Highly Correlating Distance/Connectivity-Based Topological Indices. 1: QSPR Studies of Alkanes

Mojtaba Shamsipur,[*] Bahram Hemmateenejad,[†] and Morteza Akhond[‡]

Department of Chemistry, Razi University, Kermanshah, Iran
[†] *Medicinal & Natural Products Chemistry Research Center, Shiraz University of Medical Science, Shiraz, Iran*
[‡]*Department of Chemistry, Shiraz University, Shiraz, Iran*
Received May 20, 2003

Some new topological indices based on the distance matrix and Randic connectivity (as graph invariants) are proposed. The calculation of these indices is simple and they have good discriminating ability toward alkanes. Incorporating the number of carbon atoms to one of the calculated indices gives a highly correlating topological index (Sh index) which found to correlate with selected physicochemical properties of wide range of alkanes, specially, their boiling points. Most of the investigated properties are well modeled (with $r^2 > 0.99$) by the Sh index. Meanwhile, the resulting regressions were compared with the results based on the well-established Randic and newly reported Xu indices and, in most cases, better results were obtained by the Sh index. Moreover, multiple linear regression analysis of the alkane properties *via* calculated indices gives highly correlating models with low standard errors.

Key Words : Topological index Sh, QSPR, Alkanes

Introduction

A fundamental concept of chemistry is that the structural characteristics of a molecule are responsible for its properties.[1] Topological indices are a convenient means of translating chemical constitution into numerical values which can be used for correlation with physical properties in quantitative structure-property/activity relationship (QSPR/QSAR) studies.[2,3] The use of graph invariant (topological indices) in QSPR and QSAR studies has become of major interest in recent years.[4-6]

The need to represent molecular structure by a single number arises from the fact that most molecular properties are recorded as a single number. Characterization of a molecule by a single number represents a considerable loss of information: A three-dimensional object (molecule) is described by a one-dimensional object (topological index). However, what is surprising is how much of the relevant structural information is still retained in a given topological index. Ideally, a good topological index should show low degeneracy (isomer discrimination) and high correlation ability. There are more than 120 topological indices available in the literature[4] including the famous Randic index (χ),[7] Wiener number (W),[8] Hosoya index (Z)[9] Balaban index (J),[10] Schultz index (MTI),[11] Harary index (H)[12] and so on. However, among these indices the most successful was the Randic molecular connectivity χ which is used in much of QSAR/QSPR studies.[13] Despite the large achievements attained in this field, existing topological index approaches to QSAR/QSPR need further improvement. Very recently Xu and their coworkers[14,15] proposed a new topological index based on the combination of distance sum and vertex degree of atoms. This index has been applied to QSPR study of alkanes successfully.

In this study we proposed some new topological indices based on the combination of connectivity and distance sum using Randic[7] and Balaban[10] approach. Correlation of boiling point of alkanes with these indices was studied. Introduction of number of carbon atoms in one of our indices results in very highly correlating topological indices for studying some alkane properties including normal boiling point, heat capacity, critical temperature and pressure, constants of van der Waals equation (a and b) and so on. The results were compared with the other topological indices such as Randic and Xu indices.

Theoretical Section

Let G = {V, E} be a hydrogen depleted graph of a molecule with n atom, where V is the vertex set and E the edge set.[13] The distance matrix of a graph is $D = \{d_{ij}\}_{nxn}$ whose entries d_{ij} are the distance between vertices, so that in a constitutional graph all entries on the main diagonal are 0 and the other ones are non-zero integers, the shortest distance between vertices *i* and *j*. The distance sum (s) for each vertex results by adding entries in the row or column of the distance matrix corresponding to that vertex. The adjacency matrix is $A = \{a_{ij}\}_{nxn}$ whose entries are 1 for adjacent vertices *i*, *j* and zero otherwise. The topological indices derived from only A or D result in a considerable loss of structural information in a graph G and cannot be used for heteroatom-containing molecule successfully.

The connectivity term δ represented by Randic[7] for each vertex (atom) is calculated by the following equation:

$$\delta = N^v - H \quad (1)$$

where N^v is the number of valence electrons of considered atom and H is the number of hydrogen atom adjacent to that

atom. The Randic index calculated based on the connectivity term is the most successful index applied in most of the QSAR/QSPR studies, specially for heteroatom-containing molecules.[5,13,16,17]

The recently proposed Xu parameter[15] derived from a combination of distance and adjacency matrices has a better correlation ability in QSPR study of alkanes than those indices derived from D, A or connectivity. In this paper, we were interested to define some new topological indices by combination of connectivity and distance sum.

Thus, we defined different combinations of distance sum s and connectivity d using Randic[7] and Balaban[10] approaches:

$$Sh_1 = \log\left(\Sigma\frac{s_i s_j}{\delta_i \delta_j}\right) \quad (2)$$

$$Sh_2 = \log\left(\Sigma\frac{\delta_i \delta_j}{s_i s_j}\right) \quad (3)$$

$$Sh_3 = \log(\Sigma(s_i s_j \delta_i \delta_j)^{-0.5}) \quad (4)$$

$$Sh_4 = \log\left(\Sigma\left(\frac{\delta_i \delta_j}{s_i s_j}\right)^{-0.5}\right) \quad (5)$$

$$Sh_5 = \Sigma(s_i s_j + \delta_i \delta_j)^{-0.5} \quad (6)$$

$$Sh_6 = \log\Sigma((s_i s_j) + (\delta_i \delta_j)) \quad (7)$$

$$Sh_7 = \Sigma(\delta_i \delta_j + \log(s_i s_j)) \quad (8)$$

In above equations, the sums were given over all edges (bonds) and i and j represent the two adjacent atoms. Three other proposed indices were calculated based on individual methods: The distance sums were collected in a row vector (**S**) and connectivities in another row vector (δ). The logarithm of the inner product of **S** and transpose of δ (δ^T) gives the Sh_8 index:

$$Sh_8 = \log(\mathbf{S}\boldsymbol{\delta}^T) \quad (9)$$

The inner product of transpose of **S**($\mathbf{S}^T$) and δ gives a square matrix (**Sd**). The sum over all entries of **Sd** is Sh_9 index:

$$Sd = \mathbf{S}^T\boldsymbol{\delta} \quad (10)$$

$$Sh_9 = \log\left(\sum_i \sum_j Sd_{ij}\right) \quad (11)$$

The Sh_{10} index was calculated by the summation over eigen values of **Sd** matrix. The eigen values of **Sd** are calculated by singular-value decomposition (SVD).[18] In SVD an individual matrix is decomposed to three matrices: (*i.e.*, U, Σ and V):

$$Sd = U\Sigma V^T \quad (12)$$

Σ is a diagonal matrix whose elements are the eigen values of **Sd**, **U** and **V** are row and column designate of **Sd**. In our work, **Sd** has only one eigen value and other eigen values are zero. We represented the eigen value of **Sd** by ES, so that Sh_{10} is defied as:

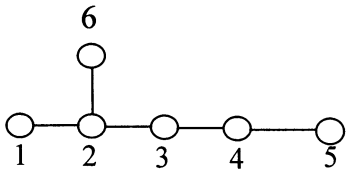

Figure 1. Labeled hydrogen depleted graph of 2-methylpentane.

$$Sh_{10} = \log(ES) \quad (13)$$

As an example, the labeled hydrogen depleted graph of 2-methylpentane is shown in Figure 1. The corresponding Sh indices of this molecule are then calculated as follows:

The distance matrix **D** and distance sum **S** vector are

$$D = \begin{vmatrix} 0 & 1 & 2 & 3 & 4 & 2 \\ 1 & 0 & 1 & 2 & 3 & 1 \\ 2 & 1 & 0 & 1 & 2 & 2 \\ 3 & 2 & 1 & 0 & 1 & 3 \\ 4 & 3 & 2 & 1 & 0 & 4 \\ 2 & 1 & 2 & 3 & 4 & 0 \end{vmatrix} \qquad S = \begin{vmatrix} 12 & 8 & 8 & 10 & 14 & 12 \end{vmatrix}$$

and the connectivity δ vector is: $\delta = \begin{vmatrix} 1 & 3 & 2 & 2 & 1 & 1 \end{vmatrix}$

$$Sh_1 = \log\left(\frac{12\times8}{1\times3} + \frac{8\times8}{3\times2} + \frac{8\times10}{2\times2} + \frac{10\times14}{2\times1} + \frac{8\times12}{1\times3}\right) = 2.22$$

$$Sh_2 = \log\left(\frac{1\times3}{12\times8} + \frac{3\times2}{8\times8} + \frac{2\times2}{8\times10} + \frac{2\times1}{10\times14} + \frac{1\times3}{8\times12}\right) = 0.22$$

$$Sh_3 = \log(12\times8\times1\times2 + 8\times8\times3\times2 + 8\times10\times2\times2 + 10\times14\times2\times1 + 8\times12\times1\times3)^{-0.5} = -0.55$$

$$Sh_4 = \log\left[\left(\frac{1\times3}{12\times8}\right)^{-0.5} + \left(\frac{3\times2}{8\times8}\right)^{-0.5} + \left(\frac{2\times2}{8\times10}\right)^{-0.5} + \left(\frac{2\times1}{10\times14}\right)^{-0.5} + \left(\frac{1\times3}{8\times12}\right)^{-0.5}\right] = 1.44$$

$$Sh_5 = (12\times8 + 1\times2)^{-0.5} + (8\times8 + 3\times2)^{-0.5} + (8\times10 + 2\times2)^{-0.5} + (10\times14 + 2\times1)^{-0.5} + (8\times12 + 1\times3)^{-0.5} = 0.51$$

$$Sh_6 = \log[(12\times8 + 1\times2) + (8\times8 + 3\times2) + (8\times10 + 2\times2) + (10\times14 + 2\times1) + (8\times12 + 1\times3)] = 2.68$$

$$Sh_7 = [\log(12\times8) + 1\times2] + [\log(8\times8) + 3\times2] + [\log(8\times10) + 2\times2] + [\log(10\times14) + 2\times1] + [\log(8\times12) + 1\times3] = 40.61$$

$$Sh_8 = \log(S.\delta^T) = 1.99$$

$$Sd = S^T.\delta = \begin{vmatrix} 12 & 36 & 24 & 24 & 12 & 12 \\ 8 & 24 & 16 & 16 & 8 & 8 \\ 8 & 24 & 16 & 16 & 8 & 8 \\ 10 & 30 & 20 & 20 & 10 & 10 \\ 14 & 42 & 28 & 28 & 14 & 14 \\ 12 & 36 & 24 & 24 & 12 & 12 \end{vmatrix}$$

$$Sh_9 = 6 \times 12 + 6 \times 24 + 2 \times 20 + 2 \times 36 + 6 \times 8 + 4 \times 16 + 2 \times 28 \;\; 3 \times 10 + 30 + 42) = 2.81$$

SS = SVD(**Sd**) = | 119.33 0 0 0 0 0 |

S = 119.33

Sh_{10} = log(119.33) = 2.08

Results and Discussion

Usually simplest QSPR studies start with studying the alkane properties. Boiling point (BP) of organic compounds is predetermined by the intermolecular interactions in the liquid state and by the difference in the molecular internal partition functions in the gas and liquid phases at boiling temperature. Therefore, it should be directly related to the chemical structure of the molecule and, thus, several methods have indeed been developed for estimation the normal BP of a compound from its structure.[17,19,20] Other physical properties such as critical temperature[21] and flash point[22] can be estimated from BP. Thus, we used normal BP of alkanes as starting point in our QSPR models. The calculated Sh indices of C_2-C_9 alkanes and their normal BPs are shown in Table 1.

In order to find the best Sh indices for modeling the alkane properties, we constructed a QSPR model for each index separately. Some indices show linear behavior and others have second ordered correlation. The resulting regression equations and the corresponding statistical parameters are represented in Table 2. In this Table, a is the intercept of the regression equation, b and c are the coefficient of first and second power of Sh indices, respectively, and r, s and f are correlation coefficient, standard error of regression and Fischer criterion of regression respectively. As shown, all

Table 1. Normal boiling points of C_2-C_9 alkanes, their calculated Sh indices and boiling points predicted by Sh index (a) and MLR (b)

No.	Alkanes	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10	Sh	Bp (°C)		
													Exp.	Pred.[a]	Pred.[b]
1	2	0	0	0	0	0.71	0.3	1	0.3	0.6	0.3	2	-88.63	-85.68	-89.70
2	3	0.78	-0.18	-0.24	0.54	0.71	1.2	7.58	1	1.51	1.06	4.35	-42.07	-44.91	-40.35
3	4	1.45	-0.38	-0.38	0.95	0.62	1.86	17.13	1.45	2.08	1.51	6.89	-0.5	-4.03	2.15
4	2M3	1.18	-0.22	-0.35	0.83	0.71	1.73	17.12	1.38	2.03	1.5	6.35	-11.73	-12.43	-12.34
5	5	1.96	-0.61	-0.49	1.26	0.53	2.37	27.97	1.78	2.51	1.83	9.38	36.07	32.94	36.710
6	2M4	1.77	-0.41	-0.45	1.17	0.61	2.25	28.77	1.72	2.46	1.82	8.95	27.85	26.77	28.32
7	22MM3	1.45	-0.24	-0.42	1.02	0.71	2.11	29.33	1.64	2.41	1.81	8.24	9.5	16.39	10.12
8	6	2.37	-0.81	-0.57	1.51	0.46	2.8	39.8	2.04	2.85	2.09	11.81	68.74	66.03	69.63
9	2M5	2.22	-0.66	-0.55	1.44	0.51	2.69	40.61	1.99	2.81	2.08	11.43	60.27	61.05	58.01
10	23MM4	2.03	-0.47	-0.51	1.35	0.58	2.58	42.27	1.93	2.76	2.05	10.98	57.99	55.06	55.52
11	33MM4	2	-0.41	-0.5	1.32	0.61	2.54	42.72	1.91	2.75	2.06	10.89	49.74	53.84	52.13
12	3M5	2.21	-0.58	-0.52	1.43	0.54	2.66	41.15	1.97	2.79	2.06	11.42	63.28	60.92	63.25
13	7	2.71	-0.99	-0.64	1.72	0.4	3.15	52.41	2.26	3.13	2.31	14.18	98.43	95.46	98.74
14	2M6	2.59	-0.87	-0.62	1.66	0.44	3.07	53.3	2.22	3.1	2.29	13.84	90.05	91.41	89.34
15	3M6	2.58	-0.79	-0.6	1.65	0.47	3.03	53.65	2.2	3.08	2.28	13.82	91.85	91.17	90.24
16	24MM5	2.43	-0.75	-0.6	1.59	0.48	2.98	54.08	2.18	3.06	2.28	13.44	80.5	86.57	79.55
17	3E5	2.56	-0.71	-0.57	1.63	0.49	2.98	54	2.18	3.06	2.26	13.76	93.48	90.45	94.06
18	22MM5	2.41	-0.65	-0.58	1.56	0.52	2.93	55.33	2.15	3.04	2.27	13.37	79.2	85.71	78.66
19	23MM5	2.43	-0.64	-0.57	1.58	0.51	2.93	55.36	2.15	3.04	2.26	13.44	89.78	86.57	87.09
20	33MM5	2.4	-0.55	-0.55	1.55	0.55	2.88	56.58	2.13	3.02	2.26	13.34	86.06	85.35	86.10
21	223MMM4	2.23	-0.48	-0.55	1.49	0.57	2.82	57.88	2.1	3	2.25	12.89	80.88	79.79	80.50
22	8	3.01	-1.15	-0.7	1.9	0.36	3.46	65.68	2.45	3.37	2.49	16.51	125.66	121.65	125.31
23	2M7	2.9	-1.06	-0.69	1.85	0.39	3.39	66.65	2.41	3.34	2.48	16.2	117.65	118.32	115.41
24	3M7	2.89	-0.98	-0.67	1.84	0.41	3.35	66.94	2.39	3.33	2.46	16.18	118.93	118.11	119.12
25	4M7	2.88	-0.96	-0.66	1.83	0.41	3.34	66.69	2.39	3.32	2.46	16.16	117.71	117.89	117.10
26	22MM6	2.75	-0.87	-0.66	1.77	0.44	3.28	68.75	2.36	3.3	2.46	15.78	106.84	113.74	107.24
27	25MM6	2.77	-0.97	-0.67	1.8	0.41	3.32	67.56	2.38	3.32	2.47	15.84	109.1	114.40	108.14
28	3E6	2.86	-0.91	-0.64	1.82	0.43	3.3	67.12	2.37	3.3	2.44	16.09	118.53	117.13	118.27
29	24MM6	2.77	-0.89	-0.65	1.78	0.44	3.27	67.78	2.36	3.3	2.45	15.83	109.43	114.29	111.68
30	23MM6	2.77	-0.84	-0.64	1.78	0.45	3.26	68.49	2.35	3.29	2.44	15.83	115.61	114.29	113.29

Table 1. Continued

No.	Alkanes	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10	Sh	Bp (°C)		
													Exp.	Pred.[a]	Pred.[b]
31	34MM6	2.76	-0.78	-0.62	1.77	0.46	3.23	68.96	2.33	3.28	2.43	15.81	117.73	114.07	114.78
32	33MM6	2.73	-0.75	-0.62	1.75	0.48	3.22	69.65	2.33	3.27	2.44	15.73	111.97	113.18	112.13
33	2M3E5	2.75	-0.79	-0.62	1.77	0.46	3.22	68.96	2.33	3.28	2.43	15.77	115.65	113.63	114.51
34	224MMM5	2.6	-0.77	-0.63	1.71	0.48	3.19	69.49	2.32	3.27	2.44	15.34	99.24	108.84	98.955
35	234MMM5	2.63	-0.72	-0.61	1.72	0.49	3.17	70.19	2.31	3.26	2.42	15.43	113.47	109.85	107.73
36	3E3M5	2.71	-0.66	-0.59	1.74	0.5	3.16	70.8	2.3	3.25	2.42	15.66	116	112.41	117.98
37	223MMM5	2.6	-0.65	-0.6	1.69	0.51	3.14	71.58	2.29	3.25	2.42	15.35	109.84	108.95	109.04
38	233MMM5	2.59	-0.62	-0.59	1.69	0.52	3.12	72.26	2.28	3.24	2.42	15.33	114.76	108.72	113.17
39	2233MMMM4	2.39	-0.51	-0.58	1.61	0.56	3.04	75.06	2.25	3.21	2.41	14.76	106	102.23	106.11
40	9	3.27	-1.3	-0.76	2.06	0.32	3.73	79.52	2.61	3.58	2.65	18.8	150.77	144.75	151.24
41	4E7	3.12	-1.03	-0.68	1.98	0.39	3.56	80.32	2.53	3.51	2.6	18.35	141.2	140.42	141.18
42	3E7	3.13	-1.06	-0.69	1.98	0.38	3.58	80.72	2.54	3.52	2.6	18.4	143	140.90	141.57
43	4M8	3.15	-1.12	-0.72	2	0.37	3.62	80.46	2.56	3.54	2.62	18.46	142.4	141.49	141.29
44	3M8	3.17	-1.16	-0.72	2.01	0.36	3.64	80.85	2.57	3.55	2.63	18.5	143.3	141.87	144.40
45	2M8	3.17	-1.22	-0.74	2.02	0.34	3.67	80.5	2.58	3.56	2.64	18.51	142.8	141.97	142.93
46	26MM7	3.06	-1.15	-0.73	1.97	0.36	3.61	81.57	2.56	3.54	2.63	18.19	135.2	138.85	138.58
47	25MM7	3.06	-1.08	-0.71	1.96	0.38	3.57	81.81	2.54	3.52	2.62	18.18	136	138.75	135.25
48	35MM7	3.05	-1.01	-0.69	1.95	0.4	3.53	82	2.52	3.51	2.6	18.15	136	138.46	136.10
49	4E2M6	3.03	-0.98	-0.68	1.94	0.41	3.51	81.59	2.51	3.5	2.59	18.08	133.8	137.77	135.32
50	24MM7	3.05	-1.04	-0.7	1.95	0.39	3.56	81.38	2.53	3.51	2.61	18.15	133.5	138.46	135.81
51	3E4M6	3.02	-0.89	-0.65	1.92	0.43	3.47	82.71	2.48	3.48	2.57	18.05	140.4	137.47	138.84
52	34MM7	3.04	-0.96	-0.68	1.94	0.41	3.51	82.54	2.51	3.5	2.59	18.13	140.1	138.26	139.21
53	3E2M6	3.02	-0.93	-0.67	1.93	0.42	3.49	82.14	2.49	3.49	2.58	18.05	138	137.47	139.04
54	23MM7	3.06	-1.03	-0.7	1.95	0.39	3.56	82.36	2.53	3.53	2.61	18.17	140.5	138.66	138.88
55	235MMM6	2.93	-0.95	-0.68	1.9	0.42	3.49	83.22	2.49	3.49	2.59	17.8	131.3	134.99	132.89
56	234MMM6	2.93	-0.86	-0.66	1.89	0.44	3.44	84.32	2.47	3.47	2.57	17.79	139	134.89	136.35
57	3E3M6	2.99	-0.83	-0.65	1.9	0.45	3.45	84.22	2.47	3.47	2.58	17.97	140.6	136.68	138.29
58	33MM7	3.03	-0.95	-0.68	1.93	0.41	3.52	83.53	2.51	3.5	2.6	18.08	137.3	137.77	138.73
59	22MM7	3.05	-1.06	-0.72	1.95	0.38	3.58	82.77	2.54	3.52	2.63	18.14	132.7	138.36	135.37
60	225MMM6	2.92	-0.99	-0.7	1.9	0.41	3.51	83.66	2.51	3.5	2.61	17.75	124	134.48	129.08
61	244MMM6	2.92	-0.87	-0.65	1.88	0.45	3.43	83.13	2.47	3.47	2.59	17.77	126.5	134.68	129.10
62	224MM6	2.91	-0.92	-0.68	1.88	0.43	3.46	83.79	2.48	3.48	2.6	17.74	126.5	134.38	129.29
63	2244MMMM5	2.75	-0.83	-0.67	1.82	0.46	3.4	85.62	2.45	3.46	2.6	17.24	122.7	129.30	118.70
64	3E23MM5	2.88	-0.72	-0.62	1.85	0.48	3.37	86.88	2.43	3.44	2.56	17.63	141.6	133.28	141.23
65	2334MMMM5	2.95	-0.7	-0.6	1.87	0.49	3.36	86.68	2.42	3.43	2.56	17.85	141.5	135.49	141.43
66	233MMM6	2.9	-0.8	-0.65	1.87	0.46	3.42	85.81	2.46	3.46	2.58	17.7	137.7	133.98	135.67
67	334MMM6	2.9	-0.76	-0.64	1.86	0.47	3.39	86.36	2.45	3.45	2.57	17.69	140.5	133.88	138.53
68	3E22MM5	2.88	-0.78	-0.64	1.86	0.47	3.39	85.38	2.45	3.45	2.57	17.65	133.8	133.48	135.26
69	223MMM6	2.91	-0.85	-0.66	1.88	0.44	3.44	85.28	2.47	3.47	2.59	17.73	131.7	134.28	135.75
70	2234MMMM5	2.73	-0.73	-0.62	1.8	0.5	3.33	86.4	2.42	3.41	2.54	17.19	133	128.78	131.12
71	2233MMMM5	2.74	-0.64	-0.62	1.79	0.51	3.31	89.92	2.41	3.42	2.56	17.22	140.27	129.09	137.97
72	3E24MM5	2.83	-0.85	-0.71	1.81	0.4	3.35	74.96	2.46	3.46	2.56	17.49	136.73	133.85	137.10
73	44MM7	3.02	-0.92	-0.68	1.92	0.42	3.5	83.17	2.5	3.49	2.6	18.05	135.20	137.48	134.95

M: methyl, E: ethyl. Example: 3E22MM5 = 3-ethyl-2,2-dimethylpentane.

indices, except Sh_2, Sh_3 and Sh_5, have $r^2 > 0.95$. As the BP of alkanes controlled by molecular weight other than the topology of molecule, we incorporated the number of carbon atoms (n) in some of the Sh indices and found that the following index is highly correlated with the BP of alkanes:

$$Sh = n + n^{1/2}Sh_1 \quad (14)$$

The relationship between the BP of C_2-C_9 alkanes (73 compounds) and proposed Sh index, along with several selected topological indices including Wiener (W),[8] Hosoya (Z).[9] Harary (H),[12] Schultz (MT1),[11] Balaban (I),[10] Randic (χ)[7] and Xu[15] indices are as follows:

$$\begin{aligned} &BP = -125.8 + 130.6\mathrm{Log}(W) \\ &\quad r^2 = 0.9635 \quad s = 8.896 \quad f = 1899 \\ &BP = -97.0 + 147.0\mathrm{Log}(Z) \end{aligned} \quad (15)$$

Table 2. Regression results of alkanes boiling points using Sh indices

Sh index	a	b	c	r^2	se	F
Sh_1	-105.28	79.76	–	0.966	8.717	1930.33
Sh_2	157	-272	–	0.819	20.13	306.78
Sh_3	-123.53	-373.46	–	0.889	15.73	545.45
Sh_4	-114.67	129.27	–	0.970	8.21	2184.8
Sh_5	26.83	717.34	-1136.1	0.804	21.05	137.86
Sh_6	-139.53	78.02	–	0.966	8.66	1957.14
Sh_7	-77.17	4.043	-.018	0.966	8.76	956.58
Sh_8	-164.57	118.90	–	0.960	9.5	1615.65
Sh_9	-183.64	90.62	–	0.962	9.16	1741.51
Sh_{10}	-169.85	116.35	–	0.950	10.61	1278.87

Table 3. Regression results of some selected alkane properties using Sh index

Property	TI	a	b	c	r^2	s	F	N
Cp	Sh	15.81	2.51	0.013	0.9914	0.6491	862.9	18
	Xu	40.99	41.76	–	0.9926	7.18	2150	18
Tc	Sh	-3.77	24.65	-0.39	0.9914	6.093	922.9	22
	Xu	57.82	65.52	–	0.9710	13.29	672	22
Pc	Sh	56.26	-3.37	0.09	0.953	1.56	183.99	18
	Xu	46.61	-6.36	–	0.9024	2.18	165.2	18
a_0	Sh	2.36	1.37	0.04	0.9986	0.434	2927.3	12
	Xu	2.038	9.28	–	0.9866	1.56	739	12
b_0	Sh	0.041	0.01	7.0×10^{-5}	0.9983	0.0027	2116	12
	Xu	0.054	0.048	–	0.9954	0.0048	2138	12

$r^2 = 0.9658$ s = 8.605 f = 2034 (16)

$BP = -133.3 + 245.7Log(H)$

$r^2 = 0.9342$ s = 11.937 f = 1022 (17)

$BP = -207.4 + 134.0Log(MTI)$

$r^2 = 0.9613$ s = 9.163 f = 1786 (18)

$BP = -297.0 + 215.1J - 26.965J^2$

$r^2 = 0.6372$ s = 28.236 f = 62 (19)

$BP = -194.5 + 119.51c - 9.51c^2$

$r^2 = 0.9891$ s = 4.891 f = 3220 (20)

$BP = -80.79 + 56.6Xu$

$r^2 = 0.9857$ s =5.791 f = 2615 (21)

$BP = -122.53 + 18.92Sh - 0.25Sh^2$

$r^2 = 0.9913$ s = 4.410 f = 3872 (22)

In general, the quality of QSPR models can be conveniently measured by the correlation coefficient r and standard error of regression s. For the proposed Sh index, r^2 and s are 0.9913 and 4.41, respectively. The root mean square error (RMS), absolute average error between the experimental values of BP and predicted values and coefficient of variation (CV) are 4.35, 3.49 and 4.16%, respectively, whereas other indices have $r^2 < 0.99$ and s > 4.5. Therefore, it is apparent that eq (22) (Sh index) represents a better model than eqs. (15-21) for modeling BP of alkanes. The plot of BP of alkanes versus the index Sh for C_2-C_9 alkanes is shown in Figure 2.

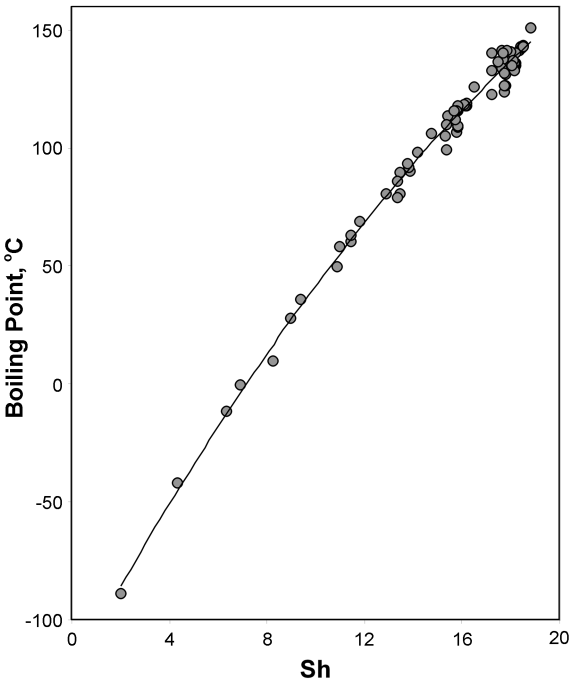

Figure 2. Plot of boiling points of C_2-C_9 alkanes versus Sh index.

In order to obtain extensive structure-property relationships, Sh index was further tested by modeling some other properties of alkanes including heat capacity C_p ($J \cdot mol^{-1} \cdot K^{-1}$) of gases at 25 °C, critical temperature T_c and pressure P_c and van der Waals constants, a_0 ($L \cdot atm \cdot mol^{-2}$) and b_0 ($L \cdot mol^{-1}$). The regression results are shown in Table 3. From this table, it is seen that all properties are fairly correlated with Sh index with ($r^2 > 0.99$), except with P_c that has a $r^2 > 0.95$. Meanwhile Sh index produces better regression results compared with the newly reported Xu index for all properties listed in Table 3, except for Cp. Since alkanes are nonpolar compounds, their physicochemical properties are inherently determined by their structures (especially molecular size), and other factors due to polarity, polarizability and hydrogen bonding have no significant effect on alkane properties. The good correlation between Sh index and alkane properties indicates that it is a suitable topological index for QSPR studies of bond additive properties.

In order to do the modeling of BPs of alkanes more accurately, we used all proposed Sh indices. Multiple linear regression (MLR) was used for selecting the best equation of fit. The result of MLR is shown in Table 4. Further improve-

Table 4. Result of MLR analysis of alkanes boiling points with Sh indices

Variable	Coefficient	Se of coefficient	T of H_0	P
Constant	-133.95	15.60	-8.59	0.000
Sh_1	31.16	10.02	3.11	0.003
Sh_2	75.28	13.44	5.60	0.000
Sh_3	175.44	26.79	6.55	0.000
Sh_7	2.98	0.32	9.30	0.000
Sh_8	422.09	62.0	6.81	0.000
Sh_{10}	-234.35	37.85	-6.19	0.000
n	-45.27	5.97	-7.58	0.000

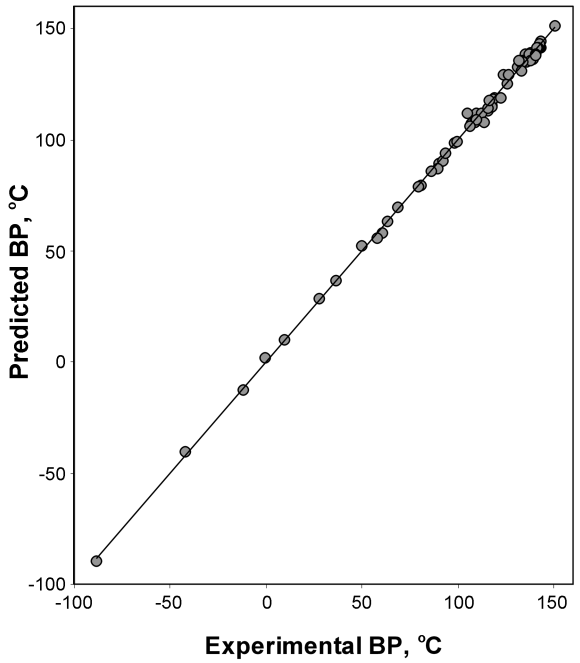

Figure 3. Plot of predicted boiling point of alkanes C_2-C_9 (using MLR) versus experimental boiling points.

ment of BP modeling of alkanes is obtained by using more than one Sh index. The standard error of this model is 2.27 °C. Absolute average error, root mean square error and relative standard deviation of prediction are 1.58, 2.16 and 2.05%, respectively. A plot of the predicted values of BP of alkanes by MLR versus the experimental values is shown in Figure 3. The high correlation coefficient of this plot (r = 0.999) confirm the excellent predictive ability of our proposed model.

Modeling through isomer series (i.e. octane isomers) is another approach to QSPR studies. Isomer discrimination is one of the fundamental properties that a topological index is expected to have. Ren[15] has done a study on selected properties of octane isomers by Randic (χ) and Xu indices. The studied properties are entropy (S), quadratic mean radius (R^2), density (D), heat of vaporization (H_v), heat of formation (H_f), Pitzer eccentric factor (ω) and octane numbers (MON). He has considered a simple linear model between octane properties and above mentioned indices, although, in some cases the quadratic relationship was found to be a better model. Thus, in this work, we repeated the modeling of octane properties by Randic and Xu indices using both the first and second ordered linear models. Moreover, the first and second ordered relationships between octane properties and Sh index were obtained. The results are shown in Table 5. From this Table it is obvious that, in some cases, the relationship between octane properties and topological indices (χ, Xu and Sh) is better modeled by a quadratic fitting. Second order relationships are necessary for modeling S, R^2, D, ω, V_c and MON by Xu index and S, D and H_v by Randic index. Meanwhile, Sh index modeled all considered octane properties by second order relationships except for H_f and ω that show a first order relationship. Sh index results in a better correlation for S ($r^2 = 0.901$), R^2 ($r^2 = 0.880$) and D ($r^2 = 0.750$) than Xu index (with respective r^2s of 0.871, 0.853 and 0.558) and Randic index (with respective r^2s of 0.825, 0.699 and 0.627). Also, Sh index has a better correlation with H_v ($r^2 = 0.779$) and H_f ($r^2 = 0.493$) than Xu index ($r^2 = 0.558$ and 0.271, respectively), and with ω ($r^2 = 0.956$), V_c ($r^2 = 0.380$) and MON ($r^2 = 0.940$) than Randic index ($r^2 = 0.819$, 0.171 and 0.630, respectively). The correlation coefficient of Sh index with S, R^2, ω and MON is > 0.9, with D and H_v is > 0.8 and there is an unsatisfactory regression between Sh and H_f and V_c.

After a series of comprehensive studies on a numerous properties of octane isomers by using of one and/or two

Table 5. First and second order correlation between octane properties and χ, Xu and Sh indices

property	Order	X			Xu			Sh		
		r^2	se	f	r^2	se	f	r^2	se	f
S	first	0.82	1.97	72.94	0.800	2.08	64.1	0.890	1.54	129.3
	2nd	0.825	2.00	35.41	0.871	1.72	50.5	0.901	1.51	68.3
R2	first	0.593	0.119	23.3	0.804	0.082	65.7	0.746	0.094	47.0
	2nd	0.699	.106	17.38	0.853	.074	43.4	0.880	0.067	55.2
D	first	0.312	0.025	7.26	0.419	0.023	11.54	0.441	0.023	12.61
	2nd	0.627	0.019	12.60	0.749	0.016	22.3	0.750	0.016	22.55
Hv	first	0.876	0.735	113.3	0.557	1.39	20.1	0.745	1.054	46.83
	2nd	0.918	0.618	83.88	0.558	1.43	9.46	0.882	1.104	26.46
Hf	first	0.723	0.679	41.8	0.270	1.10	5.91	0.489	0.922	15.33
	2nd	0.740	0.680	21.3	0.271	1.13	2.78	0.493	0.948	7.30
W	first	0.818	0.016	72.05	0.948	0.009	290.1	0.955	0.008	340.3
	2nd	0.819	0.017	33.95	0.965	0.007	204.9	0.956	0.008	163.9
Vc	first	0.113	0.015	2.04	0.472	0.012	14.3	0.261	0.014	5.67
	2nd	0.171	0.015	1.54	0.546	0.011	9.02	0.380	0.013	4.61
MON	first	0.604	16.14	21.34	0.925	7.01	173.2	0.865	9.42	89.0
	2nd	0.630	16.19	11.07	0.940	6.52	101.7	0.940	6.67	82.45

Table 6. Result of MLR of octane properties and Sh indices

Property	Predictor indices	r^2	S	F
S	Sh_1, Sh^2, Sh_3^2, Sh_4, Sh_5^2, Sh_6, Sh_7	0.973	0.98	50.04
R^2	Sh_1^2, Sh_2, Sh_2^2, Sh_3^2, Sh_5^2, Sh_6, Sh_7	0.982	0.036	52.62
D	Sh_1, Sh_2, Sh_2^2, Sh_5^2, Sh_6, Sh_7, Sh_9^2, Sh_{10}	0.963	0.0074	29.51
Hv	Sh_1, Sh_1^2, Sh_2, Sh_3^2, Sh_4, Sh_6, Sh_7, Sh_8^2	0.994	0.256	105.5
Hf	Sh_1, Sh_1^2, Sh_2, Sh_2^2, Sh_3^2, Sh_4, Sh_7, Sh_8^2, Sh_9^2	0.943	0.466	11.50
w	Sh_1, Sh_1^2, Sh_2^2, Sh_3^2, Sh_8^2, Sh_9^2, Sh_{10}	0.998	0.002	562.8
Vc	Sh_2, Sh_3^2, Sh_5^2, Sh_6, Sh_8^2, Sh_{10}	0.783	0.009	6.59
MON	Sh_2, Sh_2^2, Sh_5^2, Sh_6, Sh_7, Sh_{10}	0.994	237	272.3

topological indices, Randic *et al.*[23,24] suggested that there are no simple and accurate QSPR models for the octane isomers by the use of their topological indices. Later on, Ren[15] proposed that the small difference between octane properties is responsible for the diminished regression of octane properties with a single topological index. In order to obtain more accurate regression, we used our proposed Sh indices to modeling the octane properties by MLR (using more than two indices). In our modeling, we consider all Sh indices defined (*i.e*. Sh1-Sh10) as well as their second power. The most significant descriptors for each property were then obtained by MLR. The resulting regression models between octane properties as dependent variable, and Sh indices, as independent variable, are shown in Table 6. R^2, Hv, ω and MON are modeled with high correlation coefficients ($r > 0.99$). All other considered properties revealed correlation coefficient greater than 0.9, except Vc, which has a r^2 value $r = 0.885$. This might be due to the uncertainty in measuring the critical volume. The high correlation coefficient of modeling of octane properties by Sh indices emphasizes that a combination of these indices can extract high structural information from a molecule. In the other words, since the octane isomers have the same number of carbon atoms, their properties is determined by compactness of molecules. This means that the Sh indices are actually representative of the molecular compactness other than the molecular size.

Conclusions

A combination of distance sum and connectivity has been used to generate some new topological indices. The incorporation of number of carbon atoms in the proposed indices generates a Sh index that has high correlation with some physicochemical properties of alkanes. Meanwhile, these properties can be better modeled using a combination of the Sh indices. A multiple linear model was used for such modeling. The Sh index has a better discriminatory power of alkane isomers than the Balaban's J index. Its calculation is very simple and its correlation with BP of alkanes is better than other indices such as Randic and Xu indices. In addition, The Sh index can be extended for heteroatom-containing compounds and/or multiple bonds. The discriminating ability of Sh index among heteroatom-containing molecules (such as alcohol, amines, ethers, thiols and so on) is much better than the J index. Meanwhile, Sh index shows a good correlation with properties of these compounds. Thus, we proposed that it could be a useful parameter for QSAR/QSPR studies. Investigations on the relationships between the properties of molecules other than alkanes and the Sh indices are continued in this group and we will report on this topic in the later papers in detail.

References

1. Turro, N. J. *Angew. Chem. Int. Edit. Engl.* **1986**, *25*, 882-901.
2. Gilabert, L. E.; Sagrado, S.; Villanueva-Camanas, R. M.; Medina-Hernandes, M. J. *Anal. Chem*. **1998**, *70*, 28-34.
3. Li, H.; Xu, L.; Guo, M.; Su, Q. *Talanta* **1997**, *44*, 203-211.
4. Viswanadhan, V. N.; Mueller, G. A.; Basak, S. C.; Weinstein, J. N. *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 505-511.
5. Agatonovic-Kustrin, S.; Tacker, L. G.; Zecevic, M.; Zivanovic, L. J. *Anal. Chim. Acta* **2000**, *418*, 181-195.
6. Shamsipur, M.; Hemmateenejad, B.; Akhond, M.; Sharghi, H. *Talanta* **2001**, *54*, 1113-1120.
7. Mihalic, Z.; Trinajistic, N. *J. Chem. Educ.* **1992**, *69*, 701-712.
8. Balaban, A. T. *J. Chem. Inf. Comput. Sci*. **1985**, *25*, 334-343.
9. Balaban, A. T. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 339-350.
10. Randic, M. *J. Am. Chem. Soc*. **1975**, *97*, 6609-6615.
11. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
12. Hosoya, H. *Bull. Chem. Soc. Jpn*. **1971**, *44*, 2332-2339.
13. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
14. Schultz, H. P. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 227-228.
15. Plavsic, D.; Nikolic, S.; Trinajistic, N. *J. Math. Chem*. **1993**, *12*, 235-250.
16. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
17. Yao, Y. Y.; Xu, L.; Yuan, X. S. *Acta Chim. Sinica (in Chinese)* **1996**, *54*, 533-538.
18. Ren, B. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 139-143.
19. Zhao, Y. H.; Yang, L. H.; Wang, L. S. *Bull. Environ. Contam. Toxicol*. **1996**, *57*, 242-249.
20. Smeeks, F. C.; Jurs, P. C. *Anal. Chim. Acta* **1990**, *233*, 111-119.
21. Malinowsky, E. R. *Factor Analysis in Chemistry*; John Wiley: New York, 1991.
22. Rechsteiner, C. E. *Handbook of Chemical Property Estimation Methods*; Lyman, W. L.; Reehl, W. E.; Rosenblatt, D. H., Eds; McGraw Hill: New York, 1982.
23. Kartitzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. A. *J. Phys. Chem*. **1996**, *100*, 10400-10407.
24. Fisher, C. H. *Chem. Eng*. **1989**, *96*, 157-158.
25. Satyanarayana, K.; Kakati, M. C. *Fire Mater.* **1991**, *15*, 97-100.
26. Randic, M.; Guo, X. F.; Oxley, T.; Krishnapriyan, H. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 709-716.
27. Randic, M.; Guo, X. F.; Oxley, T. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 361-367.