

# Topological Organic Chemistry. 9. Graph Theory and Molecular Topological Indices of Stereoisomeric Organic Compounds

Harry P. Schultz<sup>\*,†</sup>

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

Emily B. Schultz<sup>‡</sup> and Tor P. Schultz<sup>§</sup>

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

Received March 29, 1995<sup>®</sup>

Vertex and valence weighted graphs and distance matrices of homonuclear and heteronuclear organic substances containing stereocenters were initial sources of steric information used to modify topological indices reported in earlier papers of this series. The modifications recognized and quantified the presence of chiral and geometrical stereocenters in the molecules.

## INTRODUCTION

Prior papers of this series<sup>1</sup> reported the use of distance (**D**) matrices as sources of molecular topological indices for a variety of homonuclear (carbon) and heteronuclear organic compounds. Although numerous authors referred to in past papers of this sequence have gathered references, reviewed papers, and created new avenues of approach to all facets of the above topics, few have been concerned with the stereochemistry of organic substances. Shelly and Munk<sup>2</sup> and Barnard et al.<sup>3</sup> and their included references overviewed the subject to 1989. Since then some seminal papers on the theme have been published by Akutsu,<sup>4</sup> Razinger et al.,<sup>5</sup> Agarwal and Gelernter,<sup>6</sup> Mata and Lobo,<sup>7</sup> and Pogliani.<sup>8</sup> Most of these have been concerned with the sophisticated objective of devising algorithms for total molecular notations of molecules possessing stereocenters.

This paper reports the results of experiments directed toward means to represent the kinds of stereocenters present in homonuclear and heteronuclear organic molecules as well as to recognize the influences of the stereocenters on the values of the molecular topological indices examined in detail in prior papers of this set<sup>1</sup>—MTI', product of row sums, determinant, permanent, and long hafnian.

These experiments were focused on two broad facets of stereochemistry. First: the isomerism that exists as a consequence of restricted rotation in appropriately substituted cyclic systems, including cycloethane (ethene). The isomerism that obtains in such systems is referred to in this article as geometrical isomerism. Second: the isomerism which is due to the presence of stereocenter(s) in some organic compounds—centers that may confer "handedness" or chirality upon molecules. Such substances are designated as optical isomers in this paper.

## COMPUTATIONS

In this presentation the order of importance of the topics is as follows: first, total structure which includes all vertices

of a graph such as has been repeatedly described in prior papers<sup>1</sup> of this series; second, geometrical isomerism which focuses upon the at least two substituted vertices responsible for reflecting isomerism due to restricted rotation within a molecule; third, optical isomerism which stems from chirality resident in an appropriately substituted stereocenter.

Molecular (multi)graphs, with hydrogen atoms suppressed and with interatomic carbon–carbon edge counts set at unity for tetrahedral and two for trigonal hybridized atoms, were derived from the representative organic compounds listed in Tables 1–4. The IUPAC system of numbering and nomenclature was used in all descriptions and experiments; the results, all invariant, hence descriptors, were also arrived at by use of the CIP numbering system; this and other related topics are presented in the splendid compendium on stereochemistry by Eliel et al.<sup>9</sup>

Although experiments with various combinations of edge (*e*), vertex (*V*), and valence (*v*) weighted graphs and **D** matrices were performed as described in the preceding paper<sup>1h</sup> of this series, only *V* and *v* weighted molecular graphs and **D** matrices were utilized for all computations reported in this paper. The *V* weighted graphs distinguished between homonuclear and heteronuclear graphs.

$$V = 1 + (\text{atomic number heteroatom} - \text{atomic number carbon})$$

$$V = 1 + (\text{atomic number heteroatom} - 6)$$

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

The *v* weighted graphs identified graphs with ionic units in the molecule, discriminating, for example, between the neutral and zwitterion forms of amino acids (Table 3; 11, 12).

*v* = edge count about each hydrogen-suppressed vertex. Additionally, each heteroatom unshared electron pair counted as one valence.

Past papers<sup>1</sup> of this sequence defined the various matrix functions studied in this article—MTI', product of the row sums (PRS), determinant (|det|), permanent (per), and long hafnian (1haf).

<sup>†</sup> Present address: POB 262, Big Horn, WY 82833.

<sup>‡</sup> Department of Forestry.

<sup>§</sup> Forest Products Laboratory.

<sup>®</sup> Abstract published in *Advance ACS Abstracts*, July 1, 1995.

**Table 1.** Descriptors Derived from the Vertex and Valence Weighted Distance Matrices of Representative Geometric Isomers and Modified for Geometric Isomerism

no.	compound	GMV <sup>a</sup>	index $D_N^N(G_{V_{VW}})$				
			S	PRS	det	per	haf
1	2-butene		124	20 736	-288	2 916	1 314
	(Z)-	+56	180	20 792	344	2 972	1 370
	(E)-	-56	68	20 680	232	2 860	1 258
2	2-pentene		223	2 021 760	1 440	82 944	9 348
	(Z)-	+84	307	2 021 844	1 524	83 028	9 432
	(E)-	-84	139	2 021 676	1 356	82 860	9 264
3	1,2-difluoroethene		1 132	5 308 416	-73 728	746 496	336 384
	(Z)-	+176	1 308	5 308 592	73 904	746 672	336 560
	(E)-	-176	956	5 308 240	73 552	746 320	336 208
4	2-butanone, enol tautomer		485	8 164 800	8 640	331 776	31 524
	(Z)-	+113	598	8 164 913	8 753	331 889	31 637
	(E)-	-113	372	8 164 687	8 527	331 663	31 411
5	3-penten-2-ol		716	1 200 225 600	-46 656	17 114 976	1 391 904
	(Z)-	+160	876	1 200 225 760	46 816	17 115 136	1 392 064
	(E)-	-160	556	1 200 225 440	46 496	17 114 816	1 391 744
6	1-ethyl-3-methylcyclopentane		508	464 950 886 400	-15 552	1 035 368 640	12 101 184
	cis-	+124	632	464 950 886 524	15 676	1 035 368 764	12 101 308
	trans-	-124	384	464 950 886 276	15 428	1 035 368 516	12 101 060
7	2,4-dimethylcyclopentanol		960	1 893 029 688 576	-104 976	4 387 588 560	49 239 576
	(1,2) (2,4) (1,4)						
	cis, cis, cis	+448	1 408	1 893 029 689 024	105 424	4 387 589 008	49 240 024
	cis, trans, trans	-172	788	1 893 029 688 404	104 804	4 387 588 388	49 239 383
	trans, cis, trans	-130	830	1 893 029 688 446	104 846	4 387 588 430	49 239 446
8	bicyclo[3.3.0]octane		814	1 893 029 688 430	104 830	4 387 588 414	49 239 430
	cis-	+112	550	685 425 033 216	-9 216	1 727 534 592	19 646 208
	trans-	-112	662	685 425 033 328	9 328	1 727 534 704	19 646 320
9	N-methylethanalimine		438	685 425 033 104	9 104	1 727 534 480	19 646 096
	syn-		262	55 296	-768	7 776	3 504
	anti-	+96	358	55 392	864	7 872	3 600
10	azomethane		166	55 200	672	7 680	3 408
	syn-	+136	400	147 456	-2 048	20 736	9 344
	anti-	-136	536	147 592	2 184	20 872	9 480
			264	147 320	1 912	20 600	9 208

<sup>a</sup> Geometric Modification Value.

Trinajstić et al.<sup>10</sup> and Gutman<sup>11</sup> described the MTI' index, which is derived from the **D** matrix and the valence (*v*) vector of a graph of *N* vertices. Gutman<sup>11</sup> also suggested that the device MTI' be replaced with the single symbol *S*. The authors agree.

$$MTI' = S = \sum_{i=1}^N (vD)_i$$

Matrix determinants as topological indices were described in the second paper<sup>1b</sup> of this series. Pettofrezzo<sup>12</sup> defined the value of the determinant of a square matrix **A** = ((*a<sub>ij</sub>*)) of order *N* as the sum of *N*! terms of the form

$$(-1)^k a_{1i_1} a_{2i_2} \dots a_{Ni_N}$$

each term holding only one element from each row and each column. The exponent *k* represents the number of interchanges of two elements necessary to place the subscripts in the order 1, 2, ..., *N*.

Matrix permanents and products of the row sums were both described in the fourth paper<sup>1d</sup> of this sequence. The definition of the permanent is similar to that for the determinant, except that all permanent values are positive (if all elements are positive); hence the  $(-1)^k$  element of the determinant expression is omitted. The product of the row sums was an early crude approximation to the permanent.

The hafnians were described in the fifth paper<sup>1e</sup> of this series as the summation over all permutations *i*<sub>1</sub>, *i*<sub>2</sub>, ..., *i*<sub>2*N*</sub> which satisfy the limitations *i*<sub>1</sub> < *i*<sub>2</sub>, *i*<sub>3</sub> < *i*<sub>4</sub> ... *i*<sub>2*N*-1</sub> < *i*<sub>2*N*</sub>; *i*<sub>1</sub> < *i*<sub>3</sub> < *i*<sub>5</sub> ... *i*<sub>2*N*-1</sub> for a *N* × *N* matrix.

## GEOMETRIC ISOMERS

After much experimentation it was concluded that the geometries about a pair of vertices responsible for geometric isomerism in appropriately substituted double bonded and cyclic systems would be indicated by geometric factors (GF) of +1 and +1 for *Z* (*cis*) geometry, and -1, -1 for *E* (*trans*) geometry. All other vertices in the graph were assigned GF values of zero. These factors were used to arrive at a geometric modification (GM) number which when incorporated into a topological index quantified the geometric isomerism of a given compound. The technique is illustrated with a specific example, 1,2-difluoroethene (Table 1;3).

1 3 4 2	CIP numbers
3 1 2 4	IUPAC numbers
$\begin{array}{c} \text{F} \\ \text{F} \end{array} \text{C} = \text{C} \begin{array}{c} \text{F} \\ \text{F} \end{array}$	
4 1 1 4	vertex ( <i>v</i> ) weights
4 3 3 4	valence ( <i>v</i> ) numbers
0 ±1 ±1 0	geometric factors (GF)
0 ±88 ±88 0	geometric modifiers (GM)—vide infra

The  $Vv$  weighted  $\mathbf{D}$  matrix was fashioned as explained in the prior paper<sup>1h</sup> of this sequence. The  $Vv$  weighted  $\mathbf{D}$  matrix for

$$\mathbf{D}_4(G_{Vv}) = \begin{bmatrix} 0 & 6 & 3 & 9 \\ 6 & 0 & 9 & 3 \\ 16 & 48 & 0 & 64 \\ 48 & 16 & 64 & 0 \end{bmatrix}$$

heteronuclear systems is usually unsymmetrical.

Early on in these trials it was appreciated that in homonuclear systems, such as 2-butene (Table 1;1), the geometric centers were usually the most heavily weighted vertices and brought the greatest amount of information to the  $\mathbf{D}$  matrix within the two rows that represented the stereocenters. In heteronuclear graphs, however, the greater weights of the heteronuclear vertices placed the bulk of the information in rows outside of the stereocenter rows, as pictured in rows 3 and 4 of the  $Vv$  weighted  $\mathbf{D}$  matrix of 1,2-difluoroethene.

Hence, in order to incorporate as much information as possible into rows representing the geometric centers as well as to treat all homonuclear and heteronuclear molecules alike, the weighted  $\mathbf{D}$  matrix was transposed, and then the transposition was added to the original  $\mathbf{D}$  matrix.

The transposed  $Vv$  weighted  $\mathbf{D}$  matrix of 1,2-difluoroethene is

$$\mathbf{D}_4(G_{Vv})^T = \begin{bmatrix} 0 & 6 & 16 & 48 \\ 6 & 0 & 48 & 16 \\ 3 & 9 & 0 & 64 \\ 9 & 3 & 64 & 0 \end{bmatrix}$$

and the sum of the original and transposed  $\mathbf{D}$  matrices is

$$\mathbf{D}_4(G_{Vv}) + \mathbf{D}_4(G_{Vv})^T = \begin{bmatrix} 0 & 12 & 19 & 57 \\ 12 & 0 & 57 & 19 \\ 19 & 57 & 0 & 128 \\ 57 & 19 & 128 & 0 \end{bmatrix}$$

The sum of the  $\mathbf{D}$  and  $\mathbf{D}^T$  matrices is always a symmetrical matrix.<sup>12</sup> Also, relatively more information is now stored in rows 1 and 2, the stereocenter rows of the summed matrices.

The GF values for the two (no. 1 and no. 2) stereocenters of the *Z* (*cis*) and *E* (*trans*) compounds are next built into two corresponding diagonal matrices.

$$\mathbf{D}_4^*(\text{GF}, Z) = \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{D}_4^*(\text{GF}, E) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Premultiplication of the sum of the  $\mathbf{D} + \mathbf{D}^T$  matrices with the diagonal GF matrices gave the data which provided the GM values for the two diastereomers of 1,2-difluoroethene; these data were obtained by summing all rows of the product

matrix,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 12 & 19 & 57 \\ 12 & 0 & 57 & 19 \\ 19 & 57 & 0 & 128 \\ 57 & 19 & 128 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 12 & 19 & 57 \\ 12 & 0 & 57 & 19 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

providing the GM value for (*Z*)-1,2-difluoroethene, +176; and in the same fashion yielding the GM number for (*E*)-1,2-difluoroethene, -176. The *S* index for 1,2-difluoroethene, calculated from its  $Vv$  weighted graph and  $\mathbf{D}$  matrix, is 1132. The GM number for (*Z*)-1,2-difluoroethene, 176, is added to the *S* index value of 1132; the sum is the *S* index for (*Z*)-1,2-difluoroethene, 1308. The *S* index for (*E*)-1,2-difluoroethene is attained in like manner:  $S = 1132 + (-176) = 956$ .

The various paired centers of isomerism were algebraically summed to arrive at the GM values for each stereocenter in substances with more than one pair of geometric centers, as in 2,4-dimethylcyclopentanol (Table 1,7); the OH is fiducial (no. 1).

1,2- <i>cis</i>	+1	+1	
2,4- <i>cis</i>		+1	+1
1,4- <i>cis</i>	+1		+1
GF	+2	+2	+2
1,2- <i>cis</i>	+1	+1	
2,4- <i>trans</i>		-1	-1
1,4- <i>trans</i>	-1		-1
GF	0	0	-2
1,2- <i>trans</i>	-1	-1	
2,4- <i>cis</i>		+1	+1
1,4- <i>trans</i>	-1		-1
GF	-2	0	0
1,2- <i>trans</i>	-1	-1	
2,4- <i>trans</i>		-1	-1
1,4- <i>cis</i>	+1		+1
GF	0	-2	0

The general technique of modified index calculation worked for a broad variety of substances that exhibit geometrical isomerism—double bonds between carbon-carbon, carbon-nitrogen, and nitrogen-nitrogen atoms (Table 1;1,2,3,4,5,9,10), substituted cyclic compounds (Table 1;6,7), and fused ring systems (Table 1;8). The derived GM values were incorporated into various molecular topological indices with no problems. However, as seen in Table 1, the larger the descriptor, the smaller the relative modification which is the consequence of eliding the GM integer with a given topological index. Were this germane to the use of a very large index, for some studies the GF value could be increased—most conveniently by a power of ten, for example.

Because the related diastereomers represent families (frequently pairs) of different substances, the pairs of geometrical isomers have different properties. Table 2 lists a series of paired *Z/E* alkene diastereomers for which the geometrically modified topological indices are listed. All indices locate the position of the double bond within the molecule, and all distinguish the *Z* from the *E* isomers; the series appears to be monotonic in character. It is interesting to note that *S* indices decrease slightly in dimension as the

**Table 2.** Descriptors Derived from the Vertex and Valence Weighted Distance Matrices of Geometrical Isomers of Representative C<sub>5</sub>–C<sub>8</sub> Alkenes

no.	compound	$R_M^{a,b}$	index $D_N^V(G_{V_{100}})$				
			S	PRS	det	per	lhaf
1	2-pentene		223	2 021 760	1 440	82 944	9 348
	(Z)-	24.9563	307	2 021 844	1 524	83 028	9 432
	(E)-	25.0248	139	2 021 676	1 356	82 860	9 264
2	2-hexene		366	274 362 660	-6 912	4 067 856	357 912
	(Z)-	29.5522	488	274 362 782	7 034	4 067 978	358 034
	(E)-	29.6371	244	274 362 538	6 790	4 067 734	357 790
3	3-hexene		364	329 204 736	-6 912	6 278 400	706 176
	(Z)-	29.6715	476	329 204 848	7 024	6 278 512	706 288
	(E)-	29.7518	252	329 204 624	6 800	6 278 288	706 064
4	2-heptene		561	51 266 779 200	32 256	273 588 480	5 207 400
	(Z)-	34.0614	731	51 266 779 370	32 426	273 588 650	5 207 570
	(E)-	34.2797	391	51 266 779 030	32 086	273 588 310	5 207 230
5	3-heptene		557	66 977 280 000	32 256	433 518 336	7 925 736
	(Z)-	34.3036	707	66 977 280 150	32 406	433 518 486	7 925 886
	(E)-	34.4178	407	66 977 279 850	32 106	433 518 186	7 925 586
6	2-octene		816	12 791 955 456 000	-147 456	25 323 892 992	350 096 256
	(Z)-	38.7959	1 044	12 791 955 456 228	147 674	25 323 893 220	350 096 484
	(E)-	38.8928	588	12 791 955 455 772	147 228	25 323 892 764	350 096 028
7	3-octene		810	17 337 889 522 800	-147 456	39 512 107 008	590 374 656
	(Z)-	38.9625	1 008	17 337 889 522 998	147 654	39 512 107 206	590 374 854
	(E)-	39.0856	612	17 337 889 522 602	147 258	39 512 106 810	590 374 458
8	4-octene		808	19 298 097 561 600	-147 456	50 581 398 528	961 352 064
	(Z)-	38.9471	996	19 298 097 561 788	147 644	50 581 398 716	961 352 252
	(E)-	39.0515	620	19 298 097 561 412	147 268	50 581 398 340	961 351 876

<sup>a</sup>  $R_M$  = molar refractivity =  $(n^2 - 1/n^2 + 2)(M/d)$  cm<sup>3</sup>/mol. <sup>b</sup> Calculated from  $n_D^{20}$  and  $d_4^{20}$  data in the *Handbook of Chemistry and Physics*, 71st ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1991.

alkene linkage moves from the exterior to the interior of a related set of isomers, but all other indices, except the constant determinant values, progress in the opposite order. Because the molar refractivity ( $R_M$ ), interpreted as the actual volume occupied by 1 mol of a substance, is a function of the number, nature, and arrangement of the atoms within a molecule, Table 2 also includes the  $R_M$  number for each isomer

$$R_M = \left( \frac{n^2 - 1}{n^2 + 2} \right) \left( \frac{M}{d} \right) \text{ cm}^3/\text{mol}$$

wherein  $d$  is the density,  $M$  is the molecular weight, and  $n$  is the refractive index of an alkene. The quantitative-structural relationships between  $R_M$  values and indices for the *Z* and *E* diastereomers were studied, with the following selected results.

**S (D) Index (Table 2, column 4; Z and E).** These sequences, like all sequences in Table 2, were unique and possessed numbers progressions that reflected the modified geometric index values related to the molar refractivity. Regression analyses for the *S* indices versus the molar refractivity ( $R_M$ ) values for all alkenes in Table 2 were studied, and unlike the analyses for the other indices were separated into *cis* and *trans* groups. Inspection of the *cis* and *trans* values revealed two diverse sets of points, as reflected in the two different equations below.

*cis*

$$R_M, \text{ cm}^3/\text{mol} = 8.4997 + 0.9553(S^{1/2})$$

$$n = 8 \quad r^2 = 0.996 \quad s = 0.3778 \quad F = 1336$$

*trans*

$$R_M, \text{ cm}^3/\text{mol} = 12.7020 + 1.0726(S^{1/2})$$

$$n = 8 \quad r^2 = 0.998 \quad s = 0.2778 \quad F = 2492$$

However, only the *cis* data for the *Z* diastereomers were compared for the remaining modified descriptors (Table 2, columns 5, 6, 7, 8). The signs of the determinant indices were recorded in Table 2, but only the absolute values were used in calculations for the modified indices and for the regression studies.

**PRS (D) Index (Table 2, column 5, Z)**

$$R_M, \text{ cm}^3/\text{mol} = 12.5706 + 1.9980(\log \text{ PRS})$$

$$n = 8 \quad r^2 = 0.999 \quad s = 0.1411 \quad F = 9613$$

**det (D) Index (Table 2, column 6, Z)**

$$R_M, \text{ cm}^3/\text{mol} = 2.5944 + 7.0185(\log \text{ det})$$

$$n = 8 \quad r^2 = 1.000 \quad s = 0.1063 \quad F = 16945$$

**per (D) Index (Table 2, column 7, Z)**

$$R_M, \text{ cm}^3/\text{mol} = 13.1842 + 2.4406(\log \text{ per})$$

$$n = 8 \quad r^2 = 0.998 \quad s = 0.2553 \quad F = 2933$$

1haf (**D**) Index (Table 2, column 8, Z)

$$R_M, \text{cm}^3/\text{mol} = 13.443 + 2.9246(\log 1\text{haf})$$

$$n = 8 \quad r^2 = 0.985 \quad s = 0.7020 \quad F = 383$$

An exceptionally good correlation was obtained with the determinant (**D**) index, and a good fit was observed with the product of the row sums (**D**) index. The five sets of geometrically modified topological indices were ordered in the following approximate sequence of decreasing reliability, with respect to the  $R_M$  values of the listed *cis*-alkenes.

$$\det > \text{PRS} > \text{per} > S > 1\text{haf}$$

Pogliani<sup>8</sup> reported a new topological *cis* connectivity  $^nX_c$  index from which was derived a *cis/trans* connectivity  $X_{ct}$  descriptor. A set of four pairs of geometrical isomers, similar but not identical to the eight pairs studied in this paper, gave an  $R$  value ( $N = 8$ ) of 0.9998 for the correlation between the  $X_{ct}$  values and the molar refractivities.

### OPTICAL ISOMERS

The general technique of determining modifications to topological indices of geometrical isomers was extended to optical isomers. The  $R$  configuration of a chiral stereocenter of a graph was designated with a chiral factor (CF) of +1; the  $S$  configuration about a stereocenter was indicated with a CF of -1. The remainder of the vertices of a graph were assigned CF values of zero.

As in the prior section, the  $Vv$  weighted **D** matrix of a graph possessing a chiral stereocenter was used first to calculate the values of the five molecular topological indices studied in this paper. There was then also derived from the **D** matrix, by use of the CF value, a chiral modifier (CM) which was used to alter a given topological index so that the contribution of the chiral center to the topological index of a molecule was recognized and quantified.

The steps used to accomplish the above objective were the same as earlier outlined: the  $Vv$  weighted **D** matrix was constructed, the **D** matrix was transposed, and then the **D** and **D**<sup>T</sup> matrices were summed. A diagonal CF matrix representing the configuration of the chiral vertex was constructed and premultiplied into the summed **D** + **D**<sup>T</sup> symmetrical matrix; the resultant product matrix row sum of the chiral vertex was used as a CM integer to adjust the original topological index to reflect the  $R$  or  $S$  configuration of the molecule.

Table 3 presents a brief illustrative list of substances and displays the results of the above techniques. 3-Methylhexane and bromochlorofluoromethane (Table 3;1,2) depict, respectively, a simple alkane and a one-carbon compound each of which possesses one chiral center. Note the amazing coincidence of the identical determinant and long hafnian values for compound 2, Table 3. The **D** matrices of the two substances beautifully illustrated how the bulk of the information presented by the **D** matrix of a graph shifts from the important chiral center in the instance of the alkane out to the peripheral heteroatoms in the case of the polyhalogen derivative of methane, justification for continued operations upon the sum of the **D** and **D**<sup>T</sup> matrices as detailed earlier in this paper.

Compounds 4 and 5 of Table 3 extend the above techniques to substances containing multiple chiral stereocenters. In 2,3-butanediol the *RR/SS* pair of optical antipodes is recognized as well as the one *meso* (*RS*) form of the molecule, the consequence of two identical chiral centers in the molecule. 2,3-Pentanediol (Table 3;5) displays the CM altered indices for the two pairs of diastereomers (*RR/SS* and *RS/SR*) which indicate the presence of two different chiral centers in 5.

3-Penten-2-ol (Table 1;5, Table 3;6,7) has both geometric and chiral centers of isomerism. Focusing only on the geometric centers (Table 1;5), the *Z* and *E* isomers of 3-penten-2-ol have geometrically modified  $S$  indices of 876 and 556, respectively. Next, observing its chiral center, 3-penten-2-ol (Table 3;6,7) has the  $S$  indices that reflect the  $R$  and  $S$  stereochemistry of each geometric isomer of the molecule. The two sets of data were elided, as illustrated for the geometrical and optical isomers of 3-penten-2-ol. The  $Vv$  weighted  $S$  index for the structure was first calculated; the  $S$  indices for the *cis* (*Z*) and *trans* (*E*) isomers were next derived. Finally, to the data for the geometrical  $S$  indices were added the modifying chiral data, +56 for the  $R$  configuration and -56 for the  $S$  configuration, summarized as follows:

3-Penten-2-ol		$S$ index
structure		716
geometry, <i>Z</i>	<i>Z</i>	876
	<i>E</i>	556
chirality, <i>ZR</i>	<i>ZR</i>	932
	<i>ZS</i>	820
	<i>ER</i>	612
	<i>ES</i>	500

1-Ethyl-3-methylcyclopentane (Table 3;8,9) presents a more sophisticated instance of melding one pair of geometrical diastereomers with two pairs of chiral diastereomers, the consequence of two different chiral stereocenters in the molecule.

Glyceraldehyde and  $\alpha$ -alanine (Table 3;10,11,12) display topological data for biologically important substances; the secondary amine (Table 3;13) conveys data for a substance with a heteroatom as a chiral center and whose optical isomers are not resolvable because of the very low energy barrier that separates the  $R$  from the  $S$  enantiomers.

The above techniques for determining the CM integers of organic substances possessing chiral centers can be used as a test for the chirality of a given vertex—in the instance of alkanes or cycloalkanes only a tertiary or quaternary vertex. If the separate row sums for the  $Vv$  weighted **D** + **D**<sup>T</sup> matrix of each vertex directly attached to a suspected chiral center all differ from one another, then the chiral nature of the tested vertex is established, but vide infra for a caveat to this observation.

It is enlightening to apply this simple algorithm to the 1,3-dimethyl- and 1,4-dimethylcyclohexanes, both of which exhibit geometrical isomerism, but only the former has chiral (no. 1 and no. 3) stereocenters. These interesting contrasts are succinctly illustrated in McMurry's textbook of organic chemistry.<sup>13</sup> Inspection of the row vector sums in the  $Vv$  weighted **D** + **D**<sup>T</sup> matrix of the three vertices attached to

**Table 3.** Descriptors Derived from Vertex and Valence Weighted Distance Matrices of Representative Optical Isomers and Modified for Chiral Isomerism

no.	compound	CMV <sup>a</sup>	index $D_N^N(G_{VW})$				
			S	PRS	det	per	lhaf
1	3-methylhexane		294	2 464 862 400	4 608	13 761 792	318 384
	(R)-	+44	338	2 464 862 444	4 652	13 761 836	318 428
	(S)-	-44	250	2 464 862 356	4 564	13 761 748	318 340
2	bromochlorofluoromethane		3 707	103 680 000	-3 317 760	9 953 280	3 317 760
	(R)-	+193	3 900	103 680 193	3 317 953	9 953 473	3 317 953
	(S)-	-193	3 514	103 679 807	3 317 567	9 953 087	3 317 567
3	2-butanol		302	933 120	1 728	35 424	4 494
	(R)-	+29	331	933 149	1 757	35 453	4 523
	(S)-	-29	273	933 091	1 699	35 395	4 465
4	2,3-butanediol		742	522 991 161	-58 320	8 039 412	622 566
	(R,R)-	+108	850	522 991 269	58 428	8 039 520	622 674
	(S,S)-	-108	634	522 991 053	58 212	8 039 304	622 458
	(R,S)-, meso	0	742	522 991 161	58 320	8 039 412	622 566
5	2,3-pentanediol		1 034	84 321 972 000	279 936	478 970 496	7 888 104
	(R,R)-	+131	1 165	84 321 972 131	280 067	478 970 627	7 888 235
	(S,S)-	-131	903	84 321 971 869	279 805	478 970 365	7 887 973
	(R,S)-	+5	1 039	84 321 972 005	279 941	478 970 501	7 888 109
	(S,R)-	-5	1 029	84 321 971 995	279 931	478 970 491	7 888 099
6	(Z)-3-penten-2-ol <sup>b</sup>		876	1 200 225 760	46 816	17 115 136	1 392 064
	(R)-	+56	932	1 200 225 816	46 872	17 115 192	1 392 120
	(S)-	-56	820	1 200 225 704	46 760	17 115 080	1 392 008
7	(E)-3-penten-2-ol <sup>b</sup>		556	1 200 225 440	46 496	17 114 816	1 391 744
	(R)-	+56	612	1 200 225 496	46 552	17 114 872	1 391 800
	(S)-	-56	500	1 200 225 384	46 440	17 114 760	1 391 688
8	cis-1-ethyl-3-methylcyclopentane <sup>c</sup>		632	464 950 886 524	15 676	1 035 368 764	12 101 308
	(R,S)-	-10	622	464 950 886 514	15 666	1 035 368 754	12 101 298
	(S,R)-	+10	642	464 950 886 534	15 686	1 035 368 774	12 101 318
9	trans-1-ethyl-3-methylcyclopentane <sup>c</sup>		384	464 950 886 276	15 428	1 035 368 516	12 101 060
	(R,R)-	+124	508	464 950 886 400	15 550	1 035 368 640	12 101 184
	(S,S)-	-124	260	464 950 886 152	15 304	1 035 368 392	12 100 936
10	2,3-dihydroxypropanal		1 768	42 326 323 200	-3 359 232	525 159 936	35 831 808
	(R)-	+92	1 860	42 326 323 292	3 359 324	525 160 028	35 831 900
	(S)-	-92	1 676	42 326 323 108	3 359 140	525 159 844	35 831 716
11	2-aminopropanoic acid		1 400	9 172 942 848	-995 328	124 250 112	9 123 840
	(R)-	+87	1 487	9 172 942 935	995 415	124 250 199	9 123 927
	(S)-	-87	1 313	9 172 942 761	995 241	124 250 025	9 123 753
12	zwitterion of 11		1 580	6 115 295 232	-663 552	82 833 408	6 082 560
	(R)-	+91	1 671	6 115 295 323	663 643	82 833 499	6 082 651
	(S)-	-91	1 489	6 115 295 141	663 461	82 833 317	6 082 469
13	ethylmethylamine		100	6 912	-144	768	312
	(R)-	+29	129	6 941	173	797	341
	(S)-	-29	71	6 883	115	739	283

<sup>a</sup> Chiral modification value. <sup>b</sup> Refer also to Table 1;5. <sup>c</sup> Refer also to Table 1;6.

the tertiary carbon atoms in the designated molecules provides a rapid affirmation of the identical chiralities of the no. 1 and no. 3 carbon atoms of 1,3-dimethylcyclohexane; by contrast, no chirality is observed in 1,4-dimethylcyclohexane.

The sum of the  $Vv$  weighted  $\mathbf{D} + \mathbf{D}^T$  matrices of 1,3-dimethylcyclohexane is

$$D_8^8(G_{VW}) + D_8^8(G_{VW})^T = \begin{bmatrix} 0 & 5 & 12 & 15 & 10 & 5 & 4 & 12 \\ 5 & 0 & 5 & 8 & 12 & 8 & 6 & 6 \\ 12 & 5 & 0 & 5 & 10 & 15 & 12 & 4 \\ 15 & 8 & 5 & 0 & 4 & 8 & 12 & 6 \\ 10 & 12 & 10 & 4 & 0 & 4 & 9 & 9 \\ 5 & 8 & 15 & 8 & 4 & 0 & 6 & 12 \\ 4 & 6 & 12 & 12 & 9 & 6 & 0 & 8 \\ 12 & 6 & 4 & 6 & 9 & 12 & 8 & 0 \end{bmatrix}$$

The three different row vector sums for those vertices directly attached to the no. 1 chiral stereocenter are no. 2, 50; no. 6, 58; no. 7, 57. The three different row vector sums for those vertices attached to the no. 3 chiral stereocenter are no. 2,

50; no. 4, 58; no. 8, 57; hence, vertices no. 1 and no. 3 are both chiral (and identical) stereocenters.

The same exercise extended to 1,4-dimethylcyclohexane provided the sum of the  $\mathbf{D} + \mathbf{D}^T$  matrices which displayed duplicate values among the row vector sums for those vertices adjacent to the tertiary (no. 1 and no. 4) carbon atoms, respectively: no. 2, 54; no. 6, 54; no. 7, 60 and no. 3, 54; no. 5, 54; no. 8, 60. Therefore, vertices no. 1 and no. 4 are identical and are also not chiral centers.

The same technique can be extended to heteroatom containing graphs; depending on the locations, numbers, and kinds of heteroatoms present in a graph, however, one should be prepared to recognize any class ( $1^\circ$ ,  $2^\circ$ ,  $3^\circ$ ,  $4^\circ$ ) of carbon atom as being a possible chiral center.

One can conjecture the rare instance of a graph containing a chiral center flanked by two different (perhaps even three or four) groups having identical  $Vv$  weighted  $\mathbf{D} + \mathbf{D}^T$  matrix vector sums, but whose individual vertex values, taken as a whole, are different. A search, limited to alkanes, demonstrated that 3,4,4-trimethylheptane fulfilled the prediction, with its carbon no. 3 chiral center flanked by row vector

**Table 4.** Descriptors with Chiral Modified *S* Values Derived from the Vertex and Valence Weighted Distance Matrix of D-2,3,4,5,6-Pentahydroxyhexanal

no.	compound	CMV <sup>a</sup>	<i>S</i> index
1	D-aldohexose no. 2C no. 3C no. 4C no. 5C	$\pm 287$ $\pm 255$ $\pm 259$ $\pm 299$	9 454
2	D-allose 2R, 3R, 4R, 5R	+1100	10 554
3	D-altrose 2S, 3R, 4R, 5R	+526	9 980
4	D-glucose 2R, 3S, 4R, 5R	+590	10 044
5	D-mannose 2S, 3S, 4R, 5R	+16	9 470
6	D-gulose 2R, 3R, 4S, 5R	+582	10 036
7	D-idose 2S, 3R, 4S, 5R	+8	9 462
8	D-galactose 2R, 3S, 4S, 5R	+72	9 526
9	D-talose 2S, 3S, 4S, 5R	-502	8 952

<sup>a</sup> Chiral modification value.

sums of no. 2, 87; no. 4, 87; and no. 8, 69. However, the individual row vector values of row no. 2 are 3, 0, 5, 12, 12, 16, 15, 6, 9, 9 and of row no. 4 are 15, 12, 7, 0, 6, 12, 16, 10, 5, 5.

Table 4 lists all the chiral stereoisomers of the open chain form of the D-aldohexoses, including D-glucose, which is of great biological significance. All the diastereomers displayed different CF integers, and hence had different modified *S* topological indices, as pictured in Table 4. There appeared to be no monotonic character portrayed within the series of data, and all attempts to relate the data to corresponding physical properties were futile—perhaps because the sugar molecules are mostly cyclic, hemiacetal in structure, not open chain.

The aldohexoses possess a number of similar chiral centers, and this experiment served as something of a test as to the capacity of these procedures to cope with such systems. These results, however, do not completely eliminate the possibility of degenerate values existing in even larger molecules with similar stereocenters.

## CONCLUSIONS

This paper describes how stereochemical, topographical ends were attained with topological means and exhibits molecular topological indices derived from the graphs and **D** matrices of compounds having the capacity for geometrical and/or chiral isomerism, modified to recognize the presence of such stereocenters and accordingly altering the corresponding indices—*S*, product of the row sums, determinant, permanent, and long hafnian. It appeared that the series of indices of similar alkene diastereomers displayed monotonicity, and as illustrated by the data of Table 2, when the

quantitative-structural comparisons with molar refractivities were examined, reasonably acceptable results were obtained. A simple algorithm for identifying chiral centers in organic molecules was also described.

## REFERENCES AND NOTES

- (1) (a) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228. (b) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 2. Graph Theory, Matrix Determinants and Eigenvalues, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 27–29. (c) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 3. Graph Theory, Binary and Decimal Adjacency Matrices, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 144–147. (d) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 4. Graph Theory, Matrix Permanents, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 69–72. (e) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 5. Graph Theory, Matrix Hafnians and Pfaffians, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 364–368. (f) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 6. Graph Theory and Molecular Topological Indices of Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 240–244. (g) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 7. Graph Theory and Molecular Topological Indices of Unsaturated and Aromatic Hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 863–867. (h) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 8. Graph Theory and Topological Indices of Heteronuclear Systems. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1151–1157.
- (2) Shelly, C. A.; Munk, M. E. An Approach to the Assignment of Canonical Connection Tables and Topological Symmetry Perception. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247–250, and references cited therein.
- (3) Barnard, J. M.; Cook, A. P. F.; Rohde, B. Storage and searching of stereochemistry in substructure search systems. In *Chemical Information Systems. Beyond the Structure Diagram*; Bawden, D., Mitchell, E., Eds.; Horwood: London, 1990; pp 29–41, and references cited therein.
- (4) Akutsu, T. A New Method of Computer Representation of Stereochemistry. Transforming a Stereochemical Structure into a Graph. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 414–417.
- (5) (a) Razingier, M.; Balasubramanian, K.; Perdih, M.; Munk, M. E. Stereoisomer Generation in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 812–825. (b) Razingier, M.; Perdih, M. Computerized Stereochemistry: Coding and Naming of Configurational Stereoisomers. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 290–296, and references cited therein.
- (6) Agarwal, K. K.; Gelernter, H. L. A Computer-Oriented Linear Canonical Notational System for the Representation of Organic Structures with Stereochemistry. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 463–479.
- (7) Mata, P.; Lobo, A. M. Implementation of the Cahn-Ingold-Prelog System of Stereochemical Perception in the LHASA Program. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 491–504.
- (8) Pogliani, L. On a Graph Theoretical Characterization of Cis/Trans Isomers. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 801–804, and references cited therein.
- (9) Eliel, E. L.; Wilen, S. H.; Mander, L. N. *Stereochemistry of Organic Compounds*; Wiley-Interscience: New York, 1994.
- (10) (a) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 160–163. (b) 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.
- (11) Gutman, I. Selected Properties of the Schultz Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1087–1089.
- (12) Pettofrezzo, A. J. *Matrices and Transformations*; Prentice-Hall: Englewood Cliffs, NJ, 1966; p 16.
- (13) McMurry, J. *Organic Chemistry*, 2nd ed.; Brooks-Cole: Pacific Grove, CA, 1988; pp 291–293.

CI950031R