Topological Organic Chemistry. 12. Whole-Molecule Schultz Topological Indices of Alkanes

Harry P. Schultz[*,†]

Department of Chemistry, University of Miami, Coral Gables, Florida 33124

Tor P. Schultz

Forest Products Laboratory, Mississippi State University, Mississippi State, Mississippi 39762

A whole-molecule modified Schultz index was devised from the application of linear filtration to the constant interval reciprocal distance matrix. The resultant descriptors of alkanes were unique within the environment of the 39 C_2−C_8 alkane studies. The utility of the descriptors was briefly examined with quantitative structural−physical relationship studies.

INTRODUCTION

The first paper[1] of this series introduced the Schultz molecular topological index (MTI, or *S*, as it will be referred to hereafter), where **v** is the valence row matrix, **D** is the distance matrix wherein $D_{ij} = l_{ij}$ if $i = j$, 0 otherwise; l_{ij} is the edge count of the shortest path between vertices i and j in graph *G*; **A** is the adjacency matrix where $A_{ij} = 1$ if vertices i and j are adjacent, 0 otherwise.

$$S = \sum_{i=1}^{N} [\mathbf{v}(\mathbf{A} + \mathbf{D})] \quad (1)$$

Subsequently an abbreviated *S*′ index that omitted the adjacency matrix from eq 1 was described by Mihalić et al.[2] Indices derived from application of the above equations were

$$S' = \sum_{i=1}^{N} [\mathbf{vD}] \quad (2)$$

found to be monotonic, single sum, and invariant. In a penetrating and comparative review Mihalić and Trinajstić[3] reported an r^2 value of 0.9964 from regression data obtained from *S* indices versus boiling points for the alkane series ethane through the octanes. They also cautioned that any regression data should always possess values above $r = 0.99$ ($r^2 = 0.98$) in order to be useful.

However, the *S* and *S*′ indices defined above suffered from a disadvantage, for the indices decreased in magnitude as the complexities of the alkane graphs increased. Bertz[4,5] fostered the desirability of increased topological index values for molecules as their complexities—chain lengths, branchings, cyclizations, etc.—increased.

Plavšić et al.[6] and Ivanciuc et al.[7] formulated an index (*H*, *RW*) as a reciprocal analogue of the Weiner index (*W*), naming the new index in honor of Frank Harary; Diudea[8]

extended the study. Equation 3 summarizes their work, where

$$RW = H = (1/2)\sum_{i}\sum_{j}[1/D]_{ij} = (1/2)\sum_{i}\sum_{j}[RD]_{ij} \quad (3)$$

RD is the reciprocal property matrix of the **D** matrix, each element being the reciprocal of an edge count of the original **D** matrix. However, unlike with the **D** matrix integers, there are not constant reciprocal edge count intervals between vertices of **RD** matrices; the reciprocals become increasingly smaller in value as the corresponding integers increase in size.

Guided by Elk[9] who wrote, "...let equal fractional parts (of a math model) be uniformly spaced," a preceding paper[10] of this series advanced the concept of the reciprocal edge count further to the idea of a constant interval reciprocal (CIR) edge count, where N_{max} could be designated as being

$$CIR = [N_{max} - N + 1]/N_{max} \quad (4)$$

within vector, matrix, or graph, and *N* was any **D** matrix integer; $N \neq 0$. The derived CIR data were inserted within matrices to form CIR(**D**) matrices that were then used to complete new models of *S* and *S*′ CIR indices by use of eqs 1 and 2. The consequent data yielded molecular topological indices that increased in value as the complexities of the molecules increased. Regression analyses[10] relating boiling points of alkanes with these indices (negative association) gave r^2 values of about 0.94−0.95, and heats of formation versus indices (positive association) gave useful r^2 values of about 0.98−0.99.

Unfortunately, in these recent studies, indeed in all past studies of this series, there were encountered instances of index degeneration—identical index values for different structures. All the constituent units in all alkanes are much the same, comprised of individual vertices of the main chain and methyl and ethyl (most commonly) substituents. A molecular topological index is merely the exact aggregate, however computed, of the targeted contributions of the structural units present. A crude parallelism may be likened
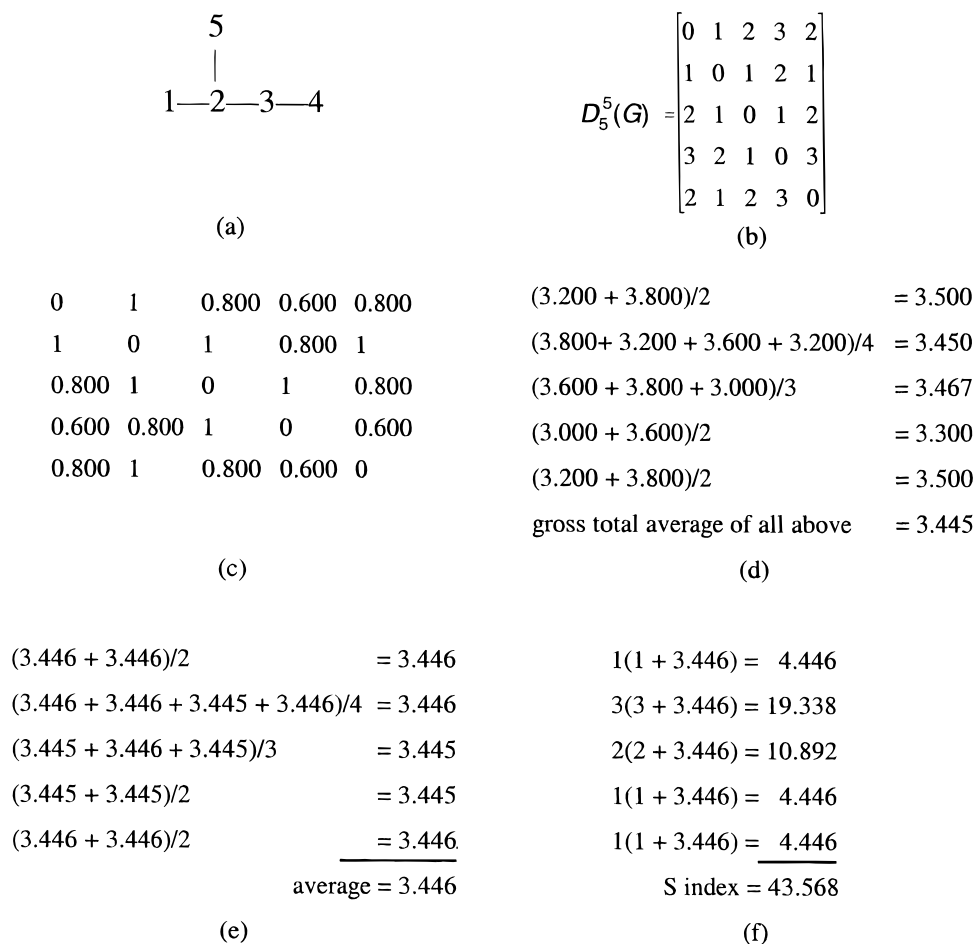
† Present address: P.O. Box 262, Big Horn, WY 82833.

$$
\begin{array}{c}
5 \\
| \\
1\!-\!2\!-\!3\!-\!4
\end{array}
$$

(a)

$$
D_5^5(G) =
\begin{bmatrix}
0 & 1 & 2 & 3 & 2 \\
1 & 0 & 1 & 2 & 1 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 3 \\
2 & 1 & 2 & 3 & 0
\end{bmatrix}
$$

(b)

0	1	0.800	0.600	0.800
1	0	1	0.800	1
0.800	1	0	1	0.800
0.600	0.800	1	0	0.600
0.800	1	0.800	0.600	0

(c)

(3.200 + 3.800)/2	= 3.500
(3.800 + 3.200 + 3.600 + 3.200)/4	= 3.450
(3.600 + 3.800 + 3.000)/3	= 3.467
(3.000 + 3.600)/2	= 3.300
(3.200 + 3.800)/2	= 3.500
gross total average of all above	= 3.445

(d)

(3.446 + 3.446)/2	= 3.446
(3.446 + 3.446 + 3.445 + 3.446)/4	= 3.446
(3.445 + 3.446 + 3.445)/3	= 3.445
(3.445 + 3.445)/2	= 3.445
(3.446 + 3.446)/2	= 3.446
	average = 3.446

(e)

1(1 + 3.446) =	4.446
3(3 + 3.446) =	19.338
2(2 + 3.446) =	10.892
1(1 + 3.446) =	4.446
1(1 + 3.446) =	4.446
	S index = 43.568

(f)

Figure 1. (a) 2-Methylbutane graph; (b) **D** matrix; (c) CIR(**D**) matrix, N_{max} graph (5); (d) first complete iteration unit; (e) last, 16th, complete iteration unit; (f) *S* index calculation.

to wrongly arriving at a property of the diatomic hydrogen molecule by reckoning and doubling the corresponding property of two separate hydrogen atoms.

In contrast and rectification of the above circumstance, a technique was sought and devised that assessed and included relative relationships among the structural units of an alkane as a totality of its own complete structure. This method consisted of constructing a whole-molecule CIR(**D**) matrix, which was then used to calculate the *S* topological index according to eq 1. The operation recognized that the whole molecule was different from the simple totality of all its parts. After much experimentation the operation chosen to accomplish this end was that of linear filtration of the vertex vectors of the CIR(**D**) matrix. Interestingly, this is the same general technique used to clarify and enhance digitally transmitted pictures, such as of astronomical objects. More distantly related instances of linear filtration or smoothing, for example, are described and reviewed in recent work by Grznar et al.[11,12]

COMPUTATIONS

Molecular graphs, hydrogen-suppressed and with the interatomic carbon−carbon edge count set at unity, were derived from the alkane series, ethane through the octane isomers, listed in Table 2. The IUPAC system of numbering was used, as displayed in Figure 1a, which pictures a modified stem-and-leaf graph of 2-methylbutane.

The **D** matrix of each graph was constructed in traditional fashion, as illustrated in Figure 1b. Each of the **D** matrix elements was then transformed into its corresponding constant interval reciprocal (CIR) number, graph based, to form the CIR(**D**) matrix as exhibited in Figure 1c. The results reported in the tables of this paper were all rounded and recorded to three decimal places. This operation was effected by the use of eq 4, with N_{max} equal to the graph vertices−5 in the instance of the CIR(**D**) matrix pictured in Figure 1c. Each vertex vector was then summed to yield its CIR total. The complete CIR data were then subjected to total linear filtration (TLF), as illustrated in Figure 1d.

Tucker and Small[13] have recently described linear filtration (LF) for simple curves, wherein the *i*th value of d_i was replaced by the average of the sum of the values of d_i and the two flanking values of d_{i-1} and d_{i+1} yielding a new value d_i' defined by eq 5. In the instances of alkane graphs each

$$d_i' = (d_{i-1} + d_i + d_{i+1})/3 \tag{5}$$

vertex may have from one through four adjacent values, depending upon the valence of vertex d_i. Equation 5 could, therefore, have as few as two and as many as five figures in a given vertex LF operation. The LF operation for 2-methylbutane is pictured in Figure 1d. The newly created set of LF data was iterated until constancy of repeated sets of iteration numbers was attained as displayed in Figure 1e. The final constant set of iterated data, consistently carried

Table 1. Descriptors Derived from Constant Interval Reciprocal (CIR)−Total Linear Filtration (TLF) Elements of **D** matrices for 4-Carbon Alkanes and Cycloalkanes

no.	compound	total iterations	TLFCIR(**D**) vertex values	descriptors S′	S
1	butane	4	2.550	15.300	25.300
2	methylpropane	7	2.701	16.206	28.206
3	cyclobutane	0	2.750	22.000	38.000
4	methylcyclopropane	11	2.793	22.344	40.344
5	bicyclobutane	5	2.894	28.940	54.940
6	tetrahedrane	0	3.000	36.000	72.000

along to three decimal places, is referred to in this paper as total linear filtration (TLF).

The presence of the digit 5 in the fourth decimal place was resolved with modification of a technique used by professional surveyors. If the progression of the iteration series was in an increasing mode, the 5 was dropped and the digit in the third decimal place was increased by 1; in a decreasing trend the 5 was dropped and the digit in the third decimal place was retained unchanged. Of the 39 compounds

reported in this study, the iterations needed to attain constancy ranged from 0 (cyclobutane and tetrahedrane, Table 1, compounds 3 and 6) to 50 (2-methylheptane and 2,2-dimethylhexane, Table 2, compounds 23 and 32). Generally, when the iteration cycles reached a steady state, all the separate values for the vertices of a compound were the same, except for a few of the more unsymmetric alkanes, especially the methyl-substituted alkanes, which had slight differences among the TLFCIR vertex values when the iterations themselves became constant. For example, such a gap was observed for 3-methylhexane (Table 2, compound 15), which displayed at iteration cycle constancy of 33 cycles a final TLFCIR range of vertex values of 4.915−4.918, 0.003 unit difference (0.061%). The average of the values, 4.917 in the above instance, was tabulated in Table 2, column 6.

A constant third place value for the TLFCIR vertex value was arrived at in yet another manner. If, instead of averaging the slightly different vertex values, one continues the iteration process and carries the arithmetic operation out to a larger number of decimal places, a point is reached wherein the

Table 2. Descriptors Derived from Constant Interval Reciprocal (CIR)−Total Linear Filtration (TLF) Elements of **D** Matrices for the 39-Alkane Sequence Ethane Through the Octanes

| | | | | | TLFCIR(**D**) vertex values | | | descriptors, based on | | | | | |
| | | | | | | | | total iter'ns | | first unit average | | bulk | |
no.	graph[a]	bp,[b] °C	(−1)H_f(g),[c] kJ/mol	total iter'ns[d]	at total iter'ns	first unit average	bulk C-1	S′	S	S′	S	S′	S
1	C2	−88.6	83.45	1	1.000	1.000	1.000	2.000	4.000	2.000	4.000	2.000	4.000
2	C3	−42.1	104.67	4	1.810	1.811	1.778	7.240	13.240	7.244	13.244	7.112	13.112
3	C4	−0.5	125.66	5	2.550	2.548	2.500	15.300	25.300	15.288	25.288	15.000	25.000
4	2MC3	−11.7	134.19	5	2.701	2.705	2.643	16.206	28.206	16.230	28.230	15.858	27.858
5	C5	36.0	146.77	9	3.262	3.256	3.200	26.096	40.096	26.048	40.048	25.600	39.600
6	2MC4	27.8	153.68	16	3.446	3.445	3.400	27.568	43.568	27.560	43.560	27.200	43.200
7	22MMC3	9.4	167.95	8	3.632	3.640	3.550	29.056	49.056	29.120	49.120	28.400	48.400
8	C6	68.7	167.03	13	3.958	3.949	3.889	39.580	57.580	39.490	57.490	38.890	56.890
9	2MC5	60.2	174.8	27	4.143	4.141	4.110	41.430	61.430	41.410	61.410	41.100	61.100
10	3MC5	63.2	172.1	14	4.208	4.202	4.118	42.080	62.080	42.020	62.020	41.180	61.180
11	23MMC4	57.9	178.3	1	4.333	4.333	4.256	43.330	65.330	43.330	65.330	42.560	64.560
12	22MMC4	49.7	186.1	20	4.395	4.398	4.333	43.950	67.950	43.980	67.980	43.330	67.330
13	C7	98.5	187.7	19	4.648	4.635	4.571	55.776	77.776	55.620	77.620	54.852	76.852
14	2MC6	90.0	194.6	38	4.826	4.820	4.800	57.912	81.912	57.840	81.840	57.600	81.600
15	3MC6	92.0	191.3	33	4.917	4.908	4.845	59.004	83.004	58.896	82.896	58.140	82.140
16	3EC5	93.5	189.3	8	5.008	4.996	4.937	60.096	84.096	59.952	83.952	59.244	83.244
17	24MMC5	80.4	201.7	8	5.008	5.005	4.932	60.096	86.096	60.060	86.060	59.184	85.184
18	23MMC5	89.7	198.0	23	5.098	5.093	5.029	61.176	87.176	61.116	87.116	60.348	86.348
19	22MMC5	79.2	205.9	33	5.096	5.096	5.061	61.152	89.152	61.152	89.152	60.732	88.732
20	33MMC5	86.0	201.2	17	5.188	5.184	5.068	62.256	90.256	62.208	90.208	60.816	88.816
21	223MMMC4	80.8	204.5	20	5.280	5.280	5.200	63.360	93.360	63.360	93.360	62.400	92.400
22	C8	125.1	208.7	27	5.331	5.317	5.250	74.634	100.634	74.438	100.438	73.500	99.500
23	2MC7	117.6	215.4	50	5.498	5.491	5.477	76.972	104.972	76.874	104.874	76.678	104.678
24	3MC7	116.5	212.5	47	5.600	5.591	5.547	78.400	106.400	78.274	106.274	77.658	105.658
25	4MC7	117.7	212.0	27	5.636	5.624	5.532	78.904	106.904	78.736	106.736	77.448	105.448
26	3EC6	118.6	210.7	37	5.740	5.724	5.692	80.360	108.360	80.136	108.136	79.688	107.688
27	25MMC6	109.1	222.5	14	5.673	5.665	5.596	79.422	109.422	79.310	109.310	78.344	108.344
28	24MMC6	109.5	219.2	21	5.774	5.765	5.694	80.836	110.836	80.710	110.710	79.716	109.716
29	23MMC6	115.6	213.8	41	5.805	5.799	5.759	81.270	111.270	81.186	111.186	80.626	110.626
30	34MMC6	117.7	212.8	19	5.873	5.865	5.760	82.222	112.222	82.110	112.110	80.640	110.640
31	3E2MC5	115.6	211.0	24	5.909	5.899	5.857	82.726	112.726	82.586	112.586	81.998	111.998
32	22MMC6	106.8	224.6	50	5.772	5.768	5.759	80.808	112.808	80.752	112.752	80.626	112.626
33	33MMC6	111.9	220.0	38	5.907	5.902	5.824	82.698	114.698	82.628	114.628	81.536	113.536
34	234MMMC5	113.5	217.3	15	5.979	5.973	5.881	83.706	115.706	83.622	115.622	82.334	114.334
35	3E3MC5	118.2	214.9	15	6.012	6.002	5.913	84.168	116.168	84.028	116.028	82.782	114.782
36	224MMMC5	99.2	224.0	34	5.943	5.943	5.886	83.202	117.202	83.202	117.202	82.404	116.404
37	223MMMC5	110.0	220.0	30	6.044	6.043	5.978	84.616	118.616	84.602	118.602	83.692	117.692
38	233MMMC5	114.8	216.3	23	6.082	6.076	5.977	85.148	119.148	85.064	119.064	83.678	117.678
39	2233MMMMC4	106.4	225.3	4	6.217	6.220	6.109	87.038	125.038	87.080	125.080	85.526	123.526

[a] C*n*, n-membered main chain; M, methyl; E, ethyl. [b] *CRC Handbook of Chemistry and Physics*, 75th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1995. [c] Clark, T.; McKervey, M. A. *Comprehensive Organic Chemistry*; Stoddart, J. F., Ed.; Pergamon: Oxford, 1979, Vol. I, p 79. [d] iter'ns = iterations.

third place becomes a constant figure. For example, in the case of 3-methylhexane cited above, 29 more iterations, each carried to 6 decimal places, gave a 3-place constant vertex value of 4.916 791−4.916 992, even though the iteration cycles had not yet attained to a new constancy! Thirty more iterations were required to reach a new level of iteration constancy, at which point the vertex values differed by 0.000 003, the difference of values ranging from 4.916 900 to 4.916 897. The final result to three decimal places was 4.917.

The iterative process was exhaustive and exhausting, and many experiments were performed to attempt to shorten the iterative procedure described and illustrated above. The results of one successful experiment merit description. It was discovered that the gross total average of the first complete iteration cycle, including the initial CIR vertex vector sums of the CIR(**D**) matrix, the intermediate figures, and the first set of d_i' values, yielded a TLFCIR value almost the same—in some instances the same—as that attained via the painstaking iteration procedure described above and illustrated in Figure 1d,e. Comparisons of the two sets of data are recorded in Table 2, columns 6 and 7. The average of only the first set of d_i' values is defined as the mean-value average value; it is slightly less "accurate" than the first unit gross total average described above, 3.445 in the example for 2-methylbutane pictured in Figure 1d.

Tucker and Small[13] also described a mode of accomplishing several individual iterations in one operation by applying a weighted average of an entire series of values, as shown in eq 6. The difference between a series of several successive

$$d_i' = (d_{i-2} + 2d_{i-1} + 3d_i + 2d_{i+1} + d_{i+2})/9 \qquad (6)$$

iterations and the bulk operation of eq 6 in their cited example was about 0.5%. In spite of the cumbersome character of eq 6, an exploratory study using 2-methylbutane was initiated. Focusing upon the first vertex and using all of the vertices of 2-methylbutane gave the bulk expression which differed (1.219%) from the total TLF value (3.446)

$$\text{bulk TLF} = (4 \times 3.200 + 3 \times 3.800 + 2 \times 3.600 + \\ 1 \times 3.000 + 2 \times 3.200)/12 = 3.400 \quad (7)$$

of 2-methylbutane listed in Table 2, column 6. The same operations were executed using the successive vertices of 2-methylbutane, so that the sequence of bulk TLF values determined from different starting points could be compared: (1) 3.400, (2) 3.440, (3) 3.420, and (4) 3.350. Similarly 2,4-dimethylpentane with a total TLF vertex value of 5.008 gave a sequence of bulk TLF vertex values of (1) 4.932, (2) 4.975, and (3) 5.064. Its structural isomer 3-ethylpentane with a total TLF vertex value also of 5.008 yielded a sequence of bulk TLF values of (1) 4.937, (2) 5.000, and (3) 5.072. Even with these relatively simple graphs it was difficult to conclude which kind of interior vertex, when expanded to attain bulk TLF individual vertex values, gave results closest to the "accurate," that is, total iterations TLF value for individual vertices. It was observed, however, that in both instances pictured above the bulk TLF vertex values were different from the two identical (5.008) total TLF vertex values of these two heptane isomers.

In spite of the fact that the results of the above described procedures were cumbersome to attain and less accurate, by

Table 3. Sequences of Increasing Complexities of *x,x*-Dimethylhexanes

Bertz[14]	2,5	2,4	2,3	3,4	2,2	3,3
Balaban, J[14]	2,5	2,4	2,2	2,3	3,4	3,3
Schultz, S, this paper	2,5	2,4	2,3	3,4	2,2	3,3
Schultz, S', this paper	2,5	2,2	2,4	2,3	3,4	3,3

about 1−2%, than the total iterations or first unit iteration averages for individual vertices, this direction of research was nonetheless pursued. The results, for consistencies of comparisons, were based on the first vertex of each graph, and were designated as bulk TLF vertex values. The individual bulk TLF vertex values were listed in column 8 and the derived S' and S TLFCIR topological index values in columns 13 and 14 of Table 2.

<div align="center">RESULTS</div>

The S indices listed in increasing order of value in Tables 1 and 2 are monotonic, single sum, and invariant—that is, descriptors. The S indices derived by applying total linear filtration (TLF) procedures to the constant interval reciprocals (CIR) of alkane **D** matrices were used to analyze in detail the three major targets of this investigation—that the S indices (1) increased in magnitude as the complexity of their corresponding alkanes increased, (2) were within this limited sample unique, and (3) reflected creditable quantitative structural—physical property relationships when regressed against positive associations.

The TLFCIR(S) indices, Table 2, column 10, differ little from the much more easily calculated S indices, Table 2, column 12, acquired via the vertex values listed in column 7 of Table 2, averages of merely the very first iteration units.

The S' indices of Tables 1 and 2, easily calculated on the way to obtaining the S indices, were also listed; however, attention was focused mainly upon the more important, in the purview of this presentation, S indices.

The increase in value of the S indices that parallels the augmentation of alkane complexity is succinctly pictured in Table 1, which lists all the various 4-carbon alkane graphs in increasing order of complexity, a table extrapolated from that published earlier by Bertz.[4,5]

Table 2 further portrays, by the standards established within this paper, the increasing complexity of the alkanes from ethane through the octane isomers due to increased chain length and substitution of methyl and ethyl groups onto the main chain. Inserting a substituent, usually methyl, upon a given straight chain increased the complexity of a molecule, as can be seen in compounds 9 and 10 of Table 2. When the alkyl group was moved from the periphery to the center of a graph, complexity increased. Compounds 11 and 12 of Table 2 demonstrate that the doubling of substituents on a carbon chain increased complexity even more, with geminal-substituted molecules more complex than the related vicinal-substituted compounds. A more scrupulous view of complexity is seen in Table 3, which duplicates a format reported by Medeleanu and Balaban.[14] The first two rows of the table present data from the reference;[14] the third row summarizes S data of this work, comparing the relative complexities of the six dimethylhexane isomers. The Bertz and Schultz sequences differ slightly from the Balaban sequence in that the two former sequences designated the geminal-disubstituted hexanes as the most complex of the pictured series.

TOPOLOGICAL ORGANIC CHEMISTRY. 12

J. Chem. Inf. Comput. Sci., Vol. 40, No. 1, 2000 **111**

Table 4. Regression Data Comparing Boiling Points and Heats of Formation with Total Linear Filtration−Constant Interval Reciprocal S' and S Indices of the 39 Alkanes, Ethane through the Octanes

property	TLFCIR index	equation	r^2	s	F
bp, °C	S' [a]	$-103.677 + 24.3024(S')^{1/2}$	0.958	10.26	833.85
	S [a]	$-118.986 + 22.0322(S)^{1/2}$	0.944	11.82	620.24
	S' [b]	$-103.709 + 24.3191(S')^{1/2}$	0.957	10.33	822.16
	S [b]	$-118.991 + 22.0407(S)^{1/2}$	0.943	11.88	613.52
	S' [c]	$-103.225 + 24.4135(S')^{1/2}$	0.957	10.31	826.73
	S [c]	$-118.844 + 22.1220(S)^{1/2}$	0.944	11.84	618.07
$(-1)H_f(g)$, kJ/mol	S' [a]	$62.714 + 17.1791(S')^{1/2}$	0.979	4.999	1756.5
	S [a]	$50.287 + 15.7533(S)^{1/2}$	0.988	3.882	2937.16
	S' [b]	$62.627 + 17.1994(S')^{1/2}$	0.980	4.953	1790.0
	S [b]	$50.229 + 15.7655(S)^{1/2}$	0.988	3.858	2974.74
	S' [c]	$62.979 + 17.2630(S')^{1/2}$	0.980	4.970	1778.05
	S [c]	$50.340 + 15.8230(S)^{1/2}$	0.988	3.806	3057.79

[a] Total TLF iterations. [b] First iteration unit average. [c] Bulk TLF computation.

Table 5. Regression Data Comparing Boiling Points and Heats of Formation with Total Linear Filtration−Constant Interval Reciprocal Values of the 39 Alkanes, Ethane through the Octanes

property	TLFCIR index	equation	r^2	s	F
bp, °C	a	$-103.686 + 37.547(\text{TLFCIR})$	0.929	13.25	485.95
	b	$-235.17 + 144.068(\text{TLFCIR})^{1/2}$	0.943	11.90	610.53
	c	$-234.662 + 144.662(\text{TLFCIR})^{1/2}$	0.945	11.65	638.68
$(-1)H_f(g)$, kJ/mol	a	$60.759 + 26.9(\text{TLFCIR})$	0.979	4.992	1761.50
	b	$60.632 + 26.996(\text{TLFCIR})$	0.980	4.950	1792.26
	c	$61.199 + 27.2246(\text{TLFCIR})$	0.981	4.854	1865.64

[a] Total TLF iterations. [b] First iteration unit average. [c] Bulk TLF computation.

Attention is invited to compound 31, Table 2 (3-ethyl-2-methylpentane), whose S number placed it between the 3,4- and the 2,2-dimethylhexanes in this paper's sequence. The second and fourth rows of Table 3 exhibit a marked similarity between the Balaban J and the Schultz S' data; only the 2,2- and the 2,4-dimethylhexane isomers are reversed.

It was most satisfying to observe in Table 2 that, contrary to prior papers of this series, all of the S values of the listed compounds were unique; none was degenerate. Compounds 16 and 17 of Table 2 (3-ethylpentane and 2,4-dimethylpentane), however, both possessed the same TLFCIR vertex values (5.008) at total iterations to constancy—eight cycles in both instances when the operations were carried from beginning to end to three decimal places. As a consequence of this identity, the S' values for the two compounds were the same (60.096), but it was noted that the S' values based upon the first iteration unit gross total average vertex values (4.996 and 5.005, respectively) differed slightly (59.952 and 60.060). It was fascinating to observe the initially different CIR vertex data for the two heptane isomers gradually merge to total linear filtration identical constancy after eight iterations. To test whether the same values of 5.008 attained through eight cycles of linear filtration to complete constancy for the vertex values might have been the happenstance consequence of rounding off intermediate iteration values to three decimal places, the total linear filtration iterations for the two compounds were repeated, carrying all operations from beginning to end to six decimal places. After 15 iterations in each instance, complete and total constancy of all vertices in the two sets of TLFCIR vertex values was attained—exactly the same values, 5.007 522! This was accepted as evidence of what was earlier in the paper referred to as coincidental identity, an identity, however, that still resulted in different S values.

The two sets of data derived via total linear iterations and via first iteration unit averages were very similar, indeed, in some instances identical, and yielded almost identical S index values, as well as having the same increasing order of S values. However, the data attained from the bulk TLF values differed somewhat, 1−2% as described above, from the two prior data sets. Although all the bulk TLFCIR(S) index values were unique, two pairs of indices, compounds 24/25 and 36/37 of Table 2, had their orders reversed in contrast with the lists of S index values described above. On the other hand, compounds 16 and 17 of Table 2, both of which had the same (5.008) total iterations TLF vertex values, had different (4.937 and 4.932) bulk TLF vertex values, as seen in column 8, Table 2.

The third objective of this study was an examination of the utility for QSPR comparisons of the S index calculated by use of constant interval reciprocal cum total linear filtration **D** matrices.

Two such studies were performed. The first, an instance of positive QSPR association, investigated the relationships of the S indices and the energies of formation of the 39 alkanes as listed in column 4 of Table 2. The second study concerned the negative association exhibited by the alkane S indices versus the boiling points in column 3 of Table 2. The regression data for the above circumstances are listed in Table 4.

The regression data demonstrating the positive association of S indices with the free energies of formation displayed in Table 4 gave a very satisfactory average r^2 value of 0.99, with a very narrow range of values. These r^2 data conform to the advice of Mihalić and Trinajstić[3] referred to above.

The negative association data of the S indices versus the boiling points of the 39 alkanes of Table 2 were also

interesting. Table 4 disclosed the r^2 average value of $r^2 = 0.94$, with a range over the three modes of arriving at the TLF($\mathbf{D}$) values of $r^2 = 0.943-0.944$.

The less valued S' index regression data, similar to the S index data, are also listed in Table 4. Overall, for energies of formation and for boiling points, r^2 values ranging from 0.96 to 0.98 were observed.

With the finalization of the above data, the targets of this study were attained. In summary and recapitulation, however, it was observed that the three TLF sets of values listed in columns 6−8 of Table 2 possessed more than the usual edge count data derived from the distance matrix. Algorithms (5 and 6) not only utilized CIR edge count data, but also incorporated TLF operations that recognized the degree of each graph vertex. Also, in the specific instance of alkanes the adjacency status of each vertex was also reflected. Therefore, in what was begun as an uncertain experiment, the TLFCIR($\mathbf{D}$) vertex values of each graph were also subjected to the regression analyses described above. As for the S' and S regression data of Table 4, the six sets of equations, r^2, s, and F values for the TLFCIR($\mathbf{D}$) data, displayed in Table 5, were similar. The boiling point regression data had an average r^2 value of 0.94; the heat of formation regression values were slightly better, with an r^2 average of 0.98. These results may be a surprising entrance into a new kind of molecular topological index that is also single-sum, invariant, and monotonic, increases in value with greater complexity of molecules, and that also gives promise of creditable QSPR relationships!

CONCLUSIONS

This paper has described a technique for applying constant interval reciprocal−total linear filtration $\mathbf{D}$ matrices to the calculation of S topological indices for the group of 39 C_2-C_8 alkanes. These whole-molecule descriptors, monotonic, invariant, and single-sum, were unique, increased in dimensions as did the complexities of molecules, and exhibited creditable QSPR positive, but less than useful negative, association relationships.

REFERENCES AND NOTES

(1) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228.
(2) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.
(3) Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701−712, and references cited therein.
(4) Bertz, S. H. The Bond Graph. *J. Chem. Soc., Chem. Commun.* **1981**, 818−820.
(5) Bertz, S. H. Branching in Graphs and Molecules. *Discuss. Appl. Math.* **1988**, *19*, 65−83.
(6) Plavsić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235−250.
(7) Ivanciuć, O.; Balaban, T. S.; Balaban, A. T. Reciprocal Distance Matrix; Related Local Vertex Invariants and Topological Indices. *J. Math. Chem.* **1993**, *12*, 309−318.
(8) Diudea, M. V. Indices of Reciprocal Properties or Harary Indices. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 292−299, and references cited therein.
(9) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundation of Chemistry. 5. The Heuristics of Primacy. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 835−840.
(10) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 11. Graph Theory and Reciprocal Schultz-Type Molecular Topological Indices of Alkanes and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 853−857.
(11) Grznar, J.; Booth, D. E.; Sebastian, P. The Use of Robust Smoothers in Nuclear Material Safeguards. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 236−240.
(12) Grznar, J.; Booth, D. E.; Sebastian, P. A Robust Smoothing Approach to Statistics of Process Control. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 241−248.
(13) Tucker, A.; Small, D. *A Unified Introduction to Linear Algebra*; Macmillan: New York, 1988; pp 48−57.
(14) Medeleanu, M.; Balaban, A. T. Real-Number Vertex Invariants and Schultz-Type Indices Based on Eigenvectors of Adjacency and Distance Matrices. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1038−1047.

CI990059T