log P = -0.6480 + 0.0341(\pm 4.7353 \times 10^{-4})MV \quad (6)$$

$n = 25$, SE = 0.5392, $R = 0.9986$, $F = 5297.698$, $Q = 1.8520$

Table 6. Proposed regression models for modelling molar volume (MV) of polyacenes

Mono-parametric models	
(15)	$MV = 288.0238 + 0.0083(7.5371 \times 10^{-4})W$
(16)	$MV = 289.1652 + 0.0031(2.8323 \times 10^{-4})Sz$
(17)	$MV = 224.4028 + 0.0517(0.0030)PI$
(18)	$MV = 33.9474 + 17.4983(0.2479)\chi$
(19)	$MV = 223.0402 + 0.1233(0.0072)Sd$
Biparametric models	
(20)	$MV = 144.0453 + 0.1338\ (0.0106)PI + 0.0140(0.0018)W$
(21)	$MV = 143.2468 + 0.1330(0.0104)PI\text{-}0.0052(6.575010^{-4})Sz$
(22)	$MV = 15.5493 + 0.0059\ (0.0030)PI + 19.3683(0.9841)\chi$
(23)	$MV = 36.0163–7.8696\ (0.3977)PI + 18.8744(0.9477)Sd$
(24)	$MV = 142.0886 + 0.3158(0.0246)Sd - 0.0138(0.0017)W$
(25)	$MV = 141.3169 + 0.3140(0.0242)Sd - 0.0051(6.433 \times 10^{-4})Sz$
(26)	$MV = 16.6022–0.0141(0.0072)Sd + 19.3825(0.9913)\chi$

Table 7. The result of regression analysis: regression parameters and quality of correlations for modelling Molar Volume (MV) of polyacenes

Model no.	Parameters	SE	R	R^2_A	F	Q
15	W	103.3461	0.9173	—	122.016	0.0089
16	Sz	103.7683	0.9166	—	120.838	0.0088
17	PI	70.0411	0.9629	—	292.717	0.0136
18	χ	17.5926	0.9977	—	1981.293	0.0567
19	Sd	69.6079	0.9634	—	296.679	0.0138
20	PI, W	36.7249	0.9904	0.9791	563.187	0.0270
21	PI, Sz	36.5111	0.9905	0.9793	569.931	0.0272
22	PI, χ	16.6017	0.9980	0.9957	2798.767	0.061
23	PI, Sd	16.4171	0.9981	0.9958	862.306	0.0608
24	Sd, W	36.3190	0.9906	0.9796	576.093	0.0273
25	Sd, Sz	36.1115	0.9907	0.9798	582.859	0.0274
26	Sd, χ	16.6022	0.9980	0.9957	2798.601	0.0601

Table 8. Correlation matrix for logP, PI, Sd, χ, MV and α of polyacenes

	LogP	MV	α	PI	χ	Sd
LogP	1.0000					
MV	0.9978	1.0000				
α	1.0000	0.9978	1.0000			
PI	0.9707	0.9629	0.9707	1.0000		
χ	0.9999	0.9977	0.9999	0.9713	1.0000	
Sd	0.9712	0.9634	0.9712	1.0000	0.9718	1.0000

$$\log P = 0.4462 + 0.1695(\pm 7.0309 \times 10^{-6})\alpha \quad (7)$$

$n = 25$, SE = 0.0020, $R = 1.0000$, $F = 5.7639 \times 10^8$, $Q = 500.0000$

These regression expressions [(6) and (7)] correspond to models (13) and (14), respectively, in Table 4.

These and earlier discussed results[24,25] show that size, shape, branching, steric effect and polarizability have a similar effect on exhibition of lipophilicity (logP). This finds further support from the fact that when MV and α are coupled with the PI index independently, no change

in the statistics of the resulting models occurs. Furthermore, when all the three parameters are used together, the resulting model suffered from the defect in that the coefficient of PI (or Sd) and MV terms were found considerably lower than the corresponding standard deviations. Such models are not statistically allowed. This means that in modelling lipophilicity (logP) of the polyacenes, PI, MV, and α or Sd, MV, or χ, Mv, and α are not independent parameters. The correlations shown in Table 8 also show that the role of PI, MV, and χ is similar to that of MV and α, which means that the parameters MV and α can be suitably exchanged by the combination of PI, Sd and χ indices. Hence, contribution due to size, shape, branching, steric effect and polarizability are contained in this combination.

The predictive potential of the models (8), (9), (12) and (22), (23), (26) in modelling lipophilicity (logP) and molecular volume (MV), respectively, is determined by examining the difference between the observed values of each of logP and MV and their estimated values. Such data are shown in Tables 9 and 10 for comparison. The residue, that is, the difference between observed and estimated parameters (logP and MV), indicates that the bi-parametric model based on the combination of PI and Sd is the most appropriate model for modelling both logP and MV.

It is worth recording that, in spite of the fact that the PI and Sd indices are highly correlated, their presence in the aforementioned model is considered statistically significant as their correlation coefficients are considerably lower than their respective standard deviation [see models (9) and (23)]. Randic[22] has argued that one should be particularly aware of a common fit in regression analysis in describing descriptors that are highly intercorrelated. He further stated that, by discarding one of the descriptors which commonly duplicates another, we may be discarding a descriptor that nevertheless may carry useful structural information in the part in which it does not parallel the other descriptor. Thus, following Randic,[22] we may safely say that the newly proposed Sd index carries some additional structural information not present in the other molecular descriptor with which Sd correlated highly. A detailed study in this respect is underway and will be published elsewhere.

Finally, predictive potentials of the proposed models are determined by estimating cross-validation parameters (Table 11). In all the cases, the predictive residual sum of squares (PRESS) was found to be considerably lower than the sum of the squares of regression values (SSY). Also, PRESS/SSY are found to be smaller than 0.1. The PRESS lower than SSY indicates that models predict better than chance and can be considered statistically significant. Similarly, PRESS/SSY < 0.1 indicates that the models have excellent predictive potential. In addition, R^2_{CV} indicates that the overall predictive ability of all the proposed models (Table 11) is excellent.

It is worth recording that uncertainty of prediction (S_{PRESS}) is found to be same as that of standard error of estimation (SE). It means that these parameters are not good parameters to be used in deciding the predictive ability of the model.

Consequent to the above, we have calculated the predictive square error (PSE) and observed that it is a better

Table 9. Estimated values of lipophilicity (logP) and their comparison with the observed values of lipophilicity (logP) of polyacenes (ref Table 4)

Compd	Log P Obs.	Log P estimated from: Model (8)		Model (9)		Model (12)	
		Est.	Res.	Est.	Res.	Est.	Res.
1	2.202	2.206	−0.004	2.205	−0.003	2.206	−0.004
2	3.396	3.408	−0.012	3.398	−0.002	3.408	−0.012
3	4.590	4.608	−0.018	4.592	−0.002	4.608	−0.018
4	5.784	5.808	−0.024	5.785	−0.001	5.808	−0.024
5	6.978	7.007	−0.029	6.978	0.000	7.007	−0.029
6	8.172	8.205	−0.033	8.171	0.001	8.205	−0.033
7	9.366	9.403	−0.037	9.365	0.001	9.403	−0.037
8	10.560	10.600	−0.040	10.558	0.002	10.600	−0.040
9	11.754	11.795	−0.041	11.752	0.002	11.795	−0.041
10	12.948	12.990	−0.042	12.945	0.003	12.990	−0.042
11	14.142	13.573	0.569	14.138	0.004	13.573	0.569
12	15.336	15.378	−0.042	15.332	0.004	15.378	−0.042
13	16.530	16.571	−0.041	16.526	0.004	16.571	−0.041
14	17.724	17.763	−0.039	17.719	0.005	17.763	−0.039
15	18.918	18.953	−0.035	18.913	0.005	18.955	−0.037
16	20.112	20.145	−0.033	20.107	0.005	20.145	−0.033
17	21.306	21.335	−0.029	21.300	0.006	21.335	−0.029
18	22.506	22.524	−0.018	22.494	0.012	22.524	−0.018
19	23.614	23.712	−0.098	23.688	−0.074	23.714	−0.100
20	24.880	24.899	−0.019	24.882	−0.002	24.899	−0.019
21	26.082	26.086	−0.004	26.076	0.006	26.086	−0.004
22	27.276	27.272	0.004	27.270	0.006	27.272	0.004
23	28.470	28.457	0.013	28.464	0.006	28.457	0.013
24	29.664	29.642	0.022	29.658	0.006	29.642	0.022
25	30.858	30.825	0.033	30.852	0.006	30.825	0.033

Table 10. Estimated values of Molar Volume (MV) of polyacenes and their comparison with the observed values of MV of polyacenes (ref Table 6)

Compd	MV Obs.	Estimated molar volume (MV) from:					
		Model (22)		Model (23)		Model (26)	
		Est.	Res.	Est.	Res.	Est.	Res.
1	26.250	73.513	−47.263	73.638	−47.388	73.515	−47.265
2	129.500	111.186	18.314	111.005	18.495	111.188	18.312
3	157.600	148.557	9.043	148.118	9.482	148.558	9.042
4	191.700	185.645	6.055	184.976	6.724	185.646	6.054
5	225.800	222.469	3.331	221.580	4.220	222.470	3.330
6	260.000	258.991	1.009	257.929	2.071	258.992	1.008
7	294.000	295.230	−1.230	294.024	−0.024	295.231	−1.231
8	382.200	331.206	50.994	329.865	52.335	331.206	50.994
9	362.200	366.879	−4.679	365.451	−3.251	366.879	−4.679
10	396.400	402.270	−5.870	400.782	−4.382	402.270	−5.870
11	430.500	418.009	12.491	435.859	−5.359	417.994	12.506
12	464.700	472.221	−7.521	470.682	−5.982	472.221	−7.521
13	498.800	506.763	−7.963	505.250	−6.450	506.762	−7.962
14	532.900	541.021	−8.121	539.564	−6.664	541.021	−8.121
15	567.000	575.016	−8.016	573.623	−6.623	575.016	−8.016
16	601.000	608.709	−7.709	607.428	−6.428	608.709	−7.709
17	635.200	642.119	−6.919	640.978	−5.778	642.119	−6.919
18	669.400	675.265	−5.865	674.274	−4.874	675.265	−5.865
19	703.500	708.109	−4.609	707.315	−3.815	708.109	−4.609
20	737.600	740.670	−3.070	740.102	−2.502	740.671	−3.071
21	771.700	772.968	−1.268	772.635	−0.935	772.968	−1.268
22	805.800	804.963	0.837	804.913	0.887	804.964	0.836
23	839.900	836.675	3.225	836.936	2.964	836.676	3.224
24	874.000	868.124	5.876	868.705	5.295	868.125	5.875
25	908.200	899.271	8.929	900.220	7.980	899.272	8.928

Table 11. Cross-validation parameters for the proposed models

Model	PRESS	SSY	PRESS/SSY	R^2_{CV}	S_{PRESS}	PSE
8	0.3536	1855.7711	0.0002	0.9998	0.1268	0.1189
9	0.0060	1852.1187	0.0000	0.9999	0.0003	0.00024
12	0.3536	1851.7711	0.0002	0.9998	0.1268	0.1189
22	6063.5411	1,542,767.4325	0.0393	0.9607	16.6017	15.5735
23	29327.3014	1,519,503.6717	0.0193	0.9807	16.5111	14.2493
26	6063.8995	1,542,767.0741	0.0039	0.9961	16.6022	15.5742

parameter than S_{PRESS} in deciding the predictive potential. The reason is that PSE is more directly related to the uncertainty of prediction.[20,21] The smaller the value of PSE the better is the predictive ability. The PSE values recorded in Table 11 shows that for modelling lipophilicity (logP), the PSE value is smallest for the model (9). This model is a bi-parametric model containing Sd and PI indices as the correlating parameters. Similarly, for modelling MV, model (21) has the lowest value for PSE. This model again is a bi-parametric model containing the same (Sd and PI) topological indices. These results, therefore, show that out of the pool of the topological indices used (W, Sz, PI, χ, Sd) only two indices, namely: Sd and PI are better indices for modelling lipophilicity (logP) as well as molecular volume (MV).

Conclusions

From the aforementioned results and discussion we conclude that:

(i) Distance-based topological indices W, Sz, χ, PI, and Sd can be successfully used for modelling lipophilicity (logP) and molecular volume (MV) of polyacenes;
(ii) Correlation analysis indicated that the newly introduced Sd index is better than W, Sz and PI, but is worse than χ;
(iii) In mono as well as bi-parametric correlations, the combination of Sd and PI indices is the most appropriate for the above purpose;
(iv) The lipophilicity (logP) of polyacenes is highly influenced by size, shape, branching, steric effect, and polarizability. However, their influence is similar;
(v) The results show that effect due to MV and α is similar to that of Sd and PI indices. Also, that PI, Sd, MV, and χ parameters cannot be taken as independent variable(s) for modelling lipophilicity (logP) and molar volume (MV);
(vi) The cross-validation technique establishes that the bimolecular model containing Sd and PI indices as correlating parameters has the highest predictive potential.

Experimental

Material and methods

All the logP values were calculated by the procedure desirable in the literature.[12,24,25] The methodology used for the experimental determination of logP is the same as described in the literature.[26,27]

In the regression analysis, the molecular polarizability (α) and molecular volume (MV) were used to characterise the overall contribution due to bulk effect or volume to the QSARs. Here, molecular volume is determined from the experimental values of molecular weight and density. The polyacenes used were of BDH Analar or equivalent quality.

Topological indices. All the distance-based topological indices, namely W, Sz, and, PI were calculated using the procedure described in the literature.[13,16–19] The software provided by Lukovits was used to calculate W, Sz and χ. The newly introduced Sd index was calculated using the procedure described below. No detailed information is, therefore, needed to calculate W, Sz, χ, PI and Sd indices. However, as below, we give the respective expresions for the calculation of these indices.

Sadhna index (Sd). The development of Sadhna index (Sd) through the elementary cuts for calculating PI index is described below.

Giving,

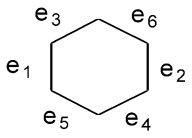
$$\mathrm{Sd} = 2h(5h+1) \quad (8)$$
$$= 2\ (\text{number of cycle present in polyacene})\ (\text{total number of edges contained in the polyacene})$$

$$\text{Therefore, Sd for } h_1 = 2(1)\ (6) = 12 \quad (9)$$

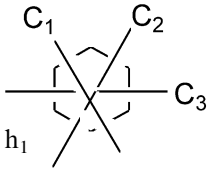
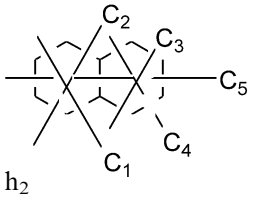
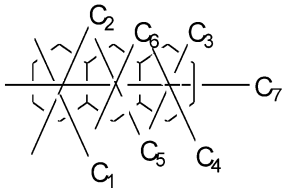
$$\text{Now, } 5h+1 = \text{total number of edges} = m$$
$$\therefore \quad \mathrm{Sd} = 2hm$$
$$\text{Also,} \quad \mathrm{Sd} = 2/5\ (m-1)m$$
$$= 2/5(\mathrm{V_E} + \mathrm{NV_E})\ (\mathrm{V_E} + \mathrm{NV_E} - 1) \quad (10)$$

where V_E is the sum of vertical edges and NV_E is the sum of non-vertical edges.

In h_1 (see the figure below); e_1 and e_2 are called the vertical edges , while e_3, e_4, e_5 and e_6



are the non-vertical edges.

Polyacene	Elementary cut	Sum of the edges on both sides of elementary cut	Number of edges involved in elementary cut	Total edges
 h_1	C_1	2+2=4	2	8
	C_2	2+2=4	2	8
	C_3	2+2=4	2	8
		Sd=12	6	24=PI
 h_2	C_1	2+7=9	2	18
	C_2	2+7=9	2	18
	C_3	2+7=9	2	18
	C_4	2+7=9	2	18
	C_5	4+4=8	3	24
		Sd=44	11	96
 h_3	C_1	2+12=14	2	28
	C_2	2+12=14	2	28
	C_3	2+12=14	2	28
	C_4	2+12=14	2	28
	C_5	6+6=12	4	48
	C_6	7+7=14	2	28
	C_7	7+7=14	2	28
		Sd=96	16	216

$$\begin{aligned}\text{Therefore, Sd} &= 2/5\ (2+4)(2+4-1)\\ &= 2/5(6)(5)\\ &= 12\\ \text{Also,}\quad \text{Sd} &= 2/5\ (m-1)m\\ &= 2/5(6-1)6\\ &= 2/5(5)(6)\\ &= 12\end{aligned}$$

Wiener index (W).[18] The Wiener index (W) of a graph G is just the sum of distances of all pairs of vertices of G:

$$\begin{aligned}\mathrm{W} = \mathrm{W}(G) &= 1/2 \sum_{v \in V(G)} \sum_{u \varepsilon V(G)} \mathrm{d}(v,u|G)\\ &= 1/2 \Sigma \mathrm{d}(v|G)\\ &\quad v \in V(G)\end{aligned} \tag{11}$$

where, $\mathrm{d}(v|G)$ is called the distance number of vertex v and is defined[17] as under

$$\mathrm{d}(v|G) = \Sigma \mathrm{d}(v,u|G) \quad u \in V(G) \tag{12}$$

Szeged index (Sz).[16,17] Let $e = uv \in E(G)$. Then we define two subsets of vertex set of G as follows:

$$\begin{aligned}N_1(e|G) &= \{x \in V(G) | \mathrm{d}(x,u|G) < \mathrm{d}(x,v|G)\}\\ N_2(e|G) &= \{x \in V(G) | \mathrm{d}(x,u|G) > \mathrm{d}(x,v|G)\}\end{aligned} \tag{13}$$

The number of elements of $N_1(e|G)$ and $N_2(e|G)$ are denoted by $n_1(e|G)$ and $n_2(e|G)$, respectively. Thus, $n_1(e|G)$ counts the vertices of G lying closer to the vertex u than to vertex v. The meaning of $n_2(e|G)$ is analogous. The vertices equidistant from both ends of the edge uv belong neither to $N_1(e|G)$ nor to $N_2(e|G)$.

Randic index (χ).[19] The connectivity index $\chi = \chi(G)$ of a graph G is defined by Randic[19] as under

$$\chi = \chi(G) = \Sigma[d_i d_j]^{-0.5} \tag{14}$$

where d_i is the valence of a vertex i, equal to the number of bonds connected to the atom i, in G, representing the graph of a compound. The meaning of d_j is analogous.

PI Index.[13–15] Let e be an edge of G connecting the vertices u and v, $e = uv \in E(G)$.

We define for $e = uv$ two quantities $n_{eu}(e|G)$ and $n_{ev}(e|G)$. $n_{eu}(e|G)$ is the number of edges lying closer to the vertex u than the vertex v, and $n_{ev}(e|G)$ is the number of edges lying closer to the vertex v than the vertex u. Edges equidistant from both ends of the edge uv are not counted.

The PI index, $\mathrm{PI} = \mathrm{PI}(G)$ of the graph G is defined[13–15] as:

$$\mathrm{PI} = \mathrm{PI}(G) = \Sigma[n_{eu}(e|G) + n_{ev}(e|G)] \quad e \in E(G) \tag{15}$$

The summation goes over all edges of G.

Regression analysis.[20,21] Simple as well as multiple regression analysis were performed using a standard in-house program (P.V.K.). In the regression equation n is the number of compounds considered, R is the correlation coefficient, SE is the standard error of estimation, F is the Fischere's ratio, Q is the quality factor, R^2_{CV} is the cross-validation correlation coefficient derived from the respective residuals sum of squares PRESS, leave one out method, S_{PRESS} is the uncertainty of prediction, PSE is the predictive squared error, SSY is the sum of the squares of response values and regression coefficients are given with 95% confidence intervals.

Acknowledgements

One of the authors (P.V.K.) would like to thank Ivan Gutman for introducing him to the fascinating field of 'Chemical Topology and Graph theory'. The authors are thankful to Istvan Lukovits for providing softwares.

References and Notes

1. *Topological Indices And Related Descriptors in QSAR and QSPR*. Devilliers, J., Balaban, A. T., Eds. Gorden and Breach Scientific Publication: Amsterdam, The Netherlands, 1999.
2. *Comparative QSAR*. Devilliers, J., Ed. Tailor and Francis: Washington, DC, 1998.
3. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
4. Gutman, I.; Cyvin, S. J. *Introduction to the Theory of Benzenoid Hydrocarbons*; Springer: Berlin, 1989.
5. Gutman, I.; Gaverilovic, N.; Bankovic, D.; Khadikar, P. V.; Deshpande, N. V.; Kale, P. P. *J. Serb. Chem. Soc.* **1994**, *59*, 519.
6. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Gutman, I. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1181.
7. Gutman, I.; Khadikar, P. V.; Khaddar, T. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 105.
8. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Natl. Acad. Sci. Lett.* **2000**, *23*, 124.
9. Ravsky, O. A.; Schaper, K. J.; Seydel, J. K. *Quant. Struct. Act. Relat.* **1995**, *14*, 433.
10. Schaper, K. J.; Zhange, H.; Raevsky, O. A. *Quant. Struct.-Act. Relat.* **2001**, *20*, 46.
11. Khadikar, P. V.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1125.
12. Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH: Weinheim, 1992.
13. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
14. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3035.
15. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 7141.

16. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
17. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
18. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
19. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
20. Draper, N.; Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley and Sons: New York, 1981.
21. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
22. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
23. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
24. Hansch, C. In *Drug Design*; Atriens, E. J., Ed. Academic: New York, 1971.
25. Hansch, C.; Leo, A. *Substitution Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
26. ACD/Chem Skech version 4.5 for Microsoft Windows; Advanced Chemistry Development.
27. Barlow, R. B. *Quantitative Aspect of Chemical Pharmacology*; Biling and Sons: Croom Helm, London, 1980.