

# A Novel Definition of the Hyper-Wiener Index for Cycles

I. Lukovits

Central Research Institute for Chemistry, Hungarian Academy of Sciences,  
P.O. Box 17, H-1525 Budapest, Hungary

W. Linert\*

Institute of Inorganic Chemistry, Technical University of Vienna, Getreidemarkt 9, A-1060 Vienna, Austria

Received December 14, 1993\*

The definition of the hyper-Wiener index  $R$  (a graph invariant) proposed by Randic for acyclic graphs, has been extended to simple cycles. The extended version of  $R$  is consistent with the definition given by Randic for trees. Quantitative structure-property relationships were derived between boiling points and values of  $R$  in a series of cyclic and acyclic alkanes.

## 1. INTRODUCTION

The hyper-Wiener index  $R$  was recently proposed by Randic.<sup>1</sup> In his definition,  $R$  is the sum of contributions  $K_{ij}$ , where subscripts  $i$  and  $j$  denote a pair of vertices.  $K_{ij}$  may be obtained by using the following two-step algorithm. (1) Remove the path (for trees there is only one path) between vertices  $i$  and  $j$ ; two nonconnected subgraphs will be obtained. (2) Multiply the number of vertices of the first subgraph by the number of vertices of the second one.  $R$  may be obtained by adding  $K_{ij}$  for all pairs of vertices  $i$  and  $j$ . There are  $N(N-1)/2$  terms altogether, where  $N$  denotes the number of vertices (i.e., the number of carbons in H-suppressed graphs). Consider the example for isobutane (2-methylpropane): the central vertex is numbered 2 and we have  $K_{12} = 1 \times 3 = 3$ ;  $K_{13} = 1 \times 1 = 1$ ;  $K_{14} = 1 \times 1 = 1$ ;  $K_{23} = 3 \times 1 = 3$ ;  $K_{24} = 3 \times 1 = 3$ ;  $K_{34} = 1 \times 1 = 1$ ; and  $R = 3 + 1 + 1 + 3 + 3 + 1 = 12$ . However, this algorithm may be used only for trees.

$R$  may be obtained by a numerical procedure as proposed by Randic *et al.*<sup>2</sup> For special subsets of graphs, the Y, X, and  $\dagger\dagger$  trees (a Y tree contains only one trivalent branching vertex, an X tree contains only one tetravalent branching vertex, and a  $\dagger\dagger$  tree contains two branching vertices, the degrees of which are maximally four),  $R$  can be obtained by using analytical formulas.<sup>3,4</sup> Note that the formula obtained for  $\dagger\dagger$  trees is also valid for Y and X trees, and the formula obtained for X trees is also valid for Y trees. For details, see the references.

It may be assumed that  $R$ , similarly to the Wiener index<sup>5</sup>  $W$ , measures the compactness of a molecule. Randic *et al.* proposed<sup>2</sup> a method by which it is possible to extend  $R$  to cover the case of simple cycles. The "path count" scheme of Randic *et al.* gives for cyclohexane  $R = 111$ , which is larger than the value  $R = 70$  obtained for  $n$ -hexane. It could be asserted that this result is incompatible with the assumption that  $R$  measures the compactness of a graph, since one would then expect the  $R$  of cyclohexane to be lower than that for  $n$ -hexane, but of course other interpretations of  $R$  may exist, by means of which the result of Randic *et al.* might be justified.

The aim of the present paper is to extend the definition of  $R$  for cyclic graphs in a manner that is consistent with the definition of  $R$  by Randic. This definition produces values for cyclic structures that are lower than those for corresponding chains. Since  $R$  has been used up to now to derive quantitative

structure-activity relationships in only one case,<sup>1</sup> we have derived a regression equation between  $R$  and the boiling points of a series of alkanes and cycloalkanes containing altogether 62 molecules.

## 2. THEORY

The expressions graph and molecular structure, vertex and atom, and edge and bond between atoms will be used interchangeably hereafter. In order to understand the relationship between  $W$  and  $R$ , the definition of  $W$  due to Wiener is given here:

$$W = \sum_e W_e \quad (1)$$

where  $W_e$  denotes a contribution from edge  $e$  belonging to a tree  $T$  and is equal to the product of the number of vertices on the "left-hand side" and the number of vertices on the "right-hand side" of edge  $e$ .<sup>5</sup> The summation includes all edges, and therefore  $W$  includes contributions from adjacent vertices only. But step 1 of the algorithm used to define  $R$  (see Introduction) is equivalent to calculating  $W_e$  of a tree  $T'$ , consisting of two subgraphs and a single edge which connects them at vertices  $i$  and  $j$ .

The "product" definition of  $W$  could also be extended for cyclic structures.<sup>6</sup> According to this definition,  $W$  is equal to the sum of edge contributions  $W_e$  of any cyclic graph:

$$W_e = \sum_{i \neq j} \frac{k_{ij}^e}{k_{ij}} \quad (2)$$

where  $k_{ij}^e$  denotes the number of all paths of minimal length which connect vertices  $i$  and  $j$  and involved edge  $e$ , and  $k_{ij}$  is the number of all paths of minimal length between  $i$  and  $j$ .  $W$  is obtained by summing contributions  $W_e$  for all edges  $e$  (eq 1). In fact,  $W_e$  may be obtained for single cycles by using simple formulas,<sup>7</sup>:

$$W_e = N^2/8 \quad \text{for } N = \text{even} \quad (3)$$

$$W_e = (N^2 - 1)/8 \quad \text{for } N = \text{odd} \quad (4)$$

where  $N$  is the number of vertices.

\* Author to whom correspondence should be addressed.

**Table 1.** Calculated Values of  $R$  and  $W$  for  $n$ -Alkanes and Cycloalkanes up to  $N = 8$ 

no. of carbons	$R$		$W$	
	chain	cycle	chain	cycle
1	0	0	0	0
2	1	1	1	1
3	5	3	4	3
4	15	10	10	8
5	35	25	20	15
6	70	51	35	27
7	126	94.5	56	42
8	210	172	84	64

The "bond contribution" definition and the usual definition of  $W$  via the distance matrix<sup>8</sup> yield identical  $W$  for any cyclic graph.<sup>6</sup> Note that  $NW_c = W$ .

In order to extend the definition of  $R$  to cycles, it is necessary to recall that step 1 of the algorithm produces a subtree  $T'$  of the original tree  $T$ , and the contributions  $K_{ij}$  are essentially bond contributions in tree  $T'$ , in which vertices  $i$  and  $j$  are adjacent. Therefore, if a contribution  $K_{ij}$  in a cycle  $C$  has to be obtained and  $i$  and  $j$  are nonadjacent vertices,  $K_{ij}$  is a bond contribution in a reduced cycle  $C'$ , in which  $i$  and  $j$  are adjacent. The algorithm will be explained by using a simple example, cyclohexane. The atoms in the ring are numbered consecutively.

In cyclohexane there are  $6(6-1)/2 = 15$  distinct pairs of vertices; in general, in a graph containing  $N$  vertices, there are  $N(N-1)/2$  distinct pairs of vertices. Out of these 15 pairs, six pairs are adjacent (in the general case,  $N$  pairs), six pairs are at distance two from one another, and three pairs are at distance three from one another.

**Contributions from Adjacent Vertices.** The adjacent vertices define six edges making up  $C_6$  (a cycle involving six vertices), and by eq 3 their contribution to  $R$  is  $R_6 = 6 \times 6^2/8 = 27$  (i.e., the value of  $W$  in cyclohexane).

**Contributions from Vertices at Distances Two from One Another.** For the contribution of vertices 1 and 3, one gets a reduced cycle  $C_5$ . The same is valid for the other pairs of vertices which are at distance two from one another. By eq 4 the sum of all six contributions is therefore  $R_5 = 6 \times (5^2 - 1)/8 = 18$ .

**Contributions from Vertices at Distances Three from One Another.** By considering contributions between pairs of vertices 1-4, 2-5, and 3-6, we arrive at a reduced cycle  $C_4$ . Hence, by using eq 3 we get  $R_4 = 3 \times 4^2/8 = 6$ .

In this way, all pairs of vertices were considered. Therefore,  $R = R_6 + R_5 + R_4 = 27 + 18 + 6 = 51$ . Table 1 lists calculated values  $W$  and  $R$  of cycloalkanes and  $n$ -alkanes up to  $N = 8$ . It should be noted that noninteger values are also possible (e.g., cycloheptane, Table 2).

In general, if  $N$  is even, then  $N - (N/2 + 1)$  cycles with  $N$  contributions plus a single cycle with  $N/2$  contributions have to be considered, yielding  $N^2 - N^2/2 - N + N/2 = N(N-1)/2$  terms altogether. If  $N$  is odd, then  $N - (N+3)/2 + 1$  cycles have to be considered, and each cycle produces  $N$  contributions; hence there are again  $N^2 + N - N^2/2 - 3N/2 = N(N-1)/2$  contributions.

### 3. NUMERICAL METHODS

The values of  $R$  for acyclic structures with no more than two branching points (a branching point is a vertex with degree  $> 2$ ) have been obtained by using a computer program for the formula<sup>4</sup> given in the Appendix (Table 2).  $R$  has been calculated for structures containing three branching atoms by using the Wiener matrix method.<sup>2</sup> The experimental data

**Table 2.** Experimental and Estimated Boiling Points ( $^{\circ}\text{C}$ ) of Alkanes and Cycloalkanes

no.	molecule	$N$	$R$	exptl bp	estd bp
1	methane	1	0	-164.0	-158.3
2	ethane	2	1	-88.6	-99.9
3	propane	3	5	-42.1	-37.1
4	butane	4	15	-0.5	11.7
5	isobutane	4	12	-11.7	1.6
6	2,2-dimethylpropane	5	22	9.5	29.2
7	2-methylbutane	5	28	27.9	40.2
8	$n$ -pentane	5	35	36.1	50.5
9	2,2-dimethylbutane	6	44	49.7	61.0
10	2,3-dimethylbutane	6	47	57.9	64.1
11	2-methylpentane	6	58	60.3	73.8
12	3-methylpentane	6	54	63.3	70.5
13	$n$ -hexane	6	70	68.7	82.5
14	2,2,3-trimethylbutane	7	69	80.9	81.8
15	2,2-dimethylpentane	7	84	79.2	90.9
16	3,3-dimethylpentane	7	76	86.0	86.3
17	2,5-dimethylpentane	7	91	80.5	94.6
18	2,3-dimethylpentane	7	83	89.8	90.4
19	2-methylhexane	7	108	90.1	102.6
20	3-methylhexane	7	99	91.9	98.5
21	3-ethylpentane	7	90	93.5	94.1
22	$n$ -heptane	7	126	98.4	109.7
23	2,2,3,3-tetramethylbutane	8	97	106.3	97.6
24	2,2,4-trimethylpentane	8	127	99.3	110.1
25	2,2,3-trimethylpentane	8	115	109.9	105.5
26	2,2,3-trimethylpentane	8	111	114.8	103.8
27	2,3,4-trimethylpentane	8	122	113.5	108.2
28	2,2-dimethylhexane	8	149	106.9	117.5
29	3,3-dimethylhexane	8	131	112.0	111.5
30	2,5-dimethylhexane	8	161	109.0	121.1
31	2,4-dimethylhexane	8	147	109.5	116.9
32	2,3-dimethylhexane	8	143	115.6	115.6
33	3-methyl-3-ethylpentane	8	118	118.3	106.7
34	2,2,4,4-tetramethylpentane	9	173	122.3	124.4
35	3,4-dimethylpentane	8	134	117.7	112.6
36	2-methyl-3-ethylpentane	8	129	115.7	110.8
37	2-methylheptane	8	185	117.7	127.6
38	3-methylheptane	8	170	119.0	123.6
39	4-methylheptane	8	165	117.7	122.2
40	2,2,3,4-tetramethylpentane	8	150	118.3	117.8
41	3-ethylhexane	9	149	140.3	117.5
42	$n$ -octane	9	164	133.1	122.0
43	2,4,4-trimethylhexane	8	210	125.7	133.4
44	2,2,5-trimethylhexane	9	189	130.4	128.6
45	2,2,5-trimethylhexane	9	217	124.1	135.0
46	2,2,4-trimethylhexane	9	198	126.5	130.7
47	2,2,3-trimethylhexane	9	190	133.6	128.8
48	2,3,3-trimethylhexane	9	181	137.7	126.5
49	2,2-dimethyl-3-ethylpentane	9	171	133.8	123.9
50	3,3,5-trimethylhexane	9	206	131.3	132.6
51	3,3,4-trimethylhexane	9	172	140.5	124.2
52	2,2-dimethylheptane	9	247	132.7	141.0
53	3,4-dimethyl-3-ethylpentane	9	178	136.7	125.8
54	3,3-diethylpentane	9	170	146.2	123.6
55	4-methyloctane	9	264	142.5	144.1
56	$n$ -nonane	9	330	150.8	154.5
57	cyclopropane	3	3	-32.7	-58.6
58	cyclobutane	4	10	12.0	-6.6
59	cyclopentane	5	25	49.2	35.0
60	cyclohexane	6	51	80.7	67.8
61	cycloheptane	7	94.5	118.5	96.4
62	cyclooctane	8	172	148.5	24.3

were collected from Kier's compilation<sup>9</sup> (structures 2-56, Table 2) and from a handbook<sup>10</sup> (structures 1 and 57-62, Table 2).

### 4. RESULTS AND DISCUSSION

It seems that both Randic's<sup>2</sup> scheme and the present scheme are consistent with the definition of  $R$  in trees, yet they produce nonequivalent values for simple cycles. The algorithm proposed in this paper produces values of  $R$  for cycles that are lower than those related to the corresponding chains. The ratio between values of  $R$  of cyclic alkanes vs  $n$ -alkanes is 1

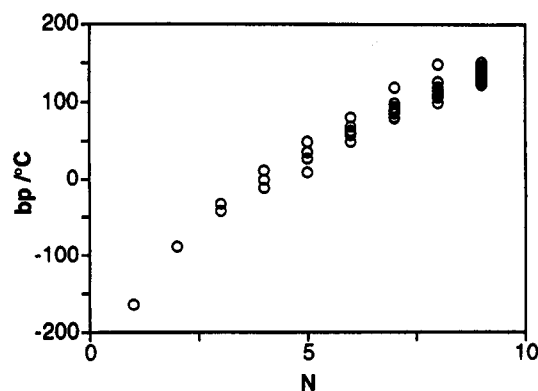


Figure 1. Boiling points of alkanes in terms of the number of carbons.

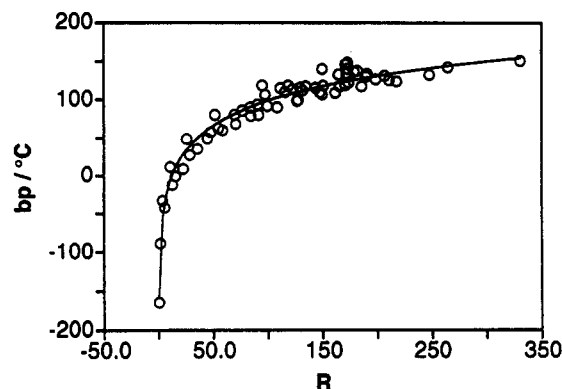


Figure 2. Boiling of alkanes in terms of the hyper-Wiener  $++$  index  $R$ .

for  $N = 2$ ; 0.6 for  $N = 3$ ; 0.67 for  $N = 4$ ; 0.71 for  $N = 5$ ; 0.73 for  $N = 6$ ; 0.75 for  $N = 7$ ; and 0.82 for  $N = 8$ . To what number this ratio converges with increasing  $N$  remains to be investigated.

Boiling points of a series of 56 alkanes and six cycloalkanes have been collected<sup>9,10</sup> (Table 2) and regressed against  $N$  and  $R$ . The best equations obtained were

$$\text{bp} = 30.5N - 130.4 \quad r = 0.970; \quad s = 14.8; \quad F = 946 \quad (5)$$

$$\text{bp} = 46.6 \ln(R + 0.4) - 115.6 \quad r = 0.982; \quad s = 11.6; \quad F = 783 \quad (6)$$

where  $r$  denotes the multiple correlation coefficient,  $s$  denotes the standard error of estimate, and  $F$  is Fisher's  $F$  value. Note that the inclusion of an additional factor in eq 6 increased  $r$  significantly as compared with  $r$  obtained without this factor, and its use is therefore justified. It also ensures that methane can be treated together with the other hydrocarbons. The experimental and calculated values of the boiling points are listed in Table 2. The dependence of the boiling point on  $N$  and  $R$  is shown in Figure 1 and Figure 2, respectively. Comparison of eqs 5 and 6 indicates that the greater part of the error sum of squares is due to additive effects, and although  $N$  cannot be used to explain the variation in any series of isomers (Figure 1), we think that  $N$  is an important graph invariant in quantitative structure-property relationships, the importance of which has often been neglected. However, there are also examples in which  $N$  does appear.<sup>11-13</sup>

Equation 6 underestimates the boiling points for cyclic structures, and therefore it seems that somewhat higher values of  $R$  would be more appropriate. On replacing for cyclohexane the value  $R = 51$  in this equation by the corresponding value of 111 produced by the Randic scheme,<sup>2</sup> we would predict a

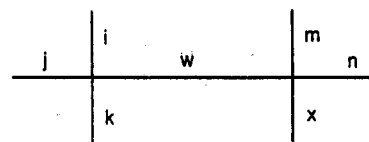


Figure 3. Scheme of an  $++$  tree.

boiling point of 104 °C. In other words, the number obtained by the Randic approach overestimates, whereas our  $\text{bp} = 68$  °C (Table 2) value underestimates the experimental value of 80.7 °C. In fact, all cyclic structures have higher boiling points than their  $n$ -alkane counterparts (Table 2), indicating that a further factor that is not accounted for by  $R$  should be included in the regression equation.

The extension of the present definition of  $R$  to polycyclic systems containing acyclic side branches has yet to be solved.

## ACKNOWLEDGMENT

Thanks are due to the "Fonds zur Förderung der Wissenschaftlichen Forschung in Österreich" (Grant No. 8795) for financial support.

## 5. APPENDIX

$R$  can be calculated for trees with two branching points, the maximal degree of which may not exceed four, for trees with a single branching point, the degree of which may not exceed six, or for chains<sup>4</sup> by using the following formula:

$$\begin{aligned} R = & [i^4 + j^4 + k^4 + m^4 + n^4 + x^4 + w^4 + 4(i^3j + ij^3 + \\ & i^3k + ik^3 + j^3k + jk^3 + m^3n + mn^3 + m^3x + mx^3 + \\ & n^3x + nx^3 + i^3m + im^3 + i^3n + in^3 + i^3x + ix^3 + \\ & j^3m + jm^3 + j^3n + jn^3 + j^3x + jx^3 + k^3m + km^3 + \\ & k^3n + kn^3 + k^3x + kx^3 + i^3w + iw^3 + j^3w + jw^3 + \\ & k^3w + kw^3 + m^3w + mw^3 + n^3w + nw^3 + x^3w + xw^3) + \\ & 12(i^2wm + j^2wm + k^2wm + i^2wn + j^2wn + k^2wn + i^2wx + \\ & j^2wx + k^2wx + iwm^2 + jwm^2 + kwm^2 + iwn^2 + jwn^2 + \\ & kwn^2 + iwx^2 + jwx^2 + kwx^2 + iw^2m + jw^2m + kw^2m + \\ & iw^2n + jw^2n + kw^2n + iw^2x + jw^2x + kw^2x) + 6(i^2j^2 + \\ & i^2k^2 + j^2k^2 + m^2n^2 + m^2x^2 + n^2x^2 + i^2m^2 + i^2n^2 + i^2x^2 + \\ & j^2m^2 + j^2n^2 + j^2x^2 + k^2m^2 + k^2n^2 + k^2x^2 + i^2w^2 + j^2w^2 + \\ & k^2w^2 + m^2w^2 + n^2w^2 + x^2w^2) - 22(i^3 + j^3 + k^3 + m^3 + \\ & n^3 + x^3 + w^3) - 6(i^2j + ij^2 + i^2k + ik^2 + j^2k + jk^2 + \\ & m^2n + mn^2 + m^2x + mx^2 + n^2x + nx^2) - 42(i^2w + j^2w + \\ & k^2w + m^2w + n^2w + x^2w + iw^2 + jw^2 + kw^2 + mw^2 + \\ & nw^2 + xw^2) - 18(i^2m + im^2 + j^2m + jm^2 + k^2m + km^2 + \\ & i^2n + in^2 + j^2n + jn^2 + k^2n + kn^2 + i^2x + ix^2 + j^2x + \\ & jx^2 + k^2x + kx^2) - 36(iwm + iwn + iwx + jwm + \\ & jwn + jwx + kwm + kwn + kwx) + 35(i^2 + j^2 + k^2 + \\ & m^2 + n^2 + x^2) + 107w^2 - 2(ij + ik + jk + mn + mx + \\ & nx) + 22(im + jm + km + in + jn + kn + ix + jx + \\ & kx) + 70(iw + jw + kw + mw + nw + xw) - \\ & 14(i + j + k + m + n + x) - 86w]/24 \quad (7) \end{aligned}$$

where  $i, j, k, m, n, x$ , and  $w$  denote the number of vertices in the respective strings (Figure 3) and  $N = i + j + k + w + m + n + x - 6$ . Note that if the length of string  $i$  (or any other string) is equal to 1, then this string does not appear as a separate string. (A string is a sequence of vertices and bonds.) A string may begin with a vertex of degree one (an endpoint) or with a branching point and end with an endpoint

or with a branching point.<sup>14</sup> Therefore, branching points are counted more than once, and this is corrected by the additional factor 6. Examples for the parameters in eq 7: (1) methane,  $i = j = k = m = n = x = w = 1$ ; (2) 2,2-dimethylpropane,  $i = j = k = w = 2$ ,  $m = n = x = 1$ , or  $i = j = k = 1$ ,  $w = m = n = x = 2$ , or  $i = m = n = x = 2$ ,  $j = k = w = 1$ , etc. The formula can easily be programmed and checked by using the examples by Randic *et al.*<sup>2</sup>

## REFERENCES AND NOTES

- (1) Randic, M. Novel molecular descriptor for structure-property studies-property studies. *Chem. Phys. Lett.* **1993**, *211*, 478-483.
- (2) Randic, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. J. Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709-716.
- (3) Lukovits, I. Formulas for the hyper-Wiener index. Submitted for publication in *J. Chem. Inf. Comput. Sci.*
- (4) Lukovits, I. Formula of the hyper-Wiener index of  $\dagger\dagger$  trees. Submitted for publication in *Comput. Chem.*
- (5) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- (6) Lukovits, I. Wiener indices and partition coefficients of unsaturated hydrocarbons. *Quant. Struct.-Act. Relat.* **1990**, *9*, 227-231.
- (7) Lukovits, I. The generalized Wiener index for molecules containing double bonds and the partition coefficients. *Rep. Mol. Theory* **1990**, *1*, 127-131.
- (8) Hosoya, H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332-2339.
- (9) Kier, L. B.; Hall, L. H. Molecular Connectivity in *Chemistry and Drug Research*; Academic: New York, 1976; pp 42, 132.
- (10) Weast, R. C. *CRC Handbook of Chemistry and Physics*, 67th ed.; CRC Press: Boca Raton, FL, 1987.
- (11) Trinajstić, N.; Randic, M.; Klein, D. J. On the quantitative structure-activity relationship in drug research. *Acta Pharm. Jugosl.* **1986**, *36*, 267-279.
- (12) Seybold, P. G.; May, M. A.; Gargas, M. J. On the physical properties of halogenated hydrocarbons. *Acta Pharm. Jugosl.* **1986**, *36*, 253-265.
- (13) Balaban, A. T.; Ciubotariu, D.; Medeleanu, M. Topological Indices and Real Number Vertex Invariants Based on Graph Eigenvalues or Eigenvectors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 517-523.
- (14) Lukovits, I. General formulas for the Wiener index. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 503-507.