

## A Topological Account of Chirality

István Lukovits<sup>\*,†</sup> and Wolfgang Linert<sup>‡</sup>

Chemical Research Center, H-1525 Budapest, P.O. Box 17, Hungary, and Vienna University of Technology, A-1060 Wien, Getreidemarkt 9/153, Austria

Received April 26, 2001

Graph invariants may differentiate structural isomers but are inappropriate to account for stereoisomerism and to distinguish between chiral structures. This work is an attempt to address this problem. A chiral function  $F$  satisfying condition  $F(D) = -F(L)$ , where  $D$  and  $L$  denote enantiomers of the same structure, has been applied in combination with Randić's index  ${}^1\chi^v$ . The resulting index  $\chi_c$  was used to explain the variance in thin-layer chromatographic retention indices.<sup>1</sup>

### INTRODUCTION

A graph invariant is a function of the underlying graph (structural formula)  $G$ , which does *not* depend on the numbering of the vertices in  $G$ . Graph invariants are highly useful in explaining physical/chemical properties of isomers in terms of the molecular structure. Since Wiener's discovery<sup>2</sup> a great number of graph invariants has been proposed. Trinajstić's book<sup>3</sup> is a detailed review on this topic.

In most cases graph invariants related to stereochemical isomers (cis/trans isomers, chiral molecules, knots, etc.) are identical. Because of this fact regression equations derived between an experimental feature and the graph invariants will fail to account for differences in the properties of stereoisomers. One method to correct this imperfection is based on graph invariants derived from geometric distances which will obviously yield different values of the invariants in case of cis/trans isomers.<sup>4,5</sup> Another approach, also concerning cis/trans isomers, is based on the triangles made up by the functional groups.<sup>6</sup>

As for chiral molecules, Ruch et al.<sup>7</sup> have considered 16 different types and devised a chirality function (see below) for each class. In this paper only one class (no. 7 in Ruch's classification) will be considered: molecules belonging to this class possess one chiral center, which means that all four groups attached to this center (a carbon atom) are different. In this class antipodes (optical isomers) of the same molecule have identical physical and chemical properties. If they interact with polarized light the angle of rotation will still be identical, but the sign of rotation will be opposite. If antipodes interact with structures being chiral themselves, the resulting complex will be diastereomeric. Because the resulting combined system is "diastereoisomeric", its properties will be different in case of  $D$  and  $L$  ligands. The difference in properties may be profound in case of interaction with cell-receptors, with one antipode producing zero effect, and the other antipode being rather efficient. In case of chiral chromatographic material, both antipodes may produce significant effects.

In deriving regression equations the simplest method to take into account the existence of the chiral center would be

to include a dummy variable, being equal to one in case of  $D$  isomers and minus one in case of  $L$  isomers. This approach, however, would become tedious in case of structures possessing more than one chiral center. Pyka<sup>1</sup> suggested another approach: she multiplied the matrix of distances  $D$  with a column vector,  $z$ , containing  $N$  (the number of vertices) entries. In case of  $L$  isomers all entries of  $z$ , including the entry related to the chiral center, are equal to 1. If a  $D$  isomer were considered, the entry related to the chiral center would be equal to  $-1$ . Entries of the resulting column vector have to be summed up to yield the chiral index  $I_{opt}$ . Therefore  $I_{opt}(L) = W$ , where  $W$  denotes the Wiener index,<sup>8</sup> and  $I_{opt}(D) < W$ . (Note that  $W$  is always positive.) As a consequence  $I_{opt}(L) \neq -I_{opt}(D)$  in Pyka's approach. Other mathematical properties of index  $I_{opt}$  have been discussed in detail elsewhere.<sup>9</sup>

Randić<sup>10</sup> has designed an index to account for the direction of circumference of planar structures. If the direction of circumference of planar structures is altered, Randić's index will have the same magnitude and opposite sign, and, therefore, it is suitable to account for binding to chiral receptor sites. On the other hand, Randić's index has not been used to establish quantitative structure–property relationships, so far.

In this paper we applied the polynomial index proposed by Ruch et al.,<sup>7,11</sup> fulfilling the condition  $F(D) = -F(L)$ , and  $F$  is the chiral function depending on the actual index. Brewster<sup>12</sup> has already proposed the same polynomial for open chain compounds. In addition to the polynomial Ruch et al.<sup>7,11</sup> proposed a second kind of chiral function. In this case the function is a linear combination of three-parametric antisymmetric functions, but the authors have not specified the actual form of the antisymmetric components. Function  $F$  can be used in combination with any graph theoretical index. As an illustration we applied this index to the high-pressure, thin-layer chromatographic (HPTLC) data measured by Pyka.<sup>1</sup>

### METHODS

Experimental Chir HPTLC retention indices  $R_f$  and  $R_M$  were taken from Pyka's work.<sup>1</sup> The figures listed in Table 1 are mean values based on 10 parallel measurements on  $R_f$ .

\* Corresponding author phone: (36)-1-325-7933; fax: (36)-1-325-7554; e-mail: lukovits@chemres.hu.

<sup>†</sup> Chemical Research Center.

<sup>‡</sup> Vienna University of Technology.

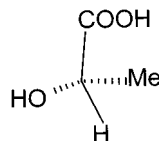


Figure 1. Stereoview of D-lactic acid.

$R_f$  and  $R_M$  are connected through the following equation:

$$R_M = \log(R_f^{-1} - 1) \quad (1)$$

If  $R_f$  is nearly equal to zero (but is still positive), then  $R_M$  will be a large positive number. With increasing values of  $R_f$ ,  $R_M$  decreases, and when  $R_f = 0.5$ , then  $R_M = 0$ . Although values  $R_f$  seem to be the primary experimental quantities, Pyka discussed the dependence of index  $R_M$  in terms of the molecular parameters, only. We shall consider the variance of both indices.

The molecular branching index,  $^1\chi$ , was introduced by Randić,<sup>13</sup> and later extended by Kier and Hall<sup>14</sup> to account for structures containing heteroatoms. According to this latter model the modified branching index,  $^1\chi^v$ , is defined as follows

$$^1\chi^v = \sum 1/(\delta_i \delta_j)^{1/2} \quad (2)$$

where  $\delta_i$  and  $\delta_j$  denote the valences of adjacent vertices  $i$  and  $j$ , respectively, and the summation has to be performed for all pairs of adjacent vertices, i.e., for all (chemical) bonds. Valence  $\delta_i$  is the difference between the number of valence electrons,  $Z_i$ , and the number of hydrogens,  $h_i$ , attached to vertex  $i$ :

$$\delta_i = Z_i - h_i \quad (3)$$

Hydrogens are neglected unless H belongs to a functional group or is bound directly to the chiral center. In this case  $\delta_H$  will be taken into account, and, for reasons to be explained in the next paragraph,  $\delta_H$  will be set equal to zero.

Let us consider a molecule with substituents A, B, C, and D being attached to a chiral center. The corresponding chirality function will be given in terms of the respective substituent constants  $a$ ,  $b$ ,  $c$ , and  $d$ :

$$F = (a - b)(a - c)(a - d)(b - c)(b - d)(c - d) \quad (4)$$

The order of A, B, C and D will determine the configuration, and obviously, if the order of any pair of substituents is interchanged, then  $F' = -F$ . If the order of a second pair of substituents is also interchanged, then the value of the resulting function  $F''$  will be identical to  $F$  ( $F'' = F$ ). Also  $F = 0$ , if at least two substituents happen to be identical. Observe that  $F$  may also be used if any of the substituents is hydrogen. In Ruch's<sup>7</sup> and Brewster's<sup>12</sup> models the chirality function was designated to account for a pseudoscalar (namely the optical rotation) of the antipodes and the contributions  $a$ ,  $b$ ,  $c$ , and  $d$  (eq 4) had to be determined *de novo* in each practical example. In contrast to that model the present approach employs (graph theoretical) substituent constants. In this paper  $F$  is based on the connectivity index and will therefore be denoted as  $\chi_c$ . The approach, however, is not restricted to branching indices.

Computation of  $\chi_c$  of D-lactic acid (Figure 1) will be used as an illustration. The following valence connectivity index

(related to groups attached to the chiral center of valence three) were used:  $^1\chi^v(-OH) = 0.33333333$ ,  $^1\chi^v(-COOH) = 0.716406078$ ,  $^1\chi^v(-CH_3) = 0.577350269$ , and  $^1\chi^v(H-) = 0.000000000$ . Using eq 4 we readily obtain  $\chi_c$  (D-lactic acid) = 0.001792123. A calculator program has been created to carry out computations.

In addition we list connectivity contributions of several other substituents attached to the chiral (valence = 3) center:  $^1\chi^v(NH_2-) = 0.333$ ,  $^1\chi^v(CH_3-S-CH_2-CH_2-) = 1.762$ ,  $^1\chi^v(HO-CH_2-) = 0.724$ ,  $^1\chi^v(CH_3-CH_2-CH-(CH_3)-) = 2.026$ ,  $^1\chi^v(HOOC-CH_2-CH_2-) = 1.690$ ,  $^1\chi^v(H-) = 0$ .

For reasons to be explained in the next section values of  $\chi_c$  were subtracted from  $^1\chi^v$  to obtain the combined index  $\chi^-$ :

$$\chi^- = ^1\chi^v - \chi_c \quad (5)$$

Using standard (calculator) programs we have derived regression equations. All equations listed were significant at the  $p < 0.001$  level. The errors of the regression coefficients have not been listed, because in case of mono-variate regression equations the F-test alone is sufficient to demonstrate the significance of the slope.

## RESULTS AND DISCUSSION

The calculated values of indices  $^1\chi^v$  and  $\chi_c$  are listed in Table 1. Note, that by selecting index  $^1\chi^v$  to characterize the substituents we also have fixed the configuration, because the set of parameters will unambiguously determine the sign of the chirality function. Of course for the same structure any other set of graph invariants might yield a different sign, and in this sense the configuration as determined by  $\chi_c$  is not absolute. However, if Ruch's<sup>7</sup> suggestion to compute the contributions  $a$ ,  $b$ ,  $c$ , and  $d$  (eq 4) *de novo* were put to use, the configuration would be absolute.

For the subseries of eight hydroxy acids rather moderate correlation coefficients between values  $R_f$ ,  $R_M$ , and  $^1\chi^v$  have been demonstrated

$$R_f = -0.064 ^1\chi^v + 0.822 \quad (6)$$

$$n = 8, r = -0.665, F = 4.76$$

$$|R_M| = -0.131 ^1\chi^v + 0.622 \quad (7)$$

$$n = 8, r = -0.674, F = 5.01$$

where  $n$  denotes the number of data points, and  $F$  is the result of the Fisher's  $F$ -test, and  $r$  is the correlation coefficient. Consideration of the absolute values of  $R_M$  allows us to discuss retention indices  $R_f$  and  $R_M$  simultaneously. Since all values of  $R_M$  were less than zero in our example, the transformation of regression equations derived for  $|R_M|$  into equations related to  $R_M$  is simple: the regression coefficient and the intercept have to be multiplied by  $-1$ . (The correlation coefficient has also to be multiplied by  $-1$ .) The correlation coefficient between  $|R_M|$  and  $^1\chi^v$  and also between  $R_f$  and  $^1\chi^v$  is negative. With increasing values of index  $^1\chi^v$  the absolute value of the chromatographic indices decreases. Factors leading to an increase in  $^1\chi^v$  will cause the overall reduction of the retention indices. The value of index  $^1\chi^v$

increases if the valence of the vertices decreases (eq 2). Replacement of carbons with vertices of higher degree (say replacement of carbon by oxygen) will decrease  ${}^1\chi^v$  and increase the values of  $R_f$  and  $|R_M|$ . Therefore the more polar or branched the molecules are the higher are values of  $R_f$  and  $|R_M|$ . Note that the group contributions to index  ${}^1\chi^v$  are additive.

On the other hand (again the absolute) values of retention indices related to L isomers are greater than the corresponding values related to D isomers. Because values of  $\chi_c$  related to L-isomers are (in most cases) positive and those related to D-isomers negative, we subtracted  $\chi_c$  from  ${}^1\chi^v$  (eq 5). With this convention we ensured that with decreasing values of the resulting index  $\chi^-$  (Table 1)  $R_f$  and  $|R_M|$  increase.

Regression eqs 8 and 9 were derived for the hydroxy acid series.

$$R_f = -0.070 \chi^- + 0.835 \quad (8)$$

$$n = 8, r = -0.844, F = 14.92$$

$$|R_M| = -0.140 \chi^- + 0.643 \quad (9)$$

$$n = 8, r = -0.843, F = 14.76$$

Observe that the absolute values of the correlation coefficients obtained by using  $\chi^-$  (eqs 8 and 9) are clearly greater than those obtained by using parameter  ${}^1\chi^v$  (eqs 6 and 7, respectively). Similar regression equations have been derived for the amino acid series (eqs 10 and 11):

$$R_f = -0.144 \chi^- + 0.998 \quad (10)$$

$$n = 10, r = -0.948, F = 70.58$$

$$|R_M| = -0.278 \chi^- + 0.948 \quad (11)$$

$$n = 10, r = -0.938, F = 58.15$$

The best (monivariate) correlation demonstrated by Pyka<sup>1</sup> was of lower quality. ( $r = 0.816$  for the hydroxy acid series, and  $r = 0.892$  for the amino acid series.) For the combined series, consisting of all 18 compounds, somewhat lower (with respect to the absolute value of) correlation coefficients were obtained. (Pyka<sup>1</sup> did not consider the combined series.)

$$R_f = -0.104 \chi^- + 0.906 \quad (12)$$

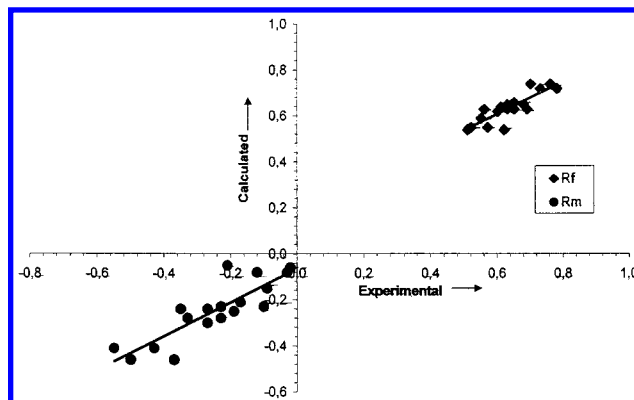
$$n = 18, r = -0.855, F = 43.29$$

$$|R_M| = -0.204 \chi^- + 0.776 \quad (13)$$

$$n = 18, r = -0.854, F = 43.04$$

The calculated vs the experimental values of the retention indices are shown in Figure 2.

The interpretation of the results (eqs 8–13) is the following. Variance in Randić's branching index explains achiral (see previous paragraphs) as well as the chiral interactions, taking place between the ligands and the (chiral) binding sites of the HPTLC plates. The achiral and chiral contributions to  $\chi^-$  are, however, different. If the differences in polarity between the functional groups A, B, C, and D



**Figure 2.** Experimental<sup>1</sup> vs calculated values of chromatographic retention indices  $R_f$  (◆, eq 12) and  $R_M$  (●, eq 13). Equation 13 is related to absolute values of  $R_M$ . Since  $R_M$  is in all cases less than zero, its value was calculated by using the formula  $R_M = 0.204 \chi^- - 0.776$ , instead of eq 13.

**Table 1.** Numerical Values of Retention Indices (See Ref 1)  $R_f$  and  $R_M$  and of the Topological Indices

compound	$R_f$	$R_M$	${}^1\chi^v$	$\chi_c$	$\chi^-$
L-lactic acid	0.76	-0.50	1.552	-0.002	1.554
D-lactic acid	0.70	-0.37	1.552	0.002	1.550
L-mandelic acid	0.65	-0.27	2.885	0.537	2.348
D-mandelic acid	0.57	-0.12	2.885	-0.537	3.422
L-hydroxyvaleric acid	0.68	-0.33	2.590	0.167	2.423
D-hydroxyvaleric acid	0.60	-0.17	2.590	-0.167	2.757
L-hydroxycaproic acid	0.69	-0.35	3.090	0.466	2.624
D-hydroxycaproic acid	0.62	-0.21	3.090	-0.466	3.556
L-methionine	0.61	-0.19	2.812	0.241	2.571
D-methionine	0.55	-0.09	2.812	-0.241	3.053
L-valine	0.63	-0.23	2.538	0.121	2.417
D-valine	0.56	-0.10	2.538	-0.121	2.659
L-leucine	0.65	-0.27	3.021	0.371	2.650
D-leucine	0.52	-0.03	3.021	-0.371	3.392
L-serine	0.78	-0.55	1.774	0.000	1.774
D-serine	0.73	-0.43	1.774	0.000	1.774
L-isoleucine	0.63	-0.23	3.076	0.411	2.665
D-isoleucine	0.51	-0.02	3.076	-0.411	3.487

increase, then the absolute value of the chirality function  $\chi_c$  will increase. In contrast to the achiral interactions accounted for by index  ${}^1\chi^v$ , the chiral part,  $\chi_c$ , is multiplicative, indicating that the interactions taking place between functional groups attached to the chiral center and *three* binding sites are synergistic. Retention indices increase if  ${}^1\chi^v$  decreases and  $\chi_c$  increases. Therefore the absolute values of retention indices of L isomers will be greater than the respective indices of D isomers (Table 1).

The lower retention indices observed with D isomers might also be a result of synergism. In this case the ligand is connected to *two* of the three binding sites; only the third coupling between the functional group and the surface cannot be realized because of geometrical constraints. Because of synergism, the contribution to the interaction might be equal to zero for D isomers. Equations 8–13 might implicate that because of the negative sign of index  $\chi_c$  in case of D isomers, retention decreases with increasing values of  $\chi_c$  with respect to the retention index of the corresponding L isomer or an achiral analogue. In reality this decrease has to be understood with respect to the average value of the retention indices related to L, D, and achiral isomers.

It has to be noted that  $\chi_c$  is neither an additive index, like  ${}^1\chi^v$ , nor a multiplicative index,<sup>15</sup> like  $\pi$ , but a combination of these. Combination of  $\chi_c$  with  ${}^1\chi^v$  (eq 5) is not obligatory;

both  $\chi_c$  and  ${}^1\chi^v$  might be considered separately in subsequent applications.

Further research is needed to clarify to which extent the substituents attached to the chiral center should be considered. In our case all bonds were taken into account, but in the case of cyclic molecules (e.g. proline) this approach will break down since the *same* chain is attached to the chiral center. Also in case of more bulky substituents, those atoms placed at a greater distance from the chiral center might not contribute to  $\chi_c$ . The solution of this question needs further investigation, but, clearly, truncation of the substituents at a definite distance would be an answer.

Finally it should be noted that our approach is capable of treating molecules containing several chiral centers, too. In this case the contributions related to various chiral centers must be added. Achiral carbons will not contribute to the sum. The use of this model to such more complicated cases will be investigated.

#### ACKNOWLEDGMENT

I.L. is indebted to Drs. M. Simonyi and T. Cserhádi (Budapest) for helpful discussions and to the Austrian Chemical Society for supporting his stay at the Technical University of Vienna. Thanks are further due to the "Fonds zur Förderung der Wissenschaftlichen Forschung in Österreich" (Project 15048-CHE) to the Training and Mobility of Researchers Network (TMR) project "Thermal and Optical Spin State Switching" (TOSS) supported by the European Community under the Contract number ERB-FMRX-CT98-0199 and to the European Science Foundation under the project "Molecular Magnets".

#### REFERENCES AND NOTES

- (1) Pyka, A. A New Optical Topological Index ( $I_{opt}$ ) for Predicting the Separation of D and L Optical Isomers by TLC. Part III. *J. Planar Chromatogr.* **1993**, 6, 282–288.
- (2) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- (3) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
- (4) Randić, M.; Jerman-Blazić, B.; Trinajstić, N. Computation of 3-dimensional Molecular Descriptors. *Comput. Chem.* **1990**, 14, 237–246.
- (5) Pyka, A. A new Optical Topological Index ( $I_{opt}$ ) for Predicting the Separation of D and L Optical Isomers by TLC Part III. *J. Planar Chromatogr.* **1993**, 6, 282–288.
- (6) Pogliani, L. On a Graph Theoretical Characterization of Cis/Trans Isomers. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 801–804.
- (7) Ruch, E.; Schönhofer, A. Näherungsformeln für spiegelungsantimetrische Moleküleigenschaften. *Theor. Chim. Acta* **1968**, 10, 91–110.
- (8) Wiener, H. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, 69, 2636–2638.
- (9) Gutman, I.; Pyka, A. New Topological Indices for Distinguishing Between Enantiomers and Stereoisomers: A mathematical Analysis. *J. Serb. Chem. Soc.* **1997**, 62, 261–265.
- (10) Randić, M. Graph Theoretical Description of Two-Dimensional Chirality with Possible Extension to Three-Dimensional Chirality. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 639–649.
- (11) Ruch, E.; Schönhofer, A.; I. Ugi. Die Vandermondesche Determinante als Näherungsansatz für eine Chiralitätsbeobachtung, ihre Verwendung in der Stereochemie und zur Berechnung der optischen Aktivität. *Theor. Chim. Acta* **1967**, 7, 420–432.
- (12) Brewster, J. H. A Useful Model of Optical Activity. I. Open Chain Compounds. *J. Am. Chem. Soc.* **1959**, 81, 5475–5483.
- (13) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
- (14) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (15) Gutman, I.; Linert, W.; Lukovits, I.; Tomović, Ž. The Multiplicative Version of the Wiener Index. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 113–116.

CI0100346