

**Alkane Ordering as a Criterion for Similarity between Topological Indices: Index J as a
“Sharpened Wiener Index”**

Alexandru T. Balaban,^a Denise Mills,^b and Subhash C. Basak^b

^a Texas A&M University at Galveston, Department of Marine Sciences, Galveston, TX 77553-1675, USA

^b Natural Resources Research Institute, University of Minnesota - Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA

Abstract. Constitutional isomers of alkanes with 6 through 9 carbon atoms are ordered differently by various topological indices. This fact allows the detection of similarity and dissimilarity between topological indices. The ordering induced by the average distance connectivity index J parallels the ordering induced by Wiener's index, but reduces the degeneracy of the latter index providing a much higher discriminating ability.

E-mail addresses: balabana@tamug.tamu.edu, and sbasak@nrri.umn.edu

Introduction

Among structural features that characterize molecular complexity, one can distinguish: *number of atoms; diversity of bonds and of hybridization types; branching leading to constitutional isomerism; cyclicity classified according to the cyclomatic number, and to the type of polycyclic systems (e.g. spirocyclic, isolated, condensed, bridged bicyclics, etc.);* simple stereochemical features leading to enantiomers or diastereomers. All these features apply equally to hydrocarbons and to molecules possessing heteroatoms. The following additional features apply to molecules consisting of more than two kinds of atoms (which are usually carbon and hydrogen, as in hydrocarbons): *diversity of atoms; distribution of heteroatoms in the molecule;* complex stereochemical features (e.g. secondary and tertiary structure of polypeptides); supramolecular features (e.g. catenanes, rotaxanes, knots, non-covalent interactions).

The topology of molecules, which includes all features italicized in the preceding paragraph, allows chemists to count all constitutional isomers, and can be correlated with physical-chemical properties. Moreover, in many instances the molecular topology can also be correlated with biological activities, although for the latter activities steric, hydrogen-bonding ability, hydrophobic, and electrical features can also play an important role. Molecular topology can be expressed numerically by topological indices for QSPR/QSAR (quantitative structure-property/activity relationships).

Topological indices¹⁻⁶ have become a focus of attention for several applications owing to their fast and easy computation for any covalently-bonded chemical structure. The recent advances in combinatorial chemistry and high-throughput screening have led to exploring huge virtual combinatorial libraries, and the molecular descriptors that can serve in such QSAR studies must satisfy two requirements: (i) they should be independent of experimentally determined parameters and (ii) they must be able to be computed with minimal cpu duration. Most often, topological indices (TIs) are used along with other molecular descriptors in multiparametric correlations between physical-chemical properties (for QSPRs) or biological activities (for QSARs) and the chemical structure of the corresponding compounds. Examples of applications for TIs in the optimization of biologically active lead compounds abound.⁷ Interestingly, even for systems that include non-covalent interactions such as hydrogen bonds in transfer RNA,

inclusion of information on these hydrogen bonds allowed topological indices to be useful in applications.⁸

The large number of extant TIs is sometimes a discouraging factor for scientists who would like to explore their use in QSAR or QSPR studies. It is, however, necessary to have at one's disposal a pool of TIs from which to select those TIs that are most closely correlated with the property or biological activity of compounds in the database. Sometimes, the calculation of TIs includes information from the database so that the TIs are specially adapted to that particular database, as advocated by Randic.⁹ The problem of characterizing branching in alkanes was reviewed by Rouvray,¹⁰ who pointed out decisive contributions by Randic with his molecular connectivity index¹¹ that was later extended by Kier and Hall.¹²⁻¹⁴ Other authors who enriched this field are Bonchev and Trinajstić who were the first to apply information theory to chemical graphs.^{15,16} Bonchev *et al.* also developed overall connectivities,¹⁷ the overall Wiener index,¹⁸ and (with a few other authors) indices obtained on the basis of hierarchically ordered connectivities.^{19,20}

Throughout the present paper, the word "isomer" will be used for brevity to indicate "constitutional isomer". It should be noted that among the nine such isomeric heptanes, two of them (isomers denoted by letters C and F in Table 1) are chiral having an asymmetric carbon atom, and therefore each of these two gives rise to a pair of enantiomers raising the total to 11 when counting also stereoisomers. Among the higher alkanes, the numbers of stereoisomers - both diastereomeric and enantiomeric ones - increases faster than the number of constitutional isomers so that already for undecanes the constitutional isomers are overtaken: there are 9 (constitutional) isomers of heptane and 11 steric isomers, as shown above; 18 isomers and 24 steric isomers of octane; 35 isomers and 55 steric isomers of nonane; 75 isomers and 136 steric isomers of decane; 159 isomers and 345 steric isomers of undecane, etc.^{21,22}

In order to avoid using TIs that include the same type of dependence on the molecular structure, and to test only TIs that have a chance to be orthogonal to each other, several attempts have been made for grouping together TIs that are similar. A first method was to explore intercorrelations between various TIs that were known at that time.²³ Later, variable clustering on hydrocarbons and on sets with diverse functionalities allowed numerous TIs to be examined,²⁴ and this study has been recently extended to a database containing about 3000 diverse structures

and to all indices included in the MOLCONN-Z program.²⁵ Another approach was to investigate what was the partial ordering of isomeric heptanes and octanes induced by various TIs based on a variety of symmetrical matrices.²⁶

The present paper is a continuation of the latter method, but it includes the ordering of alkanes with six through nine carbon atoms, and it explores a large set of TIs. We start with stating that saturated acyclic hydrocarbons (alkanes) constitute an excellent testing ground for TIs: *intramolecular comparison* of local vertex invariants (LOVIs) allows to test whether these LOVIs vary in a consistent manner with the vertex degree and the centricity of branching; whereas *intermolecular comparison* for a series of isomeric alkanes reveals whether the TIs have a logical trend of variation from the linear alkane as the least branched isomer, and the alkane most closely approximating a star graph as the most branched one. Star graphs have all their vertices except the central one with degree one (endpoints): the hydrogen-depleted graphs of isobutane and neopentane are the star graphs with 4 and 5 vertices, respectively. However, alkanes are 4-trees (quartic trees), with an upper limit of 4 for the vertex degree, and therefore the way in which branched alkanes with more than five carbon atoms come closest to star graphs is debatable.

Cayley was the first mathematician to become interested in the enumeration of alkane constitutional isomers. Later, chemists enumerated alkane isomers: Losanitsch, Hermann, Senior, Blair and Henze,²⁷ but their approach was based on cumbersome empirical methods. Polya succeeded, however, to develop a general and rigorous method for solving this problem.²⁸ His main theorem was adopted, refined, and used by Robinson *et al.*,²¹ Read,²² and many other authors.

Ordering of alkanes by various topological indices

Previously, it was noted that various topological indices led to different orderings of alkanes, and several such orderings were noted and listed.²⁹ Thus it was noted that centric indices, which (like other early first-generation TIs) had high degeneracy, led to an ordering that was different from that of other TIs known at that time.²⁹

In an early paper,³⁰ Bertz used successive line graph derivatives for alkanes as a measure of branching. He proceeded in a rigorous manner, and postulated the linear alkane as the least

branched in an isomeric series, and the star graph as the most branched graph. Then he assumed that the numbers of edges in iterated line graphs are an operational criterion for characterizing branching, and he presented the ordering of all alkanes with one through eight carbon atoms based on this idea (it is sufficient to list graph derivatives through the fifth derivative, N^V for complete isomer discrimination). He observed that among the then known TIs only one other index (Balaban's average distance-based connectivity index J ^{31,32}) led to the same ordering of alkanes with 1 through 7 carbon atoms. He also remarked that in the induced ordering of the octane isomers, index J also had the smallest deviations from N^V among all other indices.

It should be made clear that we do not claim that the ordering based on index J is the best, since one cannot know which properties or biological activities will be appropriately reflected by any particular TI. As an example for QSPR, in most instances, thermodynamic properties (such as the normal boiling point) increase with increasing number of carbon atoms but decrease with increasing branching. However, some chemical properties (such as the octane number) decrease with increasing number of carbon atoms and with increasing branching.^{33,34} Our only aim is to use the alkane ordering as a criterion for examining similarities between the numerous TIs. In other words, to paraphrase Orwell, all TIs are different, but some TIs are more different than others.

Based on the common ordering between J and N^V for alkanes with up to 7 carbon atoms, Table 1 lists a few TIs so that deviations from this ordering become prominent. In all tables, M denotes methyl, E denotes ethyl, and C_n denotes an alkane with n carbon atoms. One can observe for instance that the earliest TI, due to Wiener and denoted by W ,³⁵ orders heptanes exactly in the same manner as J and N^V , but it has a high degeneracy with two pairs of heptane isomers sharing the same value for W . However, it will become apparent that, since both W and J are distance-based TIs, they vary in the same order, i.e. whenever two alkanes with different J values have degenerate W values, they occur together, one after the other, without any gap, up to nonane isomers (such gaps start to appear with decane isomers). Therefore we claim that index J is a "sharpened Wiener index", allowing a better discrimination, but conserving the main features of the Wiener index. In other words, J imparts a sharper vision to W (Fig. 1).

We have examined over 300 TIs, namely all those contained in the POLLY and MOLCONN-Z programs, as well as triplet TIs. Topological indices have values that may

increase or decrease with increasing branching. By examining the values of these TIs for all alkanes up to and including octane isomers, we retained for further characterization only those indices that had extreme values for: (i) linear alkanes, and (ii) the most branched alkane which in these simple cases (hydrogen-depleted graphs with up to eight vertices) is the unique structure with the maximal number of vertices of degree 4. Also, we eliminated all indices that had zero values for any isomer (keeping, however, N^V) or that had degenerate values for any pair of hexane isomers. A few definitions and the corresponding notation for the TIs that will appear in the following discussion have to be presented. Table 2 contains the TIs that were selected with their notation.

Table 1. Hexane and heptane isomers, and some of their topological indices

Alkane	Structure	W	N^V	J	OW	χ	D
A	C6	35	0	2.3391	126	2.9142	2.6458
B	2M-C5	32	103	2.6270	154	2.7701	2.3664
C	3M-C5	31	239	2.7540	161	2.8081	2.2657
D	2,3MM-C4	29	906	2.9935	188	2.6427	2.0817
E	2,2MM-C4	28	3261	3.1685	197	2.5607	2.0000
A	C7	56	1	2.44747	252	3.4142	3.0551
B	2M-C6	52	112	2.67826	311	3.2701	2.7946
C	3M-C6	50	297	2.83182	333	3.3081	2.6547
D	2,4MM-C5	<u>48</u>	364	2.95322	384	3.1259	2.5261
E	3E-C5	<u>48</u>	561	2.99230	354	3.3461	2.5071
F	2,3MM-C5	<u>46</u>	1402	3.14421	411	3.1807	2.3905
G	2,2MM-C5	<u>46</u>	3546	3.15449	414	3.0607	2.4103
H	3,3MM-C5	44	5472	3.36044	440	3.1213	2.2678
I	2,2,3MMM-C4	42	8508	3.54120	510	2.9434	2.0354

^a Boldface underlined numbers indicate degeneracy, and boldface numbers indicate an ordering that differs from that induced by J and N^V .

Table 2. Symbols (POLLY abbreviations and corresponding standard abbreviations), definitions and classification of topostructural and topochemical parameters.

POLLY notation	Standard notation	Parameter Classification
		Topostructural
idw	I_D^W ---	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
midw	I_D^W I_D^E	Mean information index for the magnitude of distance Information index for the edges in the graph
w	W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
hv	H^V	Graph vertex complexity
hd	H^D	Graph distance complexity
m2	M_2	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
s1, s2	${}^h\chi$	Path connectivity index of order h = 1, 2 (for h = 1, χ is Randić's index)
j	J	Balaban's J index based on topological distance
dn2s _y	DN^2S_y	Triplet index from distance matrix, square of graph order (# of non-H atoms), and distance sum; operation y = 1-5
dn21 _y	DN^21_y	Triplet index from distance matrix, square of graph order, and number 1; operation y = 1-5
as1 _y	$AS1_y$	Triplet index from adjacency matrix, distance sum, and number 1; operation y = 1-5
ds1 _y	$DS1_y$	Triplet index from distance matrix, distance sum, and number 1; operation y = 1-5
asn _y	ASN_y	Triplet index from adjacency matrix, distance sum, and graph order; operation y = 1-5
dsn _y	DSN_y	Triplet index from distance matrix, distance sum, and graph order; operation y = 1-5
dn2n _y	DN^2N_y	Triplet index from distance matrix, square of graph order, and graph order; operation y = 1-5
ans _y	ANS_y	Triplet index from adjacency matrix, graph order, and distance sum; operation y = 1, 2
an1 _y	$AN1_y$	Triplet index from adjacency matrix, graph order, and number 1; operation y = 1-5
ann _y	ANN_y	Triplet index from adjacency matrix, graph order, and graph order again; operation y = 1-4
asv _y	ASV_y	Triplet index from adjacency matrix, distance sum, and vertex degree; operation y = 1-5

(Continued)

Table 2. Continued

dsv _y	DSV _y	Triplet index from distance matrix, distance sum, and vertex degree; operation y = 1-5
anv _y	ANV _y	Triplet index from adjacency matrix, graph order, and vertex degree; operation y = 1-5

 Topochemical

b1, b2	$\chi^{h,b}$	Bond path connectivity index of order h = 1, 2
v1, v2	$\chi^{h,v}$	Valence path connectivity index of order h = 1, 2
azv _y	AZV _y	Triplet index from adjacency matrix, atomic number, and vertex degree; operation y = 1, 3, 5
azs _y	ASZ _y	Triplet index from adjacency matrix, atomic number, and distance sum; operation y = 1, 2
asz _y	ASZ _y	Triplet index from adjacency matrix, distance sum, and atomic number; operation y = 1-5
azn _y	AZN _y	Triplet index from adjacency matrix, atomic number, and graph order; operation y = 1-4
anz _y	ANZ _y	Triplet index from adjacency matrix, graph order, and atomic number; operation y = 1-4
dsz _y	DSZ _y	Triplet index from distance matrix, distance sum, and atomic number; operation y = 1-5
dn2z _y	DN ² Z _y	Triplet index from distance matrix, square of graph order, and atomic number; operation y = 1-5

Note: In this table, operations y = 1-5 for the triplet indices are indicated by a subscript; in the text, however, these operations are indicated by a normal font, possibly in brackets when designating several related triplet indices that differ only in the operation.

Calculation of Molecular Descriptors

The majority of the topological descriptors were calculated using software developed by Basak *et al.*, including POLLY 2.3.³⁶ The topological descriptors include Wiener number,³⁷ molecular connectivity indices developed by Randic¹¹ and Kier and Hall,¹³ frequency of path lengths of varying size,¹³ information theoretic indices defined on distance matrices of graphs using the methods of Bonchev and Trinajstic,¹⁵ Roy *et al.*,³⁸ Basak *et al.*,^{39,40} as well as those of Raychaudhury *et al.*,⁴¹ parameters defined on the neighborhood complexity of vertices in hydrogen-filled molecular graphs,^{38,39} and Balaban's J indices^{31,32,42,43} as well as the triplet indices.^{24,44} Of course, for alkanes all J indices modified for the presence of heteroatoms are equal because there are no heteroatoms. The information-theoretic index $X(D)$ was defined earlier.⁴⁵ The triplets result from a matrix, main diagonal column vector, and free term column vector which are converted into a system of linear equations. The notation used to represent the vectors and matrices is as follows:

A = Adjacency matrix

D = Distance matrix

V = Vertex degree, as well as a vector with N elements equal to the corresponding vertex degree.

S = Distance sum, as well as a vector with N elements equal to the distance sum for the corresponding vertex.

N = Total number of vertices in the graph, as well as a vector with N elements equal to N.

Z = Atomic number, as well as a vector with N elements equal to the corresponding Z value, which for hydrocarbons is 6.

1 = Unity vector

After the system of N linear equations is solved, the local vertex invariants, x_i , are assembled into a triplet descriptor based on one of the following operations:

1. Summation, $E_i x_i$;
2. Summation of squares, $E_i x_i^2$;
3. Summation of square roots, $E_i x_i^{1/2}$;
4. Sum of inverse square root of cross-product over edges ij , $E_{ij}(x_i x_j)^{-1/2}$;
5. Product, $(\prod_i x_i)^{1/N}$.

Additional topological descriptors, including an extended set of molecular connectivity indices, electrotopological state descriptors, general polarity descriptors, and hydrogen bonding descriptors, were calculated by Molconn-Z 3.50.⁴⁵ A total number of 363 topological descriptors was calculated for use in the current study, but a smaller number was selected for discussion because some of TIs were eliminated as will be seen below.

Results and discussion

The astonishing result was the following: not only was J the only other index yielding the same ordering of heptane isomers as N^V , but none of these over 300 indices paralleled the ordering of alkanes with 8 or more carbon atoms according to Bertz's N^V values. However, several indices ordered the 9 heptane isomers (Table 3) and the 18 octane isomers exactly in the same sequence as the J index. In addition to the triplet index DN^2Z2 indicated in Table 4, five other triplet indices shared the ordering with J through all constitutional isomers of octane: DN^2N1 , DN^2N2 , DN^2I1 , DN^2I2 , DN^2I3 . On the other hand, for the 35 constitutional isomers of nonane, the ordering differed from that induced by J and displayed in Table 5 (with one pair for DN^2Z2 , DN^2N1 , DN^2N2 , DN^2I1 and with two pairs for the other two triplet indices mentioned above, namely DN^2I2 and DN^2I3). Unlike previous tables, nonane isomers are denoted by numbers and not by letters in Table 5, because the number of nonane isomers is larger than the number of letters in the alphabet.

Table 3. Heptane isomers and some topological indices that lead to the same ordering

Comp.	Structure	N^V	J	X(D)	DN^2Z2	AS11	ASN1	DSN1
A	C7	1	2.44747	1.01950	0.060457	4.0730	2.8511	1.7080
B	2M-C6	112	2.67826	1.10780	0.062467	4.3330	3.0331	1.8293
C	3M-C6	297	2.83182	1.15881	0.063604	4.4913	3.1439	1.9222
D	2,4-MM-C5	364	2.95322	1.20739	0.064573	4.6253	3.2377	1.9602
E	3E-C5	561	2.99230	1.20952	0.064718	4.6565	3.2596	2.0111
F	2,3MM-C5	1402	3.14421	1.27065	0.065785	4.8029	3.3620	2.0720
G	2,2MM-C5	3546	3.15449	1.28380	0.065819	4.8054	3.3638	2.0829
H	3,3MM-C5	5472	3.36044	1.34790	0.067003	4.9897	3.4928	2.1869
I	2,2,3MMM-C4	8508	3.54120	1.41474	0.068110	5.1564	3.6095	2.2500

In the Addendum (<http://wyle.nrri.umn.edu/wyle/downloads/>) that one may download from the web, one finds for isomeric alkanes with 3 through 8 carbon atoms, all TIs that are included in Table 2 and many other ones. Two triplet indices (ASV3 and ASV4) do not have all TIs for alkanes with 3 – 5 carbon atoms because one alkane in each case yields by chance two identical equations, leading to non-determination. Structures are indicated in the Addendum by Smiles Notation, and the ordering is according to increasing J values.

In Table 6 one sees, for a few topological indices (the same ones as in Table 1), how they differ from index J in ordering octane isomers. Similar and more detailed information is provided by Table 7, with the same notation of the heptane and octane isomers as in the preceding tables. The absence of gaps in W values for nonane isomers ordered according to increasing J values is observed in Table 5. Starting with decane isomers, such gaps do appear, however.

Table 4. Octane isomers and some of their topological indices ^a

Alkane	Structure	W	J	X(D)	DN ² Z ₂
A	C8	84	2.53006	0.97065	0.040417
B	2M-C7	79	2.71584	1.03795	0.041522
C	3M-C7	76	2.86207	1.08551	0.042277
D	4M-C7	75	2.91961	1.10261	0.042545
E	2,5MM-C6	74	2.92782	1.11164	0.042665
F	3E-C6	72	3.07437	1.15017	0.043280
G	2,4MM-C6	<u>71</u>	3.09883	1.16594	0.043458
H	2,2MM-C6	<u>71</u>	3.11177	1.18983	0.043758
I	2,3MM-C6	70	3.17082	1.18983	0.043754
J	3,4MM-C6	68	3.29248	1.22755	0.044270
K	3E-2M-C5	<u>67</u>	3.35488	1.24448	0.044525
L	3,3MM-C6	<u>67</u>	3.37338	1.25914	0.044563
M	2,2,4MMM-C5	66	3.38892	1.26942	0.044726
N	2,3,4MMM-C5	65	3.46423	1.28678	0.045009
O	3E-3M-C5	64	3.58321	1.32174	0.045354
P	2,2,3MMMC5	63	3.62328	1.34275	0.045579
Q	2,3,3MMMC5	62	3.70832	1.36777	0.045859
R	2,2,3,3-C4	58	4.02039	1.47446	0.046941

^a Boldface underlined numbers indicate degeneracy.

Table 5. Nonane ordering

No.	Nonane isomer	W	J	$\lambda\lambda_1$	Rank of $\lambda\lambda_1$
1	C9	120	2.59508	1.35308	1
2	2-M-C8	114	2.74669	1.34140	2
3	3-M-C8	110	2.87662	1.33562	3
4	2,6-MM-C7	108	2.91466	1.32945	5
5	4-M-C8	108	2.95482	1.33318	4
6	2,5-MM-C7	104	3.06082	1.32335	8
7	2,2-MM-C7	104	3.07299	1.31952	10
8	3-E-C7	104	3.09225	1.32731	6
9	2,4-MM-C7	102	3.15125	1.32039	9
10	2,3-MM-C7	102	3.15528	1.31937	11
11	4-E-C7	102	3.17534	1.32483	7
12	3,5-MM-C7	100	3.22305	1.31672	12
13	2,2,5-MMM-C6	98	3.28071	1.30682	17
14	4-Et-2-Me-C6	98	3.30739	1.31418	13
15	3,4-MM-C7	98	3.32476	1.31270	14
16	3,3-MM-C7	98	3.33007	1.30933	16
17	2,3,5-MMM-C6	96	3.37660	1.30617	19
18	3-E-2-M-C6	96	3.41009	1.31015	15
19	4,4-MM-C7	96	3.43105	1.30633	18
20	2,2,4-MMM-C6	94	3.46726	1.29963	21
21	3-E-4-M-C6	94	3.49948	1.30594	20
22	2,3,4-MMM-C6	92	3.57583	1.20792	23
23	2,4,4-MMM-C6	92	3.57675	1.29248	26
24	2,2,3-MMM-C6	92	3.58873	1.29462	25
25	3-E-3-M-C6	92	3.61739	1.29849	22
26	3-E-2,4-MM-C5	90	3.67762	1.29483	24
27	2,3,3-MMM-C6	90	3.70209	1.29051	27
28	2,2,4,4-MMMM-C5	88	3.74642	1.28186	32
29	3-E-2,2-MM-C5	88	3.79291	1.28728	29
30	3,3,4-MMM-C6	88	3.80240	1.28626	30
31	3,3-EE-C5	88	3.82468	1.29007	28
32	2,2,3,4-MMMM-C5	86	3.87760	1.27916	33
33	3-E-2,3-MM-C5	86	3.91921	1.28209	31
34	2,3,3,4-MMMM-C5	84	4.01374	1.27403	34
35	2,2,3,3-MMMM-C5	82	4.14473	1.26641	35

Boldface characters indicate degenerate W values without gaps in J, and in the last

column an order for index $\lambda\lambda_1$ that differs from that induced by index J.

Table 6. Octane isomers and some of their topological indices

Alkane	Structure	W	N ^V	J	OW	χ	D
A	C8	84	2	2.53006	462	3.9142	3.4641
B	2M-C7	79	114	2.71584	572	<i>3.7701</i>	3.2238
C	3M-C7	76	307	2.86207	622	<u>3.8081</u>	3.0706
D	4M-C7	75	356	2.91961	636	<u>3.8081</u>	3.0178
E	2,5MM-C6	74	242	2.92782	709	3.6259	2.9761
F	3E-C6	72	638	3.07437	683	3.8461	2.8536
G	2,4MM-C6	<u>71</u>	586	3.09883	771	3.6639	2.8221
H	2,2MM-C6	<u>71</u>	3573	3.11177	781	3.5607	2.8473
I	2,3MM-C6	70	1488	3.17082	790	3.6807	2.7775
J	3,4MM-C6	68	1936	3.29248	837	<u>3.7187</u>	2.6726
K	3E-2M-C5	<u>67</u>	2078	3.35488	848	<u>3.7187</u>	2.6118
L	3,3MM-C6	<u>67</u>	5795	3.37338	859	3.6213	2.6213
M	2,2,4MMM-C5	66	4252	3.38892	968	3.4165	2.5912
N	2,3,4MMM-C5	65	3173	3.46423	981	3.5534	2.5284
O	3E-3M-C5	64	8286	3.58321	921	3.6820	2.4785
P	2,2,3MMMC5	63	9700	3.62328	1049	3.4814	2.4422
Q	2,3,3MMMC5	62	11664	3.70832	1065	3.5040	2.3905
R	2,2,3,3-C4	58	27702	4.02039	1312	3.2500	2.2039

^a Boldface underlined numbers indicate degeneracy. Italicized boldface numbers indicate an ordering that differs from that induced by J.

Table 7. Ordering of heptane and octane isomers

J, N^V , etc.	W	I_D^E	I_D^W	Z	χ	D	MTI	OW	AZ(RDS)1
Heptane isomers									
A	a	a	a	a	a	a	a	a	a
B	b	b	b	<u>d</u>	<u>d</u>	b	b	b	b
C	c	c	c	c	c	c	c	c	c
D	de	d	<u>e</u>	g	<u>f</u>	d	d	<u>e</u>	d
E	de	e	<u>d</u>	<u>b</u>	<u>b</u>	e	e	<u>d</u>	e
F	fg	g	f	<u>e</u>	<u>e</u>	g	g	f	f
G	fg	f	g	<u>h</u>	<u>h</u>	<u>f</u>	<u>f</u>	g	g
H	h	h	h	f	g	h	h	h	h
I	i	i	i	i	i	i	i	i	i
Octane isomers									
A	a	a	a	a	a	a	a	a	a
B	b	b	b	<u>ef</u>	g	b	b	b	b
C	c	c	c	c	<u>cd</u>	c	c	c	c
D	d	d	d	d	<u>cd</u>	d	d	d	d
E	e	e	e	<u>kl</u>	<u>k</u>	e	e	<u>f</u>	<u>f</u>
F	f	g	f	<u>b</u>	<u>b</u>	f	fg	<u>e</u>	<u>e</u>
G	gh	<u>h</u>	g	<u>i</u>	<u>i</u>	<u>h</u>	<u>h</u>	g	g
H	gh	f	h	<u>no</u>	<u>m</u>	g	fg	h	<u>i</u>
I	i	i	i	i	i	i	i	i	<u>h</u>
J	j	j	j	<u>ef</u>	<u>ef</u>	j	j	j	j
K	kl	<u>l</u>	k	<u>h</u>	<u>ef</u>	k	lm	k	k
L	kl	<u>k</u>	l	<u>kl</u>	l	<u>n</u>	<u>k</u>	l	l
M	m	m	m	<u>q</u>	<u>q</u>	<u>l</u>	lm	<u>n</u>	<u>n</u>
N	n	n	n	<u>m</u>	n	<u>m</u>	n	<u>o</u>	<u>m</u>
O	o	o	o	gh	<u>h</u>	o	o	m	o
P	p	p	p	<u>p</u>	<u>p</u>	p	p	p	p
Q	q	q	q	<u>no</u>	<u>o</u>	q	q	q	q
R	r	r	r	r	r	r	r	r	r

Letters A-I for heptanes indicate the structures displayed in Tables 1 and 2. Letters A-R for octanes denote the structures presented in Tables 3 and 4. Entries with two letters denote degeneracy. The place occupied by structures A-I or A-R ordered according to various indices is denoted by lower case letters, starting with the letter a. Boldface entries indicate an ordering that differs from that induced by J.

Discussion of the results

Several interesting observations on the data presented in Tables 1, 3-7 can be made. One is that the Randić (χ) and Hosoya (Z) indices assign a lower priority to branching at the extremes of a longer chain than to branching closer to the middle of the chain, therefore these indices differ from most of the other ones which put more emphasis on the vertex degree and/or larger topological distances. Hexane isomers are the smallest alkanes where ordering differences can be observed. The Hosoya (Z) and Randić indices (χ) as well as indices $V1$ and $V2$ lead to ACBDE ordering, and most other indices to the “normal” order (ABCDE). However, the ordering of heptanes induced by χ (AECBFDHGI) differs from that induced by Z or by the information-theoretic index I_{pc} (AECBFHDG).

The information-theoretic indices I_D^W and I_D^E differ in their ordering of heptane isomers in that the former has one pair of isomers (ED) and the latter another pair (GF) in a different order than that indicated in Table 7. Concentrating on this last Table, one can observe that although the information-theoretic index I_D^W (based like W and J on topological distances) has one pair of heptane isomers in a different order from that induced by J , it orders octane isomers exactly like J . In all cases, according to our selection criteria, the normal (linear) alkane and the most branched alkane appear always at the extremes of the ordering for all TIs included in Table 7. There is also an interesting agreement for all TIs of Table 7 on the place occupied by 3-methylhexane among heptane isomers, as well as 3-methylheptane, 4-methylheptane and 2,2,3-trimethylpentane among octane isomers.

In addition to the analysis of alkane ordering due to Bertz,³⁰ we need to analyze the contribution of Milan Randić who, in 1998, published an elaborate discussion on this topic.⁴⁷ He investigated a few older TIs, namely the J index, the hyper-Wiener index and the Harary index, as well as the leading eigenvalues of four matrices: adjacency, distance, Wiener, and path matrices. Only the last TI, denoted by $\lambda\lambda_1$, has no degeneracy up to, and including decanes, and therefore Randić proposed to define quantitatively molecular branching by $\lambda\lambda_1$. He had noted, however, that J and Wiener’s index parallel the ordering induced by $\lambda\lambda_1$ up to, and including octanes, but that starting with nonanes there are discrepancies between W and $\lambda\lambda_1$. We illustrate the

parallelism between the ordering of nonanes by J and W (which is much more degenerate than J) in Table 5. The same Table also indicates the divergence between the ordering induced by $\lambda\lambda_1$. It had previously been demonstrated that J for alkane isomers starts to have degenerate values only for dodecane isomers, whereas for $\lambda\lambda_1$ it is only known that decanes have non-degenerate values. All these arguments lead to the conclusion that, for the time being, index J seems to be a better measure for branching than $\lambda\lambda_1$.

Alkane ordering induced by triplet indices

Since triplet indices are both powerful and very diverse, and since they have been little investigated, this section will present their partition according to the alkane ordering induced by them. Although one can have many triplet indices derived from less familiar square symmetrical matrices (Szeged matrix, reciprocal distance matrix, resistance distance matrix, reverse Wiener matrix, detour matrix. etc.), we consider only two such matrices, namely the adjacency matrix (**A**) and the distance matrix (**D**). Also, though one could imagine many other operations grouping together into a single number the local vertex invariants, we restrict ourselves to the five operations indicated above and denoting by numbers 1 through 5 the last digit of the triplet indices. In the following, for brevity, we will include in brackets this last indicator (operations 1 through 5) so that one may group together several triplet indices derived from the same matrix and pair of vectors, but differing in the last operation (1 through 5). From the outset, we have eliminated all indices that lead to zero or negative values for such indices, or to indices that do not have the linear alkane with an extremal value (smallest or largest).

Based on hexane ordering, there are four groups of triplet indices:

- 1) order ACBDE for AZV(1-5); ANV(1-5); AZN(1-4); ANZ(1-4); AN1(1-4); ANN(1-4).
- 2) order ABCED for ASV(4) and AZS(2).
- 3) order ACBED for ASV(5).
- 4) order ABCDE, i.e. the "normal order" for all remaining ones, which will be partitioned in turn according on their ordering of heptane isomers.

The groups of triplet indices grouped under item 4 above is subdivided into nine subgroups as follows on the basis of heptane isomer ordering:

- 1) order AECBFDHGI for AZV(1,3,5); ANV(1-3,5); AZN(1-4); ANZ(1,3,4); AN1(1-4), ANN(1-4).
- 2) order ADECBFHGI for ANZ(2).
- 3) order ABCEGDFHI for DSN(5).
- 4) order ABCDEFGHI for ASV(2).
- 5) order ABCDEGFI for $DN^2S(2)$; $DN^2I(4)$; $DSI(4,5)$; $DSN(4)$; $DN^2N(4)$; $DSZ(4,5)$; $DN^2Z(4)$.
- 6) order ABCDGEFHI for AZS(2).
- 7) order ABCDFEGHI for $DN^2Z(3)$.
- 8) order ABCDEFGHI for ASV1; AZS(1); ASZ(2-5); AS1(4); DS1(3); ASN(2,4); DSN(3); $DN^2N(5)$; ANS(1,2); DSZ(3); $DN^2Z(5)$.
- 9) order ABCDEFGHI (considered by Bertz and by us to be the “normal” ordering) for the following most numerous triplet indices: DSV(1,2); ASZ(1,3,5); $DN^2S(1,3-5)$; $DN^2S(3-5)$; $DN^2I(1-3)$; AS1(1,3,5); DS1(1,2); ASN(1,3,5); DSN(1,2); $DN^2N(1-3)$; DSZ(1,2); $DN^2Z(1,2)$.

From the last group of triplet indices, it was shown above that six of them, namely $DN^2Z(2)$; $DN^2N(1,2)$; $DN^2I(1-3)$, led to the same ordering of octanes as index J. Starting with the 35 nonanes, the ordering induced by triplet indices differed from that induced by J in at least one pair of nonane isomers.

Conclusions

It was shown that the ordering of constitutional isomers of alkanes with six to eight carbon atoms could serve for observing relatedness between topological indices (TIs). Starting from ideas first advocated by Bertz, we compared the ordering of alkane constitutional isomers with up to nine carbon atoms with that induced by index J. It was found that the ordering of alkane isomers allowed the grouping of TIs into a few classes, and that numerous triplet TIs paralleled the ordering induced by index J up to octanes.

Acknowledgement. We thank Dr. C. Rücker for his useful comments on the manuscript.

References

1. Devillers, J.; Balaban, A. T. (editors), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, Reading, UK, 1999.
2. Balaban, A. T. QSAR and computational methods in drug discovery in *Encyclopedia of Analytical Chemistry* (R. A. Meyers, Ed.), Wiley, Chichester, 2000, vol. 8, pp. 7288-7311.
3. Balaban, A. T. A personal view about topological indices for QSAR/QSPR, in *QSAR/QSPR Studies by Molecular Descriptors* (M. Diudea, Ed.), Huntington, New York, 2000, pp. 9-37.
4. Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, 2000.
5. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH, New York, 2000 (vol. 11 in the Series of Methods and Principles of Medicinal Chemistry, Eds. R. Mannhold, H. Kubinyi & H. Timmerman).
6. Randic, M. Topological Indices, in *Encyclopedia of Computational Chemistry* (P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III and P.R. Schreiner, Eds.), Wiley, Chichester, 1998, pp. 3018-3032.
7. Grassy, G; Calas, B.; Yasri, A.; Lahana, R.; Woo, R.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immunosuppressive compounds, *Nature Biotechnol.* 1998, **16**, 748-752.
8. Bermudez, C.; Daza, E. E.; Andrade, E. Characterization and comparison of *Escherichia coli* transfer RNA by graph theory based on secondary structure, *J. theor. Biol.* 1999, **197**, 193-205.
9. Randic, M. On computation of optimal parameters for multivariate analysis of structure-property relationships. *J. Comput. Chem.* 1991, **12**, 970-980.
10. Rouvray, D. H. A rationale for the topological approach to chemistry, *J. Mol. Struct. (Theochem)*, 1995, **336**, 101-114.
11. Randic, M. On characterization of molecular branching, *J. Am. Chem. Soc.* 1975, **97**, 6609-6615.

12. Kier, L.B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
13. Kier, L.B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*, Wiley, London, 1986.
14. Kier, L.B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*, Academic Press, San Diego, 1999.
15. Bonchev, D; Trinajstic, N. Information theory, distance matrix, and molecular branching, *J. Chem. Phys.* 1977, **67**, 4517-4533.
16. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*. Research Studies Press - Wiley, Chichester, 1983.
17. Bonchev, D. Overall connectivities/topological complexities: a new powerful tool for QSAR/QSPR. *J. Chem. Inf. Comput. Sci.* 2000, **40**, 934-941.
18. Bonchev, D. The overall Wiener index – a new tool for characterization of molecular topology. *J. Chem. Inf. Comput. Sci.* 2001, **41**, 582-592.
19. Balaban, A.T.; Mekenyan, O.; Bonchev, D. Unique description of chemical structures based on hierarchically ordered extended connectivities (HOC procedures). I. Algorithms for finding graph orbits and canonical numbering of atoms. (Following papers in the series are in the same issue). *J. Comput. Chem.* 1985, **6**, 538-551.
20. Mekenyan, O.; Bonchev, D.; Balaban, A. T. Unique description of chemical structures based on hierarchically ordered extended connectivities (HOC procedures). V. New topological indices, ordering of graphs, and recognition of graph similarity. *J. Comput. Chem.* 1984, **5**, 629-639.
21. Robinson, R.W.; Harary, F.; Balaban, A. T. The number of chiral and achiral alkanes and mono-substituted alkanes. *Tetrahedron*, 1976, **32**, 355-361.
22. Read, R. C. The enumeration of acyclic chemical compounds, in *Chemical Applications of Graph Theory*, (A. T. Balaban, ed.), Academic Press, 1976, pp. 25-61.

23. Motoc, I.; Balaban, A.T. Topological indices: intercorrelations, physical meaning, correlational ability. *Rev. Roum. Chim.* 1981, **26**, 593-600.
24. Basak, S.C.; Balaban, A.T.; Grunwald, G. D.; Gute, B.D. Topological indices: their nature and mutual relatedness. *J. Chem. Inf. Comput. Sci.* 2000, **40**, 891-898.
25. Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D; Balaban, A.T. Investigation of alkane branching (and resulting partial ordering) by topological indices. *Math. Chem. (MATCH)* 2000, **42**, 155-180.
26. Balaban, A.T.; Mills, D.; Basak, S. C. Interrelationship of topological indices evidenced by clustering *J. Chem. Inf. Comput. Sci.* 2002 (to be published).
27. The history of alkane isomer enumeration is presented with extensive bibliography in two chapters of *Chemical Applications of Graph Theory* (A. T. Balaban, ed.), Academic Press, London, 1976.
28. Polya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*. Springer, Berlin, 1987 (translated from Polya's original paper published 50 years earlier in German in *Acta Math.* 1937, **68**, 145, with Read's comments and up-dated literature survey).
28. Blair, C. M.; Henze, H.R. The number of structurally isomeric alcohols of the methanol series, *J. Am. Chem. Soc.* 1931, **53**, 3042-3046.
29. Balaban, A. T. Chemical graphs. XXXIV. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* 1979, **53**, 355-375.
30. Bertz, S. H. Branching in graphs and molecules, *Discr. Appl. Math* 1988, **19**, 65-83.
31. Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* 1982, **80**, 399-404.
32. Balaban, A.T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* 1983, **55**, 199-206.
33. Balaban, A.T.; Motoc, I. Chemical Graphs. XXXVI. Correlations between octane numbers and topological indices of alkanes. *Math Chem (MATCH)*, 1979, **5**, 197-218.

34. Balaban, A.T.; Kier, L. B.; Joshi, N. *Math. Chem. (MATCH)*, 1992, **28**,13-27: Structure-property analysis of octane numbers for hydrocarbons (alkanes, cycloalkanes, alkenes).
35. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* 1947, **69**, 17-20.
36. Basak, S. C.; Harriss, D. C.; Magnuson, V. R. POLLY 2.3, University of Minnesota: Duluth, MN, 1998.
37. Wiener, N. *Cybernetics*. Wiley, New York, 1948.
38. Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V.R., Neighborhood complexities and symmetry of chemical graphs and their biological applications. In, *Mathematical Modelling in Science and Technology*. 4th Internat. Conf. Zurich. (X. J. R. Avula, R. E. Kalman, A. I. Liapis, and E. Y. Rodin, Eds.). Pergamon Press, New York, pp. 745-750, 1983.
39. Basak, S. C. Use of molecular complexity indices in predictive pharmacology and toxicology: A QSAR approach. *Med. Sci. Res.*, 1987, **15**, 605-609.
40. Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R. Determining structural similarity of chemicals using graph-theoretic indices. *Discr.Appl. Math.* 1988, **19**, 17-44.
41. Raychaudhury, C.; Ray, S. K.; Roy, R.; Ghosh, J. J.; Basak, S. C. Discrimination of isomeric structures using information theoretic indices. *J. Comput. Chem.* 1984, **5**, 581-588.
42. Balaban, A.T. Chemical graphs. 48. Topological index J for heteroatom-containing molecules taking into account periodicities of element properties. *Math. Chem. (MATCH)* 1986, **21**, 115-122.
43. Balaban, A. T.; Filip, P. , Computer program for topological index J (average distance sum connectivity). *Math. Chem. (MATCH)* 1984, **16**, 163-190.
44. Filip, P. A.; Balaban, T. S.; Balaban, A.T. A new approach for devising local graph invariants: Derived topological indices with low degeneracy and good correlational ability. *J. Math. Chem.* 1987, **1**, 61-83.
45. Balaban, A. T.; Balaban, T. S. New vertex invariants and topological indices of chemical graphs based on information on distances. *J. Chem. Inf. Comput. Sci.* 1991, **31**, 383-397.

46. Molconn-Z. Hall Associates Consulting, Quincy, MA.
47. Randic, M. On structural ordering and branching of acyclic saturated hydrocarbons. *J. Math. Chem.* 1998, **24**, 345-358.