Topological Organic Chemistry. 10. Graph Theory and Topological Indices of Conformational Isomers

Harry P. Schultz*,†

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

Emily B. Schultz[‡] and Tor P. Schultz[§]

Department of Forestry and Forest Products Laboratory, Mississippi State University, Mississippi State, Mississippi 39762

Received February 27, 1996[⊗]

Distance matrices constructed from valence and vertex weighted graphs were the sources of modified information that recognized and quantified conformational isomerism exhibited by linear organic molecules. An algorithm is presented that defines the relative complexities of conformers.

INTRODUCTION

The prior paper¹ of this series described experiments wherein operations upon valence and vertex weighted distance matrices yielded topological indices that differentiated between geometric isomers. This present paper utilizes the same techniques used with geometrically modified topological indices to recognize and describe conformational isomerism within linear molecules. Paramount among researchers who have investigated the subtle territory of linear rotamer or conformational isomerism are Randić, Kleiner, and DeAlba.² Others who have directed attention to the topic of conformational analysis include Fella et al.,³ Tomczak,⁴ and Smellie et al.⁵ The superb stereochemistry compendium by Eliel et al.⁶ examines conformational isomerism in extensive detail.

This paper records the results of experiments directed to recognizing and quantifying conformational isomerism in linear homonuclear (carbon) and heteronuclear organic molecules. Molecular topological indices described in detail in the prior papers of this series—the *S* index, product of row sums, determinant, permanent, and long hafnian—are the vehicles of transmission for the conformational information. Thus, topographical ends were sought with topological means.

The topic was studied in two sections. The first set of experiments was directed to assessing the influences of simple substituents on the values of conformationally modified indices of substituted ethanes. The second portion of this report examined the C_4 through C_7 alkane conformational isomers so penetratingly studied and described in Randić et al.²

COMPUTATIONS

Molecular graphs, with hydrogen atoms suppressed and with interatomic edge counts set at unity for all atoms, were derived from the representative organic compounds listed in Tables 1 and 2. The IUPAC system of numbering and nomenclature was used in all experiments; all results were

invariant, hence descriptors, and were confirmed by use of the CIP numbering system.

Valence (v) and vertex (V) weighted graphs and distance (\mathbf{D}) matrices were utilized for all experiments, as described in detail in the preceding article¹ of this series.

The V weighted graphs differentiated homonuclear from heteronuclear vertices within graphs.

V = 1 + (atomic number heteroatom -

atomic number carbon)

V = 1 + (atomic number heteroatom - 6)

$$V = C, 1; N, 2; O, 3; F, 4$$

v = edge count about each hydrogen-suppressed vertex

with additionally, each heteroatom unshared electron pair counted as one valence. Such ν weighted graphs provided information about the bond status (single, double, triple) of each edge in a molecule as well as adding to information that distinguished homonuclear from heteronuclear vertices.

The practice of projecting the conformational isomers onto a graphitic lattice described by Randić et al.² was used in this work. The descriptions of rotamer status, such as synperiplanar (syn, sp) and antiperiplanar (anti, ap), were all supplanted with the succinct and unique symbols of C and T, conforming to the usages and drawings of Randić et al.²

The definitions of the various topological indices studied in this paper were detailed in the prior paper¹ of this sequence and are briefly summarized here. The S index is the sum of the product of the valence vector and the distance matrix of order N.¹

$$S = \sum_{i=1}^{N} (\mathbf{v}\mathbf{D})_i$$

Pettofrezzo 7 defined the determinant (det) of a square matrix of order N as

$$\det = (-1)^k a_{1i_1} a_{2i_2} ... a_{Ni_N}$$

The permanent (per), sometimes referred to as the positive

[†] Present address: P.O. Box 262, Big Horn, WY 82833.

[‡] Department of Forestry.

[§] Forest Products Laboratory.

[⊗] Abstract published in Advance ACS Abstracts, July 1, 1996.

Table 1. Descriptors of Selected Substituted Ethanes Derived from the Vertex and Valence Weighted Distance Matrices and Modified to Represent the Individual syn (C) and anti (T) Conformations of Each Graph

		index $\mathbf{D}_N^N(G_{V_{UW}})$						
no.	compound	RMN ^a	S	PRS	\det^b	per	1haf	
1	butane		44	2 304	-48	256	104	
	C	+26	70	2 330	74	282	130	
	T	-26	18	2 278	22	230	78	
2	1,3-butadiene		120	20 736	-432	2 304	936	
	C	+42	162	20 778	474	2 346	978	
	T	-42	78	20 694	390	2 262	894	
3	1,2-diaminoethane		128	36 864	-768	4 096	1 664	
	C	+44	172	36 908	812	4 140	1 708	
	T	-44	84	36 820	724	4 052	1 620	
4	1,2-diaminiumethane ion		56	9 216	-192	1 024	416	
	C	+32	88	9 248	224	1 056	448	
	T	-32	24	9 184	160	992	384	
5	1,2-ethanedialdiimine		288	186 624	-3888	20 736	8 424	
	C	+66	354	186 690	3 954	20 802	8 490	
	T	-66	222	186 558	3 822	20 670	8 358	
6	1,2-ethanediol		356	186 624	-3888	20 736	8 424	
	C	+74	430	186 698	3 962	20 810	8 498	
	T	-74	282	186 550	3 812	20 662	8 350	
7	1,2-ethanedial		648	746 496	-15552	82 944	33 696	
	C	+102	750	746 598	15 654	83 046	33 798	
	T	-102	546	746 394	15 450	82 842	33 594	
8	oxalic acid		1874	133 885 737 216	-14929920	2 058 089 472	159 376 896	
	C	+190	2064	133 885 737 406	14 930 110	2 058 089 662	159 377 086	
	T	-190	1684	133 885 737 026	14 929 730	2 058 089 282	159 376 706	
9	1,2-difluoroethane		800	589 824	-12288	65 536	26 624	
	C	+116	916	589 940	12 404	65 652	26 740	
	T	-116	684	589 708	12 172	65 420	26 508	

^a Rotamer modification value. ^b Absolute values used in calculation of modified indices.

Table 2. Descriptors Derived from the Vertex and Valence Weighted Distance Matrices of Linear C₄-C₇ Alkanes Modified To Represent the Various Syn (C) and Anti (T) Conformations of Each Graph

		RMV^a	index $\mathbf{D}_N^N(G_{V_{UW}})$					
no.	compound		S	PRS	\det^b	per	1haf	rel potential energy, kcal/mo
1	butane		44	2 304	-48	256	104	
	C	+26	70	2 330	74	282	130	6.1
	T	-26	18	2 278	22	230	78	0
2	pentane		100	235 200	256	9 216	1 448	
	CC	+88	188	235 288	344	9 304	1 536	12.0
	CT	0	100	235 200	256	9 216	1 448	6.0
	TT	-88	12	235 112	168	9 128	1 360	0
3	hexane		190	35 283 600	-1280	549 952	49 760	
	CCC	+202	392	35 283 802	1482	550 154	49 962	
	CTC	+78	268	35 283 678	1358	550 030	49 838	12.0
	TCC	+62	252	35 283 662	1342	550 014	49 822	11.9
	TTC	-62	128	35 283 538	1118	549 890	49 698	6.0
	TCT	-78	112	35 283 522	1102	549 874	49 682	5.9
	TTT	-202	-12	35 283 398	1078	549 750	49 558	0
4	heptane		322	7 326 498 816	6144	42 720 256	1 000 000	
	CĈCC	+384	706	7 326 499 200	6528	42 720 640	1 000 384	
	CTCC	+208	530	7 326 499 024	6352	42 720 464	1 000 208	17.9
	TCCC	+176	498	7 326 498 992	6320	42 720 432	1 000 176	
	CTTC	+32	354	7 326 498 848	6176	42 720 288	1 000 032	12.0
	TTCC	0	322	7 326 498 816	6144	42 720 256	1 000 000	11.9
	TCTC	0	322	7 326 498 816	6144	42 720 256	1 000 000	11.9
	TCCT	-32	290	7 326 498 784	6112	42 720 224	999 968	11.8
	TTTC	-176	146	7 326 498 640	5968	42 720 080	999 824	6.0
	TTCT	-208	114	7 326 498 608	5936	42 719 048	999 792	5.9
	TTTT	-384	-62	7 326 498 432	5760	42 719 872	999 616	0

^a Rotamer modification value. ^b Absolute values used for calculation of modified indices.

determinant, was described in detail by Minc.8

$$per = a_{1i_1} a_{2i_2} ... a_{Ni_N}$$

The product of the row sums (PRS) was an early, rough approximation to the permanent. Caianiello9 described the hafnian (1haf) as the summation over all permutations i_1 , i_2 , ..., i_{2N} which satisfies the limitations $i_1 < i_2$, $i_3 < i_4$... $i_{2N-1} < i_2$ i_{2N} ; $i_1 < i_3 < i_5 ... < i_{2N-1}$. These five indices were the carriers of the conformational information presented in our paper.

The geometries about a pair of single bonded vertices responsible for the conformational isomerism were indicated by rotamer factors (RF) of +1, +1 for C geometry; -1, -1for T geometry. All other vertices of the graphs were

Randić et al.,2 increasing sequences molecular ID numbers ϕ Values this paper, decreasing sequences all index values C_5 C_6 C_7 C_5 C_6 C_7 C_4 C_6 C_7 C T CC CCCC CC**CCC** \mathbf{C} CC CCC C **CCC** CTCT TCC CTCC CT TCC TCCC CTC CTCC TT CTC TCCC TT CTC TCC TCCC **CTCC** TCT **TCCT** TTT TCCT TTC CTTC TTCC TCT TTC TCT TTCC TTCC TTT CTTC TTC CTTC TTT TCTC **TCTC** TCTC TCCT TTCT TTCT TTTC TTTC TTTC TTCT TTTT TTTT TTTT

Table 3. Conformational Sequence Comparisons of Linear C₄-C₇ Alkanes as Projected onto a Graphite Lattice

assigned RF values of zero. These factors were used to calculate a rotamer modification number (RMN) which was then incorporated into a topological index, using the same principles as described in the previous paper.¹

A specific example is pictured using 1,2-difluoroethane (Table 1, 9).

The Vv weighted **D** matrix was constructed

$$\mathbf{D}_{4}^{4}(G_{Vvw}) = \begin{bmatrix} 0 & 2 & 2 & 4 \\ 2 & 0 & 4 & 2 \\ 16 & 32 & 0 & 48 \\ 32 & 16 & 48 & 0 \end{bmatrix}$$

which was then transposed

$$\mathbf{D}_{4}^{4}(G_{Vvw})^{\mathrm{T}} = \begin{bmatrix} 0 & 2 & 16 & 32 \\ 2 & 0 & 32 & 16 \\ 2 & 4 & 0 & 48 \\ 4 & 2 & 48 & 0 \end{bmatrix}$$

and the two matrices were summed, as explained, and justified in the preceding publication. In heteronuclear graphs the greater weights of the heteronuclear vertices placed the bulk of the information in rows outside of the conformational rows, as seen in rows no. 3 and no. 4 of the Vv weighted \mathbf{D} matrix of 1,2-difluoroethane. The sum of the \mathbf{D} and \mathbf{D}^{T} matrices provided relatively more information in rows no. 1 and no. 2, the conformational centers of the compound. The technique also provided a more equitable comparison between homonuclear and heteronuclear graphs.

$$\mathbf{D}_{4}^{4}(G_{Vvw}) + \mathbf{D}_{4}^{4}(G_{Vvw})^{\mathrm{T}} = \begin{bmatrix} 0 & 4 & 18 & 36 \\ 4 & 0 & 36 & 18 \\ 18 & 36 & 0 & 96 \\ 36 & 18 & 96 & 0 \end{bmatrix}$$

The RF values for the C and T conformations of 1,2-

difluoroethane were built into two diagonal matrices.

Premultiplication of the $\mathbf{D} + \mathbf{D}^{T}$ matrices with the diagonal RF matrices gave the information for the rotamer modified numbers (RMN) of the two conformers of 1,2-difluoroethane.

$$\mathbf{D}_{4}^{4}(RMN, C) = \begin{bmatrix} 0 & 4 & 18 & 36 \\ 4 & 0 & 36 & 18 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{D}_{4}^{4}(RMN, T) = \begin{bmatrix} 0 & -4 & -18 & -36 \\ -4 & 0 & -36 & -18 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The content of each matrix was summed. The rotamer modification number (RMN) for the S index of the C rotamer of 1,2-difluoroethane was +116; that for the T conformer was -116. These modifications were added to the S index value (800) of 1,2-difluoroethane to give the C conformer modified value of 916 and the T modified value of 684 for the S indices of the two rotamers of 1,2-difluoroethane. The other topological indices were calculated in the same fashion. All five molecular topological indices reported in this paper exhibited the same arrangements, but the larger index values possessed a lower ratio of response to the rotamer modification values.

All the substituted ethane derivatives of Table 1 have only one single, central bond about which rotamer modification exists. In the instance of larger molecules, such as pentane (Table 2, 2), the various paired centers of conformational isomerism were algebraically summed to arrive at RF values for each vertex. For example, the CC conformation of pentane displays a set of rotamer factor values (RFV) calculated as follows.

The all-C conformations of all normal C_4 – C_7 alkanes were included in this study, even though in the larger molecules

Table 4. Compaction Indices $(CI)^a$ for All Conformers of Normal C_4 – C_7 Alkanes Superimposed onto Graphitic (G) and Pentagonal (P) Lattices

со	nformer	G	P	
	С	0.88	1.18	
	T	0.25	0.31	
C_5	CC	2.01	3.44	
	CT	0.58	0.73	
	TT	0.29	0.38	
C_6	CCC	5.22		
	TCC	1.59	2.12	
	TTC	0.94	1.17	
	CTC	0.67	0.86	
	TCT	0.42	0.53	
	TTT	0.28	0.35	
C_7	CCCC			
	TCCC	3.42		
	TTCC	2.03	2.34	
	CTTC	1.72	2.27	
	TCCT	1.68	2.42	
	CTCC	1.15	1.55	
	TTCT	0.87	1.09	
	TTTC	0.71	0.85	
	TCTC	0.54	0.65	
	TTTT	0.29	0.38	

 $^{\it a}\,{\rm CI}={\rm sum}$ vertical distances of vertices from baseline/baseline length ${\rm C_1-C_N}.$

whose path graphs were imbedded in the honeycomb lattice the terminal vertices would curl around and superimpose on the first carbon atoms of C_6 and larger compounds. Although structurally meaningless, the mathematics involved is rational, and in such instances one could in a practical vein imagine a very tight spiral. The conformers were also pictured by imprinting each rotamer on a regular pentagon lattice; the interior angles of 108° of the pentagon are very nearly of the same value as the normal tetrahedral valence angle of 109.5° . Randić et al.² have pictured on pages 278 and 280 of their presentation, conformations of the C_4-C_7 normal alkanes superimposed on graphitic lattices.

RESULTS

Table 1 lists a series of related ethane derivatives, with the no. 1 and no. 2 vertices successively substituted with C, N, O, and F atoms, at increasingly higher carbon oxidation levels for N and O. Note the different index values exhibited for the two congeneric graphs of butane and 1,2-diaminiumethane ion (Table 1; 1, 4) as well as the degenerate PRS and long hafnian values for glyoxal diimine and ethylene glycol (Table 1; 5, 6). No regard was given to the additional geometric isomerism extant in the trigonal hybridized imino groups of 1,2-ethanedialdiimine (Table 1, 5).

Because, as Randić et al.² observed, most thermodynamic physicochemical properties belong to a mixture of conformations, not to a single rotamer, the order of progression of the index values for the various alkane rotamers listed in Table 2 was contrasted with the data of Randić et al.,² the latter being the authoritative basis for the comparisons. Table 3 shows the sequence relations, from all-C to all-T rotamers, for the two Randić sets of values,² molecular ID numbers, and the folding (ϕ) values, compared with the conformational sequences arrived at in this study.

As expected, all sequences were the same for butane and pentane but differed for hexane and heptane in all sets of comparisons. These latter series were similar in that both sequences, with one exception—the TTT location in the Randić ϕ series for hexane—subdivide each progression into

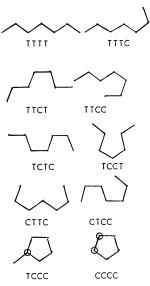


Figure 1. Possible conformations of seven-vertex chains superimposed on a pentagonal lattice. See also Figure 2, ref 2.

about equal thirds—mostly C bonds, equal numbers of C and T bonds, and mostly T bonds. The Randić numbers, however, unlike one of this paper's sequences (TTCC = TCTC, for heptane), have no degenerate values. It is interesting to observe that the identical topological index values reported in this paper for the heptane rotamers TTCC and TCTC have one each of interior and exterior T and C sets of values but that the compaction indices (CI, vide infra) are different, indicating dissimilarity between the two conformers.

In order to ascertain the degree of similarity/dissimilarity among the rotamers, a compaction index (CI) was devised. The CI index was graphically (10 cm = one unit length)determined by dividing the distance (the baseline) between the two termini of a conformer into the sum of the vertical distances from the baseline, extended if necessary, of all vertices of a graph. The lower the CI quotient, the less puckered was the conformer, with all-T conformers having the lowest CI values and all-C rotamers the highest CI numbers. Graphs with similar quotients were conformationally more similar to one another than graphs that possessed widely divergent CI quotients. Table 4 summarizes these data, and also includes data obtained from graphs superimposed onto pentagonal lattices. As expected, the pentagonal lattice CCC rotamer of hexane and the TCCC and CCCC rotamers of heptane fold the initial and terminal vertices together, and hence their compaction indices are meaningless. In the instance of the heptane, TCCT, CTTC, and TTCC conformers, a different sequence of CI numbers was observed between values derived from the graphitic and pentagonal lattices. Although the same general similarity/ dissimilarity comparisons were observed as were pictured in the Randić Figures 5 and 6, there was not an exact parallelism with the data printed in the Randić Table 7. Additionally, there was only modest agreement between sequences pictured in Tables 3 and 4 of this paper.

As a final experiment the rotamer modified topological indices were compared with the rotational energy barriers of the various conformers studied. From a plethora of data, Eliel et al.⁶ designate the syn/anti energy difference about the C2–C3 bond of butane as lying between 4.5–6.1 kcal/mol. The data, gathered from a variety of spectrographic, molecular mechanics, and ab initio sources, were focused mainly on the methyl—methyl interactions about the middle

bond of butane. To more completely utilize such data the work of Wiberg and Murcko¹⁰ was additionally searched and averaged to provide from one source all the energy data (kcal/mol) used for these syn/anti group interaction calculations. The following values (kcal/mol) were used in this study: methyl-methyl, 6.1; methylene-methylene, 5.9 (about the hexane C3—C4 eclipsed bond); methylene-methyl, 6.0 (by interpolation of the above values). The calculated syn/anti relative potential energy data for the various conformers of butane, pentane, hexane, and heptane are listed in Table 2.

Figure 1 pictures possible conformations of heptane; stepwise removals of its terminal vertices progressively form, in turn, the rotamers of hexane, pentane, and butane. The terminal vertex overlapping of the CCC rotamer of hexane and the CCCC and TCCC conformers of heptane afford meaningless information, hence their energy data are omitted from Table 2.

Table 2 shows the sequence parallelism between rotamer modified indices and relative potential energy contents of all conformers of butane, pentane, hexane, and heptane, including the identical energy values for the TTCC and TCTC rotamers of heptane, the two rotamers that presented degenerate modified topological indices.

A quantitative assessment of the relationship between the modified index value and the energy content of each conformer was attempted. The *S* index was examined, for it was the easiest value to arrive at; the determinant index was analyzed, because in past studies of this series it generally gave the best statistical results.

First, all the collective data from all points reflected by butane through heptane rotamers were studied; no relationship was observed in the instances of using modified *S* or determinant index data.

Second, the hexane and heptane sets of rotamers were separately examined. (Too few data points were available in the smaller butane and pentane portions of data.)

S (D) Index (Table 2, column 4)

hexane

$$E$$
, kcal/mol = $0.6404 + 0.04358(S)$

$$n = 5$$
 $r^2 = 0.996$ $s = 0.3579$ $F = 778.49$

heptane

$$E$$
, kcal/mol = $2.0816 + 0.03013(S)$

$$n = 8$$
 $r^2 = 0.990$ $s = 0.5765$ $F = 622.32$

det (D) Index (Table 2, column 4)

hexane

no fit

heptane

$$E, \text{kcal/mol} = -173.35 + 0.03013(\text{det})$$

$$n = 8$$
 $r^2 = 0.990$ $s = 0.5765$ $F = 622.32$

The heptane *S* and determinant modified index comparisons gave similar equations and the same statistical data.

These results are tempered by the caveat that some of the conformers would have had 1,5 terminal carbon atom repulsions, as seen in the rotamers of CC pentane, TCC hexane, and TTCC and TCCT heptane, although calculations by Eliel et al.⁶ indicate less than 0.4% of the high energy g⁺g⁻ rotamer of pentane is present in its conformational equilibrium. When, however, in spite of this circumstance, the Eliel et al.⁶ range of 3.3–3.7 kcal/mol was included to reflect the energy of the 1,5 dimethyl repulsions, the hexane and heptane conformational energy sequences were thrown completely awry. This was to be expected, for each modified topological index represented only the conformational status along the bonds of the alkane chain and not across its nonbonded ends.

CONCLUSIONS

A technique has been presented that identifies and quantifies the molecular topological indices of the various conformational isomers of the C_4 – C_7 normal alkanes. The ease of execution of the procedures described in this paper presents an interesting mode of applying topological techniques to attain topographical ends.

ACKNOWLEDGMENT

The authors thank the referees for constructive and positive comments about this paper.

REFERENCES AND NOTES

- Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry.
 Graph Theory and Molecular Topological Indices of Stereoisomeric Compounds. J. Chem. Inf. Comput. Sci. 1995, 35, 864–870.
- (2) Randić, M.; Kleiner, A. F.; DeAlbe, L. M. Distance/Distance Matrices. J. Chem. Inf. Comput. Sci. 1994, 34, 277–286.
- (3) Fella, A. L.; Nourse, J. G.; Smith, D. H. Conformation Specification of Chemical Structures in Computer Programs. *J. Chem. Inf. Comput. Sci.* 1983, 23, 43–47.
- (4) Tomczak, J. An Explicit Representation of Molecular Geometry and Topology for Small Molecules. J. Chem. Inf. Comput. Sci. 1995, 35, 199–202.
- (5) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage. *J. Chem. Inf. Comput. Sci.* 1995, 35, 285–294.
- (6) Eliel, E. L.; Wilen, S. H.; Mander, L. N. Stereochemistry of Organic Compounds; Wiley-Interscience: New York, 1994; pp 597–834.
- (7) Pettofrezzo, A. J. *Matrices and Transformations*; Prentice-Hall: Englewood Cliffs, NJ, 1966; pp 13–29.
- (8) Minc, H. Permanents. Encyclopedia of Mathematics and Its Applications; Rota, G.-C.; Ed.; Addison-Wesley: Reading, MA, 1978; Vol. 6, pp 1–119.
- (9) Caianiello, E. R.; Proprietà Pfaffiani e Hafniani. (Properties of Pfaffians and Hafnians). Recerca, Napoli 1956, 7, 25–31.
- (10) Wiberg, B. K.; Murcko, A. M. Rotational Barriers. 2. Energies of Alkane Rotamers. An Examination of Gauch Interactions. *J. Am. Chem. Soc.* 1988, 110, 8029–8038.

CI960027V