

Counts of All Walks as Atomic and Molecular Descriptors

Gerta Rücker and Christoph Rücker*

Institut für Organische Chemie und Biochemie, Universität Freiburg, Albertstrasse 21, D-79104 Freiburg, FRG

Received March 8, 1993

A family of graph theoretical descriptors for atoms and for molecules is introduced: the counts of all walks of increasing length (all walks as opposed to self-returning walks and to paths). These descriptors, while extremely easily calculated by a Morgan-type summation procedure, exhibit high discriminating power among isomers and among structural situations of atoms. A systematic search for degeneracies is performed within tree graphs. The walk-based topological index *twc* (total walk count) is less degenerate than Balaban's *J*, and a code for graphs based on all walks seems to be more discriminating than Trinajstić's OSC (ordered structural code). The new descriptors are successfully used as alternative variables in the description of ¹³C NMR data for usual and highly branched alkanes.

INTRODUCTION

Topological indices (TI; a better term would be graph theoretical indices) are simple numbers which describe, in a condensed manner, certain structural properties of graphs, such as size and branchedness. They are graph invariants; that is, they are independent of the particular drawing or numbering of the graph's vertices. Since the simple constitutional formula of a molecule, a highly abstracted picture of that molecule, can be looked upon as a graph, such numbers correspond to properties of molecules also. In chemistry, TI's are widely used for two purposes. First, in chemical documentation they characterize chemical structures, thus allowing storage and retrieval. Second, they allow one to describe or predict experimental properties of molecules solely on the basis of the structural formula, provided that numerical values of that property for several similar structures are known (structure–property relationships). The underlying dogma is that similar structures will exhibit similar properties. The problem then is to describe structures in terms of those descriptors that best reveal important similarities and differences. Since different properties may depend on different structural features, they may require different TI's. For both purposes a TI should be as discriminant as possible while at the same time easily calculated. For leading references on the use and limitations of TI's, see recent reviews and books.^{1–12}

Whereas for decades much work was devoted to the development of better and better TI's, which are descriptors for molecules, corresponding descriptors for atoms in molecules (vertex invariants, atomic topological indices, ATI's) received some attention only recently.^{7,13–16} Such atomic descriptors could serve similar purposes as above: They characterize an atom's environment in simple numbers, thus allowing storage and retrieval of structural situations of atoms, and they may be correlated to certain experimental properties of atoms in molecules. Since the rational design of ATI's is a very recent development, correlations to any experimental properties are scarce.

Reduction of a graph to a simple number necessarily results in loss of information, and thus a loss in discrimination among nonisomorphic graphs. Therefore all TI's are more or less degenerate.^{17–21} Sometimes in order to reduce the loss of information instead of a number a sequence of numbers is used.

We reasoned that the essential of a graph, its connectivity, is preserved in its adjacency matrix and the powers thereof, that is, in the counts of walks of different length. Walk counts for a long time have found a single but very successful use in the canonical numbering of atoms in molecules by means of the Morgan algorithm.²² In 1986 walk counts were used as basic graph invariants in a Randić-type construction of TI's by Razinger.^{23b} In the present paper we want to draw the reader's attention to the counts of walks of different length starting from individual atoms, a family of fundamental descriptors for atoms in their own right, and to corresponding descriptors for molecules which, we think, deserve more exploration by chemists. It has to be emphasized that the walk counts dealt with here are the counts of *all* walks, in contrast to previous approaches which (in spite of their titles) actually deal with special walks only, self-returning walks or self-avoiding walks (paths).^{24,25}

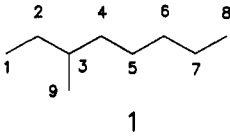
We will prove that atomic walk counts are identical to Morgan's extended connectivities, will explore the limits of their discriminating power, and will give examples of applications where walk counts are superior to existing descriptors.

ATOMIC WALK COUNTS

In our previous work we encountered the higher powers of the adjacency matrix of a graph as very useful tools for the perception of topological symmetry.²⁶ The adjacency matrix *A* of a simple graph of *n* vertices, corresponding to a saturated hydrocarbon of *n* C atoms, is a square matrix (*n* × *n*) with entries *a_{ij}* = 1 if atom *i* is connected to atom *j*, and *a_{ij}* = 0 otherwise. The *k*th power of this matrix, *A^k*, has entries *a_{ij}^(k)* which are integers giving the number of walks of length *k* bonds from atom *i* to atom *j*.²⁷ A walk is any sequence of adjacent graph edges; in walking from *i* to *j*, it is allowed to go back and forth or to visit vertices repeatedly. In contrast, a path is a sequence of adjacent graph edges without repetition. The sum of all entries in row *i* of matrix *A^k* gives the number of all walks of length *k* starting from atom *i*. We call this quantity the atomic walk count of length *k* of atom *i* and denote it by *awc_k*(*i*).

$$\text{awc}_k(i) = \sum_j a_{ij}^{(k)} \quad (1a)$$

Trivially, the number of walks of length 1 starting from atom



#	A ¹	awc ₁	A ²	awc ₂
1	0 1 0 0 0 0 0 0 0	1	1 0 1 0 0 0 0 0 0	2
2	1 0 1 0 0 0 0 0 0	2	0 2 0 1 0 0 0 0 1	4
3	0 1 0 1 0 0 0 0 1	3	1 0 3 0 1 0 0 0 0	5
4	0 0 1 0 1 0 0 0 0	2	0 1 0 2 0 1 0 0 1	5
5	0 0 0 1 0 1 0 0 0	2	0 0 1 0 2 0 1 0 0	4
6	0 0 0 0 1 0 1 0 0	2	0 0 0 1 0 2 0 1 0	4
7	0 0 0 0 0 1 0 1 0	2	0 0 0 0 1 0 2 0 0	3
8	0 0 0 0 0 0 1 0 0	1	0 0 0 0 0 1 0 1 0	2
9	0 0 1 0 0 0 0 0 0	1	0 1 0 1 0 0 0 0 1	3

#	A ³	awc ₃	A ⁴	awc ₄
1	0 2 0 1 0 0 0 0 1	4	2 0 4 0 1 0 0 0 0	7
2	2 0 4 0 1 0 0 0 0	7	0 6 0 5 0 1 0 0 4	16
3	0 4 0 4 0 1 0 0 3	12	4 0 11 0 5 0 1 0 0	21
4	1 0 4 0 3 0 1 0 0	9	0 5 0 7 0 4 0 1 4	21
5	0 1 0 3 0 3 0 1 1	9	1 0 5 0 6 0 4 0 0	16
6	0 0 1 0 3 0 3 0 0	7	0 1 0 4 0 6 0 3 1	15
7	0 0 0 1 0 3 0 2 0	6	0 0 1 0 4 0 5 0 0	10
8	0 0 0 0 1 0 2 0 0	3	0 0 0 1 0 3 0 2 0	6
9	1 0 3 0 1 0 0 0 0	5	0 4 0 4 0 1 0 0 3	12

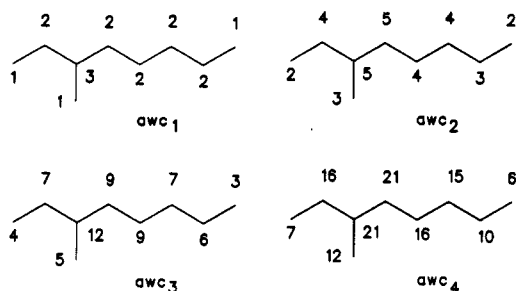


Figure 1. awc_k of 3-methyloctane. Top: Calculation as sums over rows in adjacency matrix powers. Bottom: Calculation by iterative summation over all neighbors, starting from the degrees.

i equals the number of paths of length 1, $p_1(i)$, or the number of first neighbors $n_a(i)$, which is the degree of i .

$$\text{awc}_1(i) = \sum_j a_{ij} = p_1(i) = n_a(i) = \deg i \quad (1b)$$

Figure 1 gives, as an example, the graph of 3-methyloctane (1), the first four powers of its adjacency matrix, and the walk counts of all vertices therein, up to walks of length 4. Note that vertices 2 and 5, which are endspectral and thus isocodal (have the same counts of self-returning walks),^{24,26,28} are easily differentiated by awc₃ already.

While the walk counts in Figure 1 may be obtained as sums of rows in the adjacency matrix powers, there is another and far more convenient way to obtain them. For a walk of length 1 starting at atom i , there are $\deg i$ possibilities. If we arrive at a particular atom j by a walk of length k from atom i , then there are $\deg j$ different possibilities to elongate this walk one step further. Thus summing up the degrees of all end atoms

Table I. Atomic Walk Counts, Atomic Walk Count Sums, Molecular Walk Counts, Total Walk Count, and Sequence of Ordered Atomic Walk Count Sums for 3-Methyloctane (1)

no.	awc ₁	awc ₂	awc ₃	awc ₄	awc ₅	awc ₆	awc ₇	awc ₈	awc ₉
1	1	2	4	7	16	28	65	114	237
2	2	4	7	16	28	65	114	264	500
3	3	5	12	21	49	86	199	349	724
4	2	5	9	21	37	85	149	341	649
5	2	4	9	16	36	63	142	248	520
6	2	4	7	15	26	57	99	220	430
7	2	3	6	10	21	36	78	135	291
8	1	2	3	6	10	21	36	78	157
9	1	3	5	12	21	49	86	199	376
mwc	16	32	62	124	244	490	968	1948	twc: 1942
Ordered Atomic Walk Count Sums									
157	237	291	376	430	500	520	649	724	

j of walks of length k gives the number of all walks of length $k + 1$. Thus we obtain

$$\text{awc}_1(i) = \deg i \quad (2a)$$

$$\text{awc}_2(i) = \sum_{j(i)} \deg j \quad (2b)$$

$$\text{awc}_3(i) = \sum_{j(i)} \sum_{l(j)} \deg l \quad (2c)$$

$$\text{awc}_4(i) = \sum_{j(i)} \sum_{l(j)} \sum_{m(l)} \deg m \quad (2d)$$

or

$$\text{awc}_k(i) = \sum_{j(i)} \text{awc}_{k-1}(j) \quad (2e)$$

where $j(i)$ denotes a first neighbor of atom i , and the summation index $j(i)$ says that the summation is extended over all such neighbors, and so on.

Equations 2a and 2e exactly duplicate the definition of Morgan's extended connectivities (EC's), which are obtained from the degrees of all atoms by iterative summation over all neighbors.²² As proven here, the awc's and the EC's are identical (this was first noticed by Razinger^{23a}). Therefore the awc's can be obtained very conveniently by using Morgan's summation procedure, without any matrix multiplication. This is shown for the example graph in Figure 1, bottom.

It is convenient to display the awc's as a rectangular matrix of n rows and $n - 1$ columns, corresponding to the n atoms and the lengths of walks from 1 to $n - 1$, respectively; see Table I. The Cayley-Hamilton theorem states that the n th and all higher powers of the adjacency matrix are linear combinations of the lower powers. For the purpose of discriminating vertices in graphs of n vertices, it is therefore sufficient to consider walk counts up to awc _{$n-1$} .

CHARACTERIZATION OF ATOMIC ENVIRONMENTS BY WALK COUNTS

As far as we know, no system for the unique but simple graph theoretical description of atomic environments has been proposed, not even in tree graphs. This is a task much more difficult than the unique description of trees themselves, and it is by no means sure that such a system can exist. There are far more different vertices in trees than there are different trees of a particular size (more rooted trees than trees¹²), and even the TI's for tree graphs are all known to be more or less degenerate¹⁷⁻²¹ (give the same number for two nonisomorphic

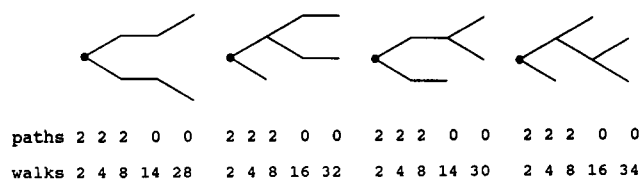


Figure 2. Some structural situations of atoms that are differentiated by walk counts but not by path counts.

trees). If such a system does exist, the code for a single atom would suffice for reconstruction of the whole tree, and it would be the natural basis for a substructure search system for atoms in tree molecules.

Randić characterized atomic environments in alkanes by a code giving the counts of paths of increasing length p_1, p_2, p_3, \dots starting from the atom i under consideration.²⁹ Path counts, however, suffer from several shortcomings.

First, whereas walk counts are easily calculated for acyclic or polycyclic graphs alike as shown above, path counts are more difficult to obtain. For acyclic graphs they require the construction of the distance matrix (which is done using the adjacency matrix powers or by the ingenious method of Trinajstić³⁰), while in cyclic graphs the paths have to be traced individually and then counted, a task increasing exponentially with increasing n . Therefore, though a human may be more used to recognizing paths rather than walks, walk counts are more fundamental and computationally more easily obtained than path counts.

Second, Randić noticed that while some of the path codes of atoms in alkanes are unique, others are not. The latter situation arises since generally there is more than one possibility of attaching a certain number of new atoms to a certain number of sites in a molecule, where the sites may or may not be structurally equivalent. In other words, not only the numbers of more remote neighbors, but their connectivities ("the distribution of neighbors"²⁹) are important.

Third, there is an upper limit for the length of paths within each molecule. Therefore, if two structural situations are not differentiated by the path counts up to this maximal length, then there is no remedy in using longer codes. Figure 2, taken from Randić's work,²⁹ gives an example of four different atomic environments having identical path codes p_1, p_2, p_3 (first line below each structure). On the other hand, even in small molecules walks of arbitrary length exist. We define the atomic walk count sequence of a vertex i in a graph as the sequence $awc_1(i), awc_2(i), \dots, awc_{n-1}(i)$. That is, atoms in trees of different size are differentiated already by the length of the sequence. The above situations (Figure 2) are in fact differentiated by the walk count sequences, as seen in the second line below each structure.

Alternatively, Randić described atomic environments focusing attention on the self-returning walks (walks starting and ending at the same atom, their numbers corresponding to the diagonal elements $a_{ii}^{(k)}$).^{24a} However, endospectral or isospectral atoms^{26,28} (2 and 5 in Figure 1 and Table I) are not differentiated by the counts of self-returning walks. If we want to look at a structure closely, there is no reason to disregard the information given by all the other walks (their numbers given by the nondiagonal elements $a_{ij}^{(k)}$).^{24b}

We tested the discriminating power of the atomic walk count sequences using a sample of tree graphs composed as follows: (a) all trees of up to 11 vertices, including the alkane graphs from methane through all undecanes (436 trees containing 3047 different vertices; most of the structures may be found in ref 12); (b) plus all 4-trees of 12 vertices (dodecane graphs, 355 trees containing 3326 different vertices; for

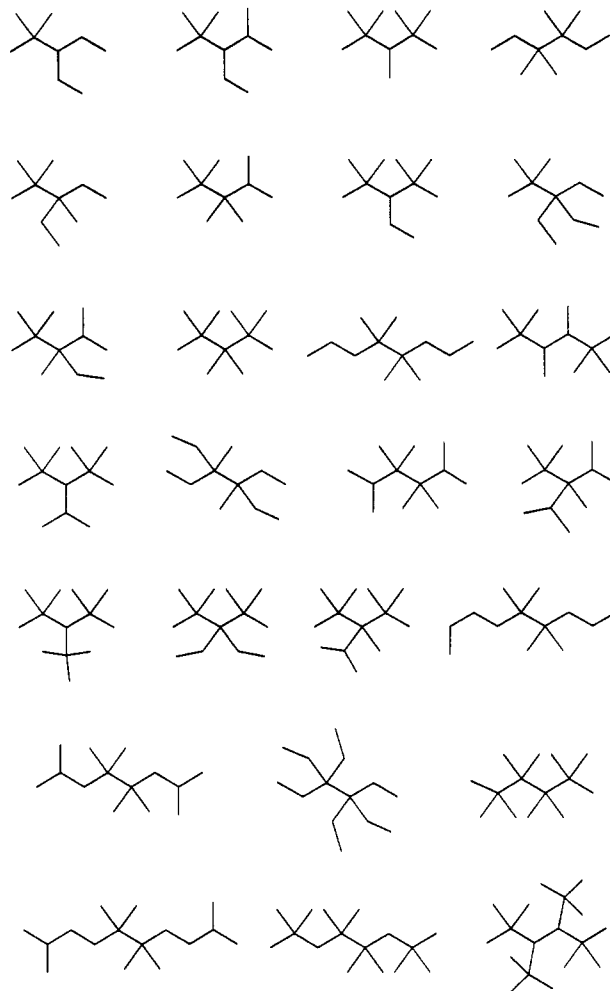


Figure 3. Highly branched Rüchardt alkanes included in the samples for testing discriminatory power of descriptors and for NMR correlations.

structures see ref 12); (c) plus the graphs of some highly branched higher alkanes, as described by Rüchardt³¹ (10 trees of size $n = 13-18$ containing 46 different vertices, structures appear in the lower part of Figure 3); (d) plus four pairs of "difficult" trees having pairwise identical values of several graph invariants, from a recent paper by Randić³² (8 trees of size $n = 15-20$ containing 81 different vertices).

All different atomic environments in this sample (total number 6500, including endospectral and isospectral atoms) are differentiated by their walk count sequences, except one pair of endospectral vertices within endospectral tree 2 ($n = 11$, Figure 4a), and three pairs of isospectral vertices in the pair of isospectral graphs 4 and 5 ($n = 12$, Figure 5a).³³

Generally, if the environments of two atoms are identical within the first m neighbor spheres, the first m awc_k 's and thus the walk count sequences up to the m th element are identical. The reverse, however, different environments have different walk count sequences, is not always true. The counterexamples shown in Figures 4a and 5a are very special endospectral/isospectral situations.^{28,34}

Analogous cases were uncovered when we additionally tested all those endospectral trees of up to 16 vertices, as given by Knop et al.³⁵ (another 47 trees containing 587 different vertices). While most of the pairs of endospectral points in these graphs are discriminated by the walk count sequences (see Figure 4b), some are not (8 cases, nos. 1120, 2450, 3105, 3120, 7638, 18690, 19164, and 19188 in ref 35). Additional examples for this situation in larger trees are found in a recent collection of trees constructed by Ivanciuc and Balaban³⁶ (8

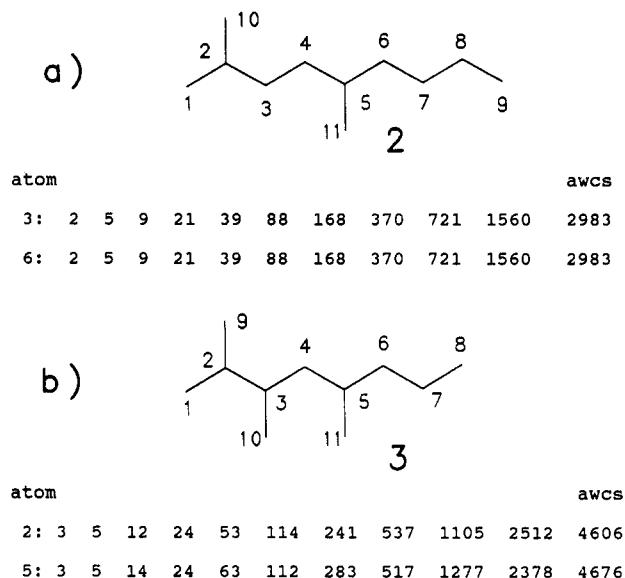


Figure 4. Examples of endospectral atoms (a) that are not and (b) that are differentiated by walk count sequences.

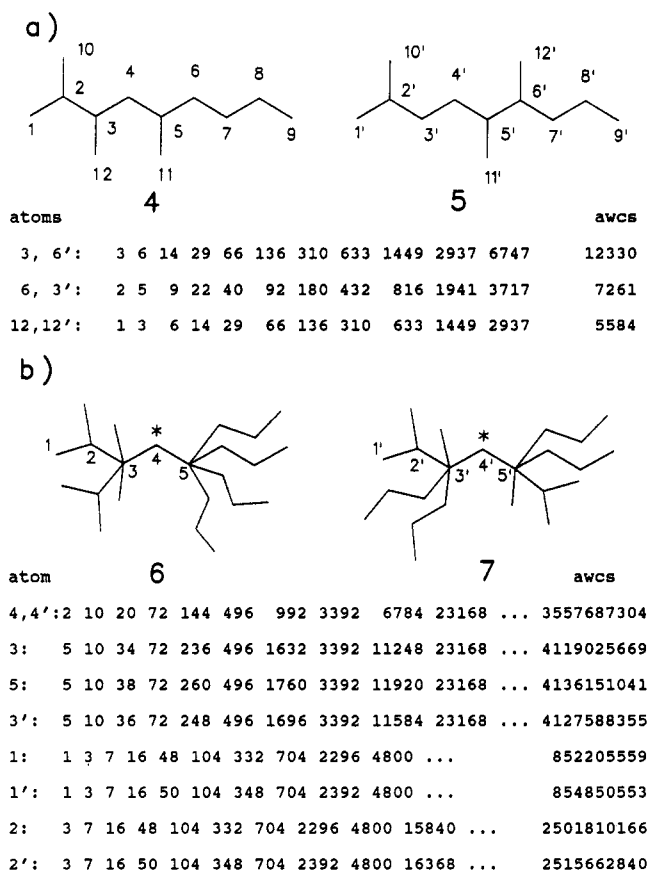
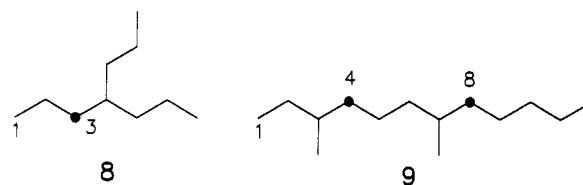


Figure 5. Examples of isospectral atoms that are not (a and first line in b) and that are (remaining cases in b) differentiated by walk count sequences and awcs numbers.

pairs of isospectral trees of size $n = 22$ – 30 , 142 different vertices). The smallest example of non-unique atomic walk count sequences from this collection is shown in Figure 5b, where it is very plausible that the isospectral vertices (marked with asterisks) cannot be discriminated by their walk counts.

We conclude that walk count sequences are highly discriminating but not unique for describing atomic environments in trees. Since the cases of non-uniqueness we were able to find are endospectral/isospectral situations, it would be tempting to conjecture that non-isospectral vertices in trees are always differentiated by walk count sequences. However,



Common elongated walk count sequence:

2 5 9 22 39 95 168 409 723 1760 3111 7573 13386 32585...

Figure 6. Common elongated walk count sequence of an atom in a decane and two endospectral atoms in a tetradecane.

the following interesting observation calls for caution: If the walk count sequences are not truncated after awc_{n-1} , then two vertices in separate trees even of different size may have identical such elongated walk count sequences; see trees 8 and 9 in Figure 6, a decane and a tetradecane. This is the only example of its type known to us; at present the mathematical significance (if any) of this coincidence is not clear.

ATOMIC WALK COUNTS AND BRANCHING

Branching of alkane trees is an old and still not completely solved issue. While the perception of branching is essentially intuitive, there are several attempts to describe branching of trees rigorously in mathematical terms.^{37–39} However, there has been no attempt to measure the branching of an atom's environment.

The walk counts of length k , being extended connectivities, effect an intra- and intermolecular ordering of atomic environments according to the extent of branching within a certain graph theoretical distance. The $awc_k(i)$ values (for constant k) increase with the degree of atom i , with the degrees of its first neighbors, etc. Therefore they increase with increasing branching and with increasing centrality, in cyclic graphs with increasing cyclicity, in multigraphs with increasing number of multiple edges. In short, the awc_k 's are a measure of the complexity of a vertex's environment.

The relative ordering of atoms by awc_k may be different for different k , according to the fact that the complexities of the closest environment and of the more remote environment of two atoms may vary independently. The ordering may even oscillate indefinitely, a problem well-known in canonical numbering by the Morgan algorithm^{22,23,40} (examples in Figure 7, intramolecular and intermolecular concurrences).

So what is the "true" order of branching in such situations? The answer depends on the rules applied for comparing sequences of numbers: A chemist may decide that any secondary carbon atom is in a more branched environment than any primary carbon atom, regardless of the number of more distant neighbors. This corresponds to an ordering by the value of the first differing position (a procedure often used in chemical documentation).⁴¹ A mathematician may decide that two sequences as given in Figure 7 are not comparable, since an element in the first sequence is sometimes larger, sometimes smaller than the corresponding element in the second sequence.

Alternatively, if a single number is to be used as a measure of branching, a suitable number may be obtained by summing the awc_k 's after a weighting procedure. We call the simple sum of all nonweighted awc_k 's of atom i (up to $k = n - 1$) the atomic walk count sum of atom i , $awcs(i)$. This number for each atom is given in the last column of Table I. It is easily seen that these numbers give an intuitively pleasing ordering of the atoms by the complexities of their respective molecular surroundings. Thus in Figure 5b the molecular environment

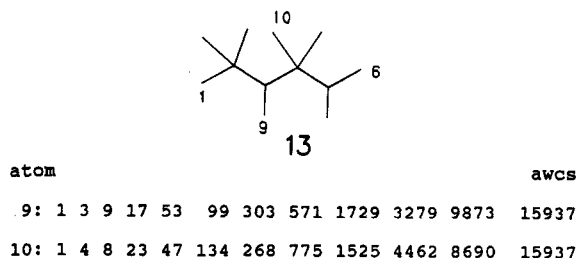
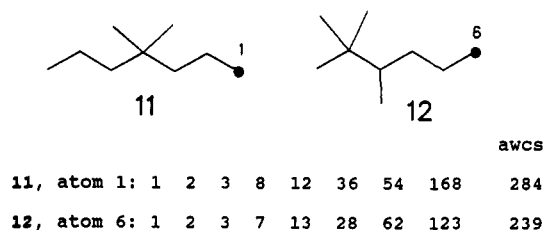
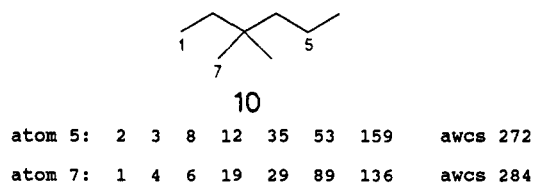


Figure 7. Examples of oscillations in the relative ordering of atoms by walk counts of increasing length.

of atom 3' (=5') in tree 7 is intermediate between those of atoms 3 and 5 in tree 6 (endospectral, and isospectral to 3'), correspondingly the walk count sequence and awcs number of the former atom is the average of those of the latter ones. Note that the endospectral atoms 2 and 5 in 1 (Table I) or the pairs of isospectral atoms 1 and 1', 2 and 2' in Figure 5b are differentiated even by their awcs numbers.

Generally, even the awcs numbers are highly discriminating among the atoms within a molecule and can therefore be used for canonical numbering. In our sample of 872 tree graphs, few cases of degeneracy of awcs within a tree were found. These are, of course, the cases of identical walk count sequences mentioned above (e.g. Figure 4a), the only other is 13 (Figure 7), where summation over differing walk counts accidentally produces identical awc's for vertices 9 and 10. The numbering induced by awcs is in most, but not all cases identical to Balaban's HOC numbering.⁴¹

CHARACTERIZATION OF TREES BY WALK COUNTS

Barysz and Trinajstić characterized trees by what they call "ordered structural code" (OSC).⁴² Such a code for a tree of n vertices is a sequence of n numbers SC_i (one for each vertex i) which are the sums of the elements $a_{ii}^{(k)}$ through the first n powers of the adjacency matrix, i.e. sums of the counts of self-returning walks. These numbers (which clearly cannot differentiate isospectral or endospectral atoms) are arranged in increasing order, and the sequences so obtained do differentiate many isospectral trees. Therefore the authors conjectured that the OSC's are unique within trees. Recently Ivanciuc and Balaban showed this conjecture to be wrong by constructing counterexamples.³⁶ We guessed that sequences of the counts of all walks (as opposed to the self-returning walks) might be more discriminant.

Razinger constructed a TI derived from walk counts by the Randić procedure: for each edge the product of the reciprocal square roots of the values for the indicent vertices is taken; then the contributions of all edges are summed.^{23b} Such a

procedure is generally known to result in very discriminating TI's (χ , J , ID), but lacks a stringent theoretical justification. We therefore decided to do without such a procedure and were amazed to find (see below) that walk counts in themselves qualify as highly discriminating descriptors for molecules as well.

From Table I it is obvious that descriptors for molecules can be obtained from the awc's in several ways, two very simple ones being as follows.

(1) We sum the elements in each column, thus obtaining a sequence of molecular walk counts of increasing length, $mwc_1, mwc_2, \dots, mwc_{n-1}$ (bottom line in Table I); then, by summing all mwc_k 's and dividing by 2 a simple integer number graph invariant, the total walk count of a graph (twc) is obtained.⁴³

$$twc = \frac{1}{2} \sum_{k=1}^{n-1} \sum_i awc_k(i) = \frac{1}{2} \sum_i \sum_j \sum_k a_{ij}^{(k)}$$

Note the formal correspondence between twc and the Wiener number

$$W = \frac{1}{2} \sum_i \sum_j d_{ij}$$

Twc can be looked upon as a measure of a tree's complexity, increasing both with increasing size and increasing branching (in contrast to many important TI's which decrease with increasing branching, such as W , χ , ID). In fact the differences in twc among the members of a family of trees (among isomeric alkanes) completely reflect Bonchev and Trinajstić's intuitive rules of branching.³⁹

(2) Alternatively we may sum the elements in each row, obtaining awcs for each atom (last column in Table I). If the latter numbers are arranged in increasing order, we obtain a sequence (the ordered atomic walk count sums sequence) which is analogous to the OSC, the difference being all walk counts vs self-returning walk counts.

We tested for uniqueness within our sample of trees (described above, 872 trees including those of Randić,³² Knop,³⁵ and Ivanciuc/Balaban³⁶) (i) the sequences of mwc 's, (ii) the two's, and (iii) the sequences of ordered atomic walk count sums.

The results are as follows:

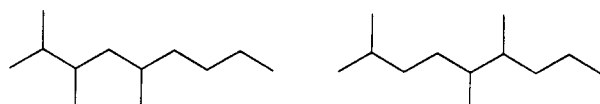
(i) All trees in our sample have different sequences of mwc 's, except the two trees in Figure 5a; see Figure 8a. Again it would be tempting to conjecture that within a family of trees (constant n) isospectral trees only may have identical mwc sequences.

(ii) Even the two's, integer number graph invariants, are highly discriminating. Notably the two's are widely different between the two members of each Ivanciuc/Balaban and each Randić pair of trees (e.g. Figure 8b).

The first and only degeneracies of twc found are in the $n = 12$ family; see the pair of trees in Figure 5a/8a and two coincidences in Figure 8c. Thus with respect to discriminating power twc is even better (at least for trees) than Balaban's index J , which has a degeneracy for a pair of trees of $n = 10$ and six degeneracies for dodecanes.¹⁹ As could be expected, due to its extreme ease of calculation twc is less discriminating than Razinger's $I(LW)$ which has no degeneracy for the dodecanes.^{23b,44}

The only case where it cannot be decided by twc or mwc sequences which one of two trees is more branched (Figure 8a) is undecidable by the Bonchev/Trinajstić rules and by the largest eigenvalue⁴⁵ either. Why is twc such a discriminating descriptor? The reason is probably that the range of integers

a)



4

5

Common sequence of mwc's:

22 48 98 212 444 958 2028 4372 9302 20046 42760 40145

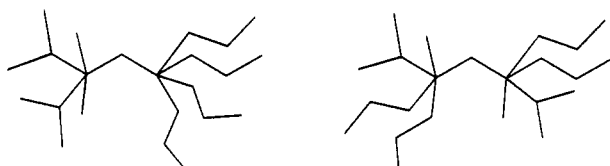
twc

Sequences of ordered awc's:

4:1501 3004 4634 4634 4957 5046 5584 7261 9804 10541 10994 12330

5:2141 2941 2941 4697 5584 6010 6485 7261 7725 9555 12330 12620

b)



6

7

OSC's:

6: 5389823⁴ 35741183⁴ 35743230⁴ 117126144² ...7: 5389823⁴ 35741183⁴ 35743230⁴ 117126144² ...

Sequences of mwc's:

6: 44 114 272 732 1824 4944 12416 33680 84736 229792 ...

7: 44 114 272 732 1824 4936 12416 33616 84736 229408 ...

twc's:

6: 18656968661

7: 18651678673

Sequences of ordered awc's:

6: 371056635⁴ 852205559⁴ 857497594⁴ 1770280966² ...7: 368411641⁴ 854550553⁴ 854852600⁴ 1778843152² ...

c)



14

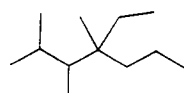
15

mwc sequences:

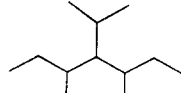
14: 22 46 92 194 394 836 1708 3638 7454 15912 32652 31474

15: 22 46 94 196 404 844 1744 3646 7542 15770 32640 31474

twc



16



17

mwc sequences:

16:22 52 112 260 574 1326 2946 6794 15132 34868 77766 69926

17:22 50 110 254 566 1308 2922 6750 15090 34850 77930 69926

twc

Figure 8. Examples of trees that are differentiated (a) not by mwc sequences, but by sequences of ordered awc's, (b) not by OSC's, but by mwc sequences, twc's and ordered sequences of awc's, and (c) by mwc sequences, but not by twc's.

accessible to the members of a family of trees of constant n increases rapidly with increasing n , even more rapidly than does the number of trees. This is shown in Table II, where

Table II. Number of Trees, Minimal and Maximal Twc for a Particular n , and Integers/Tree for All Trees of up to 11 Vertices

n	no. of trees	twc		integers/tree
		min	max	
1	1	0		
2	1	1		
3	1	5		
4	2	16	18	1
5	3	44	70	8.7
6	6	111	245	22.3
7	11	268	1 161	81.2
8	23	627	4 396	163.8
9	47	1439	25 740	517
10	106	3250	103 329	944
11	235	7259	722 215	3042

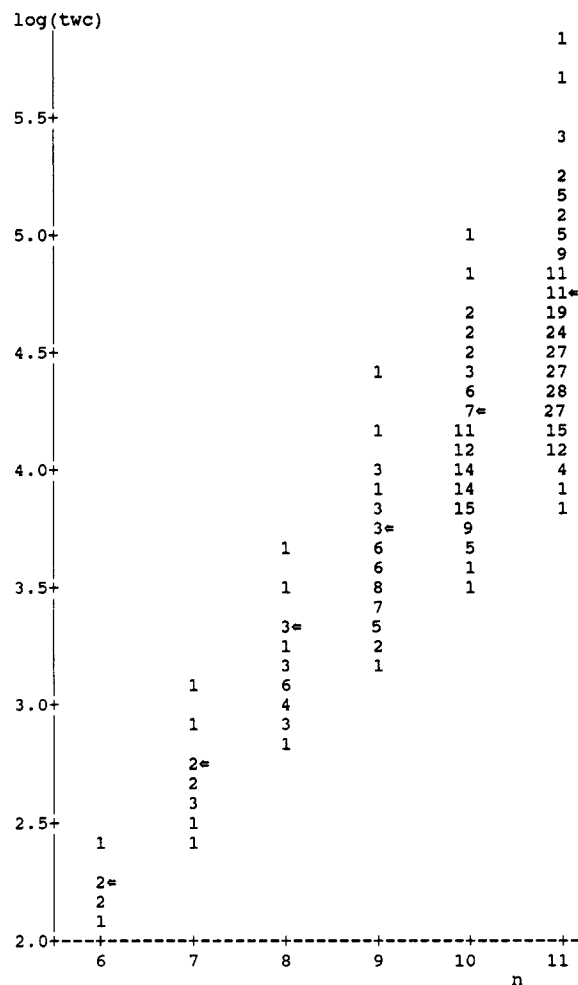


Figure 9. Plot of $\log(\text{twc})$ vs n for all trees of $n = 6-11$. Numbers in the plot are frequencies, showing the distribution of twc within a family (constant n). Arrows indicate the position of the respective most highly branched alkane.

for $n = 1-11$ the number of trees, the minimal and maximal twc in each family (chain and star graph, respectively), and the number of integers per tree (the range divided by the number of trees) are given. Thus the probability that the twc's for two nonisomorphic trees of equal n coincide decreases with increasing n , provided there is no clustering. In fact there is some clustering; see the distribution of trees over the twc range as shown in Figure 9.

Starting from $n = 7$ the ranges of twc accessible to a family overlap, which means that a smaller tree may have an identical or higher twc than a larger tree, if higher branching in the former counterbalances the size effect. This is completely logical and consistent with twc as a measure of complexity. The smallest example of identical twc's of two trees from

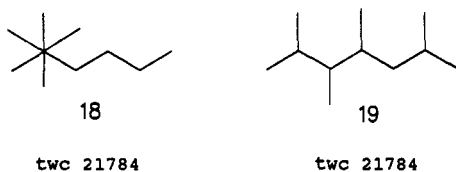


Figure 10. Two trees of different size having identical twc.

different families found in our sample is shown in Figure 10 (18 and 19).

(iii) The sequences of ordered awcs's are widely different for *all* trees in the sample, including the Ivanciuc/Balaban trees which have pairwise identical OSC's, and the trees in Figure 5a, as shown in Figure 8a,b. That is, at present we do not know of a counterexample to the uniqueness of these sequences among trees.

A graph theoretical scheme that simple must be expected to have weak sides. One weak side of the present approach is immediately seen. Among cyclic graphs the regular ones of a fixed n as well as the vertices therein cannot be discriminated, all extended connectivities being identical, e.g. in cubane, cuneane, and octabisvalene.

APPLICATIONS: CORRELATION OF NMR DATA WITH WALK COUNTS

An experimentally observable property of atoms whose value is determined by the molecular environment is the NMR chemical shift. Randić and Miyashita found a linear correlation between the ^{13}C chemical shift sum in alkanes (i.e. the average chemical shift) and the integer number graph theoretical descriptor M which is a linear combination of n and the path counts P2 and P3.^{46,47} In spite of the seemingly impressive quality of their correlation ($r = 0.999$, 96 alkanes), there is room for improvement in view of the facts that (i) three highly branched alkanes fell well apart from the straight line and were excluded and (ii) the correlation seems to be dominated by the sheer size effect ($\text{C}_2\text{--C}_{31}$). The quality of a structure-property correlation is obviously strongly de-

pendent on the particular sample of compounds used. Unfortunately the authors did not publish the structures of the alkanes in their set of compounds, but since for each size from C_{13} to C_{31} only one or two compounds were included, it has to be suspected that these compounds were the respective n -alkane and in some cases another lowly branched alkane. For higher n -alkanes a *perfect* linear correlation between the chemical shift sum and M (or even n) is to be expected. Thus the correlation found seems to be partially due to the almost trivial fact that in long unbranched (or nearly so) alkanes each CH_2 group contributes a constant increment to the ^{13}C chemical shift sum. Highly branched alkanes, on the other hand, seemed to behave distinctly different.

We wondered how walk counts perform in such a correlation, in particular, whether it is possible to obtain a common correlation for alkanes of low and of high branching. In our sample of compounds we deliberately included as many as possible highly branched alkanes and excluded long n -alkanes. Thus our sample contains methane through the butanes (5 alkanes), all pentanes through octanes and several nonanes as given by Lindeman and Adams⁴⁸ (59 alkanes), and the highly branched alkanes as described by Rüchardt et al.³¹ (26 alkanes, Figure 3). In this combined sample the best single variable for correlation with the ^{13}C chemical shift sum is mwc_2

$$\Sigma\delta = 5.98\text{mwc}_2 + 16.86$$

$$(r^2 = 0.9328, s = 27.46, F = 1221, N = 90)$$

while M ($r^2 = 0.9048, s = 32.69$) is only marginally better than n ($r^2 = 0.9043, s = 32.78$). Figure 11 gives a plot of the chemical shift sums vs mwc_2 . Different symbols are used for usual alkanes and for Rüchardt alkanes. On the basis of Figure 11, we see no reason to regard the highly branched Rüchardt alkanes as something very special.

In a two-variable model containing mwc_2 and mwc_3 , $r^2 = 0.9663$ is obtained for the same sample of alkanes:

$$\Sigma\delta = 15.88\text{mwc}_2 - 4.10\text{mwc}_3 - 32.25$$

$$(s = 19.45, F = 1247, N = 90)$$

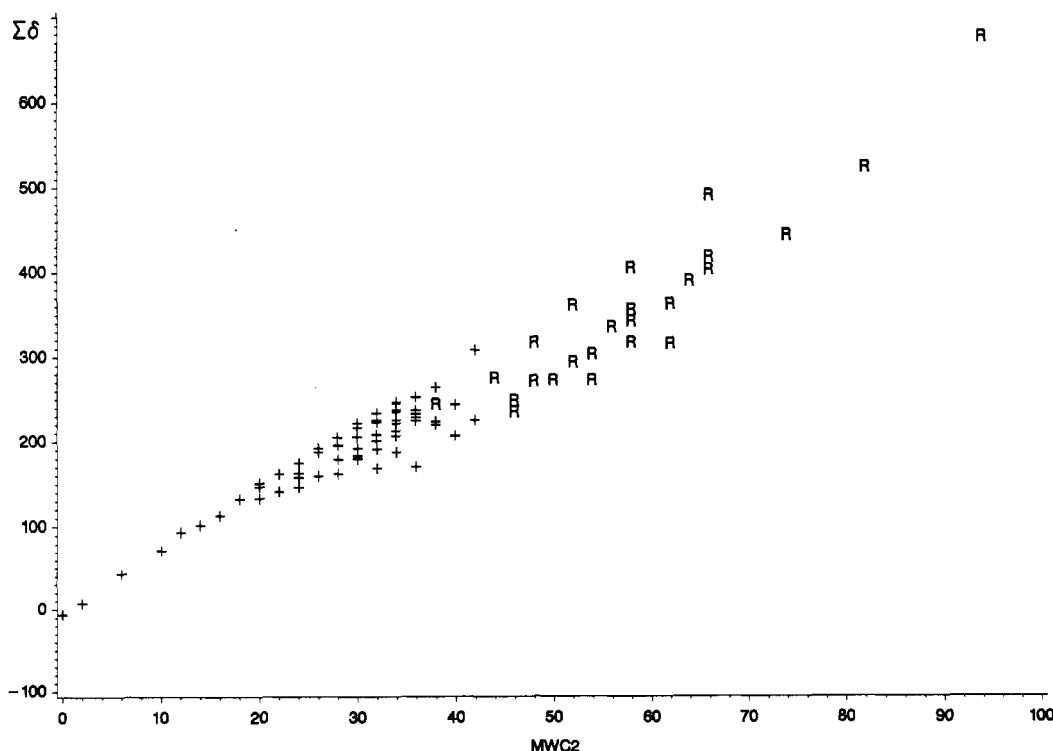


Figure 11. Plot of ^{13}C chemical shift sum vs mwc_2 for the alkanes in our sample: (+) a usual alkane; (R) a Rüchardt alkane.

awc ₁ -awc ₄					δ range	N ^a	awc ₁ -awc ₄					δ range	N ^a	awc ₁ -awc ₄					δ range	N ^a	awc ₁ -awc ₄					δ range	N ^a
1	1	1	1	5.7	1	2	2	4	4	15.9	1	3	3	9	9	25.0	1	4	4	16	16	28.0	1				
1	2	2	4	15.4	1	2	3	5	8	24.8	1	3	4	10	14	29.9	1	4	5	17	22	30.3	1				
1	2	3	5	13.0	1	2	3	6	9	22.2	1	3	4	11	15	27.9	1	4	5	18	23	30.6	1				
1	2	3	6	13.5-14.3	17	2	3	6	10	22.6-22.8	8	3	4	11	16	28.0-28.1	4	4	5	18	24	30.1-30.2	2				
1	2	3	7	13.9-14.3	8	2	3	6	11	23.0-23.3	5	3	4	11	17	28.4-28.4	2	4	5	18	25	30.1	1				
1	2	3	8	14.8-15.4	5	2	3	6	12	23.7-24.1	3	3	4	11	18	28.9-29.2	2	4	5	19	24	30.9	1				
1	2	4	6	11.5	1	2	3	7	10	20.8	1	3	4	12	16	25.7	1	4	5	19	25	31.0	1				
1	2	4	7	10.9-11.4	8	2	3	7	11	19.4-20.2	4	3	4	12	17	25.3-25.4	2	4	5	20	25	32.4	1				
1	2	4	8	10.5-11.9	6	2	3	7	12	20.8-20.8	3	3	4	12	18	25.7	1	4	5	20	28	32.3	1				
1	2	4	9	11.8-13.0	2	2	3	8	11	18.1	1	3	4	13	17	25.3	1	4	6	18	28	32.3	1				
1	2	4	10	14.4-14.5	2	2	3	8	12	17.3-17.3	2	3	4	13	20	24.3	1	4	6	18	30	32.7	1				
1	2	4	11	17.1	1	2	3	8	13	17.0	1	3	5	11	19	36.8	1	4	6	19	29	32.8	1				
1	2	4	12	18.8	1	2	3	8	14	18.0	1	3	5	11	21	34.0	1	4	6	19	30	32.5	1				
1	2	5	7	8.5	1	2	4	6	12	34.1	1	3	5	12	20	34.3	1	4	6	19	31	33.0	1				
1	2	5	8	7.7-8.1	3	2	4	6	14	31.6	1	3	5	12	21	34.6-34.7	2	4	6	20	30	32.8	1				
1	2	5	9	7.5-7.9	2	2	4	7	13	31.7	1	3	5	12	22	31.9-34.8	2	4	6	20	32	34.0-34.1	2				
1	2	5	10	7.1-9.3	3	2	4	7	14	32.0-32.1	4	3	5	12	23	32.2-32.8	2	4	6	20	35	34.5	1				
1	2	5	11	10.4-10.5	2	2	4	7	15	29.4-32.4	3	3	5	12	24	32.4	1	4	6	21	33	35.0-35.7	2				
1	2	5	12	10.5-10.9	3	2	4	7	16	29.5-33.0	5	3	5	13	21	32.1-32.3	2	4	6	22	34	37.1	1				
1	2	5	14	13.4	1	2	4	7	17	29.9-30.0	2	3	5	13	22	32.0-32.6	2	4	6	23	35	37.6	1				
1	3	3	9	24.1	1	2	4	7	18	31.0	1	3	5	13	23	29.1-29.8	2	4	6	23	41	37.0	1				
1	3	4	10	21.9	1	2	4	8	14	29.0-29.7	2	3	5	14	22	30.2-31.9	2	4	6	24	36	38.3	1				
1	3	4	11	22.3-22.9	9	2	4																				

^a N = number of observations.

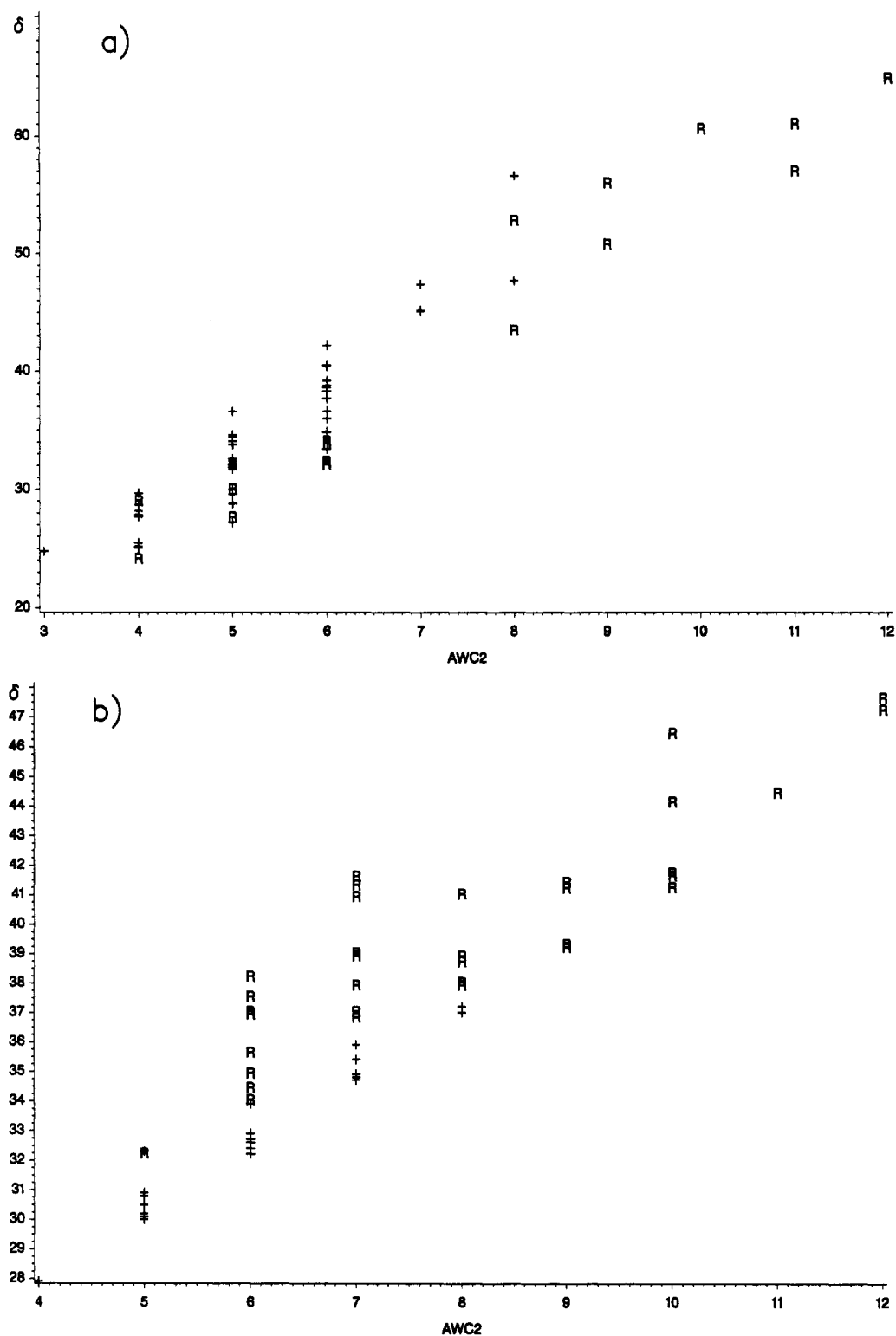


Figure 12. Plot of ^{13}C chemical shift vs awc_2 for (a) the 3° and (b) the 4° C atoms in our sample: (+) an atom in a usual alkane; (R) an atom in a Rüchardt alkane.

Table IV

Lindeman/Adams type scheme				walk count based scheme		
	variables	r^2	s	N	variables	r^2
1° C	$n_\gamma, n_\delta, n_{\alpha 2}, n_{\alpha 3}, n_{\alpha 4}$	0.9850	0.746	125	$\text{awc}_2, \text{awc}_3, \text{awc}_4, \text{sum}_2, \text{sum}_3$	0.9874
2° C	$n_\gamma, n_\delta, n_{\alpha 2}, n_{\alpha 3}, n_{\alpha 4}$	0.9955	0.571	117	$\text{awc}_2, \text{awc}_3, \text{awc}_4, \text{sum}_2, \text{sum}_3$	0.9957
3° C	$n_\gamma, n_{\alpha 2}, n_{\alpha 3}, n_{\alpha 4}$	0.9841	0.838	53	$\text{awc}_2, \text{awc}_3, \text{sum}_2, \text{sum}_3$	0.9847
4° C	$n_\gamma, n_{\alpha 2}, n_{\alpha 3}, n_{\alpha 4}$	0.9851	0.313	24	$\text{awc}_2, \text{awc}_3, \text{sum}_2, \text{sum}_3$	0.9921

Thus though our sample includes many "difficult" highly branched alkanes and takes less profit from the trivial correlation with n than the sample of ref 47 did, our approach leads to a good correlation and thus illustrates the value of walk counts.

Table V

		r^2	s
1° C	$\text{awc}_2, \text{awc}_3, \text{awc}_4, \text{sum}_2$	0.9850	0.746
4° C	$\text{awc}_2, \text{sum}_2, \text{sum}_3$	0.9908	0.246
or even	$\text{awc}_2, \text{sum}_3$	0.9867	0.296

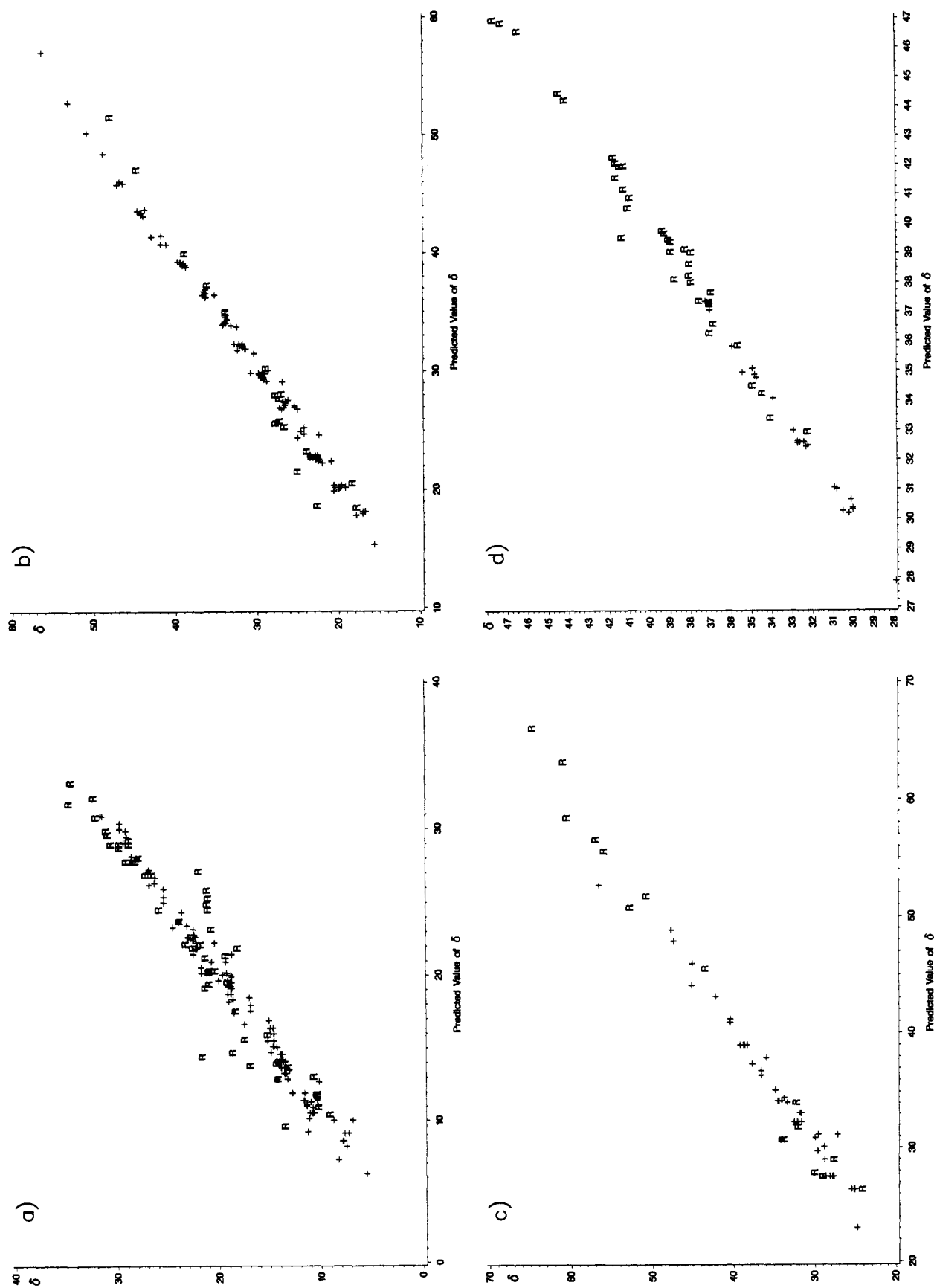


Figure 13. Plot of actual chemical shift vs chemical shift calculated using eq 3 for (a) 1° , (b) 2° , (c) 3° , and (d) 4° C atoms. Symbols are as in Figure 12.

Table VI

	Lindeman/Adams type scheme			<i>N</i>	walk count based scheme		
	variables	<i>r</i> ²	<i>s</i>		variables	<i>r</i> ²	<i>s</i>
1° C	<i>n</i> _γ , <i>n</i> _δ , <i>n</i> _{α2} , <i>n</i> _{α3} , <i>n</i> _{α4}	0.9467	1.546	185	awc ₂ , awc ₃ , awc ₄ , awc ₅ , sum ₂	0.9567	1.393
2° C	<i>n</i> _γ , <i>n</i> _δ , <i>n</i> _{α2} , <i>n</i> _{α3} , <i>n</i> _{α4}	0.9863	0.993	138	awc ₂ , awc ₃ , awc ₄ , awc ₅ , sum ₂	0.9874	0.952
3° C	<i>n</i> _γ , <i>n</i> _{α2} , <i>n</i> _{α3} , <i>n</i> _{α4}	0.9770	1.454	71	awc ₂ , awc ₃ , sum ₂ , sum ₃	0.9807	1.332
4° C	<i>n</i> _γ , <i>n</i> _{α2} , <i>n</i> _{α3} , <i>n</i> _{α4}	0.9739	0.732	60	awc ₂ , awc ₃ , awc ₄ , awc ₅	0.9898	0.458

Clearly, however, one would be much more interested in the individual chemical shifts of atoms.

Grant and Paul⁴⁹ and Lindeman and Adams⁴⁸ (L/A) in their classical work on the ¹³C NMR shifts of alkanes very successfully described the environment of an atom in a molecule in terms of the graph theoretical descriptors *n*_α, *n*_β, *n*_γ, ..., meaning the number of carbon atoms 1, 2, 3, ... bonds away from the atom under consideration, and, less coarsely, of *n*_{α1}, *n*_{α2}, ..., meaning the number of primary, secondary, ... atoms one bond away. For trees, these descriptors are identical with path counts.

Randić used the path counts *p*₁, *p*₂, *p*₃ in a classification scheme for carbon atoms.²⁹ Within a so-defined class the δ values showed a span of 6.3 ppm at most for the L/A data. He noticed that the path codes often do not differentiate structural situations, and that in such cases it makes a difference how the neighbors in a particular sphere are distributed over the neighbors in the innermost sphere. We found that this information can be coded as a quantity closely related to the variance of the degrees of the inner-sphere neighbors, namely, the sum of the squares of the degrees of these atoms:

$$\text{sum}_2(i) = \sum_{j(i)} (\text{deg } j)^2, \quad \text{sum}_3(i) = \sum_{j(i)} \sum_{l(j)} (\text{deg } l)^2$$

Alternatively, this information is contained in the higher walk counts. In fact, it can be derived from eq 2 that sum₂ is a component of awc₄, sum₂ and sum₃ are components of awc₅, etc.

As expected, classification of carbon atom structural situations by walk counts results in grouping by chemical shift values also. Within a class defined by awc₁, awc₂, awc₃, and awc₄, the maximal span of δ values is 3.2 ppm for primary (1°), 4.8 ppm for secondary (2°), 2.9 ppm for tertiary (3°), and 1.9 ppm for quaternary (4°) C atoms for the combined data set of Lindeman/Adams and the structurally diverse Rüchardt alkanes (Table III).⁵⁰

Though a classification scheme allows some predictions to be made, an equation is clearly more desirable. Therefore a linear regression model was tried. The most important variables are awc₁ and awc₂. In Figure 12a,b for 3° and 4° C atoms the dependence of δ on awc₂ is shown.⁵¹ Figure 12 demonstrates (i) how wide a range of alkane structural situations was pioneered by the Rüchardt group, (ii) that chemical shifts in the Rüchardt alkanes show a more pronounced scattering but at least with respect to awc₁ and awc₂ do not systematically deviate from the trends found in usual alkanes, and (iii) what chemical shifts can be expected for even more highly branched alkanes (theoretically, awc₂ for 4° C atoms may be as high as 16).

The Lindeman/Adams scheme may be expressed as four linear equations (one for 1°, 2°, 3°, 4° carbon atoms, respectively) in five (1°, 2° C) or four (3°, 4° C) variables, containing 22 parameters altogether. We first asked the following question: How do walk counts instead of path counts perform in such a correlation for the original L/A data set?

Table VII

		<i>r</i> ²	<i>s</i>
1° C	awc ₂ , awc ₃ , awc ₅ , sum ₂	0.9496	1.503
2° C	awc ₂ , awc ₃ , awc ₅ , sum ₂	0.9874	0.952
4° C	awc ₂ , awc ₃ , awc ₅	0.9838	0.577
or even	awc ₂ , awc ₅	0.9807	0.629

In a pool of possible L/A descriptors we included *n*_β, *n*_γ, *n*_δ, *n*_ε, *n*_{α2}, *n*_{α3}, and *n*_{α4}. In a pool of walk-based descriptors we included awc₂, awc₃, awc₄, awc₅, sum₂, and sum₃. All multilinear least-squares correlations of the ¹³C chemical shift vs one to five of these descriptors were generated, separately for 1°, 2°, 3°, and 4° carbon atoms (i.e. stratification by *n*_α or awc_i), and separately for the two descriptor pools. The best correlations using as many variables as in the original Lindeman/Adams scheme are shown in Table IV (*N* = number of observations).⁵²

Thus in all these cases the walk counts are marginally better descriptors than the path counts. For two of the four types of C atoms a walk-based correlation using fewer variables is as good or better in terms of *r*² or *s* than the Lindeman/Adams type scheme (see Table V).

Encouraged by these results we next asked how the walk count based descriptors behave for the highly branched Rüchardt alkanes, for which the Lindeman/Adams scheme is known sometimes to give large deviations between calculation and experiment. We therefore generate the correlations as above for the complete set of compounds (methane through butanes plus Lindeman/Adams plus Rüchardt alkanes)⁵² (Table VI).

Again for three of the four types of C atoms fewer variables can be used without loss in *r*² as compared to the Lindeman/Adams type scheme (Table VII). Thus the highly branched Rüchardt alkanes pose problems to both schemes, which is reflected in lower *r*² (larger deviations between calculation and experiment), while the walk-based descriptors are still better than the paths.

The best regression equations for the complete data set are as follows:

$$1^\circ: \quad \delta = 16.46\text{awc}_2 - 5.65\text{awc}_3 + 0.52\text{awc}_4 + 0.40\text{awc}_5 - 1.44\text{sum}_2 - 4.02$$

$$(r^2 = 0.9567, s = 1.393, F = 791, N = 185) \quad (3a)$$

$$2^\circ: \quad \delta = 16.72\text{awc}_2 - 3.49\text{awc}_3 + 0.16\text{awc}_5 - 1.48\text{sum}_2 - 2.42$$

$$(r^2 = 0.9874, s = 0.952, F = 2061, N = 138) \quad (3b)$$

$$3^\circ: \quad \delta = 12.84\text{awc}_2 - 3.83\text{awc}_3 - 0.96\text{sum}_2 + 0.54\text{sum}_3 + 7.14$$

$$(r^2 = 0.9807, s = 1.332, F = 838, N = 71) \quad (3c)$$

$$4^\circ: \quad \delta = 2.99\text{awc}_2 - 2.02\text{awc}_3 - 0.29\text{awc}_4 + 0.34\text{awc}_5 + 31.12$$

$$(r^2 = 0.9898, s = 0.458, F = 1341, N = 60) \quad (3d)$$

For comparison, s values for the original L/A scheme for the same data set are 2.064 for 1° C, 1.187 for 2° C, 1.965 for 3° C, and 1.085 for 4° C. Parts a–d of Figure 13 give plots of actual chemical shifts vs shifts calculated by eqs 3a–d for all the data.

It was not possible to find a satisfactory common correlation for all four types of C atoms.

CONCLUSION

In our opinion, introduction of another couple of graph theoretical descriptors, as proposed here, is justified by the fundamental structural nature of the walks. We feel that the walk counts are a reasonable compromise between conceptual and calculational simplicity on the one hand and discriminating power and correlational ability on the other.

EXPERIMENTAL SECTION

The program MORGAN is written in FORTRAN 77. It calculates all the information given in Table I for a given graph. Since FORTRAN handles integer variables up to $2^{31}-1$ only, the variables awc_k , $awcs$, $mwcs$, and twc were defined as real*8 variables, thus assuring exactness even for trees of 30 vertices. When run on the IBM 3090 computer in the Universität Freiburg Computer Center, MORGAN used 0.016, 0.016, 0.025, and 0.041 s of CPU time for a tree of $n = 9$ (Table I), 10, 20, and 30 vertices, respectively.

Sequences of numbers were sorted by the first, second, ... element using the statistics software SAS; then every two successive such sequences were tested for identity. Regression analyses were performed using SAS.

The ^{13}C NMR data were taken from refs 48 (solvent dioxane) and 31 (solvent $\text{CDCl}_3/\text{CCl}_4$). For three compounds (2,2,4,4-tetramethylpentane, 2,2,3,3-tetramethylpentane, and 2,2,3,4-tetramethylpentane, 15 different C atoms) both refs 31 and 48 report a ^{13}C NMR spectrum. For 13 atoms the difference in δ between the two reports is 0.4 ppm or less; for the remaining atoms the difference is 0.6 and 2.9 ppm. In these cases the L/A data were used. It was decided that the data from these two sources could be pooled. The assignment for the 3° and the 4° C in 2,3,3,4-tetramethylpentane in ref 48 is obviously erroneous; it was therefore interchanged. The data for tri-*tert*-butylmethane were taken from ref 31f rather than ref 31d. For this choice see also ref 53. Where more than one value is given for a particular C atom (due to stereoisomerism) in a source, the average was taken.

REFERENCES AND NOTES

- Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vol. 2.
- Rouvray, D. H. The Modeling of Chemical Phenomena Using Topological Indices. *J. Comput. Chem.* **1987**, *8*, 470–480. Rouvray, D. H. The Limits of Applicability of Topological Indices. *J. Mol. Struct. (Theochem)* **1989**, *185*, 187–201.
- Balaban, A. T. Topological Indices and Their Uses: A New Approach for the Coding of Alkanes. *J. Mol. Struct. (Theochem)* **1988**, *165*, 243–253.
- Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indices in Organic Chemistry. *Russ. Chem. Rev.* **1988**, *57*, 337–366.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, U.K., 1986.
- Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2, Chapter 9.
- Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164–175.
- Randić, M.; Hansen, P. J.; Jurs, P. C. Search for Useful Graph Theoretical Invariants of Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60–68.
- Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155–168. Randić, M. In Search of Structural Invariants. *J. Math. Chem.* **1992**, *9*, 97–146.
- Randić, M. Chemical Structure-What is She? *J. Chem. Educ.* **1992**, *69*, 713–718.
- Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701–712.
- Trinajstić, N.; Nikolić, S.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Computational Chemical Graph Theory*; Ellis Horwood: Chichester, U.K., 1991.
- 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. Diudea, M. V.; Minailiuc, O.; Balaban, A. T. Regressive Vertex Degrees (New Graph Invariants) and Derived Topological Indices. *J. Comput. Chem.* **1991**, *12*, 527–535. Balaban, A. T. Using Real Numbers as Vertex Invariants for Third-Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23–28.
- Hall, L. H.; Mohnney, B.; Kier, L. B. The Electrotopological State: An Atom Index for QSAR. *Quant. Struct.-Act. Relat.* **1991**, *10*, 43–51. Hall, L. H.; Mohnney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76–82. Bonchev, D.; Kier, L. B. Topological Atomic Indices and the Electronic Charge in Alkanes. *J. Math. Chem.* **1992**, *9*, 75–85.
- Ivanciuc, O. Definition of a Vertex Topological Index in the Case of 4-Trees. *Rev. Roum. Chim.* **1989**, *34*, 1361–1368.
- Filip, P. A.; Balaban, T. S.; Balaban, A. T. A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlation Ability. *J. Math. Chem.* **1987**, *1*, 61–83.
- Szymanski, K.; Müller, W. R.; Knop, J. V.; Trinajstić, N. On the Identification Numbers for Chemical Structures. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1986**, *20*, 173–183.
- Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. On the Determinant of the Adjacency-Plus-Distance Matrix as the Topological Index for Characterizing Alkanes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 83–84.
- Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404. Balaban, A. T.; Quintas, L. V. The Smallest Graphs, Trees, and 4-Trees with Degenerate Topological Index. *J. Math. Chem.* **1983**, *14*, 213–233.
- Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer Discrimination by Topological Information Approach. *J. Comput. Chem.* **1981**, *2*, 127–148.
- Razinger, M.; Chrétien, J. R.; Dubois, J. E. Structural Selectivity of Topological Indexes in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23–27.
- Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
- (a) Razinger, M. Extended Connectivity in Chemical Graphs. *Theor. Chim. Acta* **1982**, *61*, 581–586. (b) Razinger, M. Discrimination and Ordering of Chemical Structures by the Number of Walks. *Theor. Chim. Acta* **1986**, *70*, 365–378.
- (a) Randić, M. Random Walks and Their Diagnostic Value for Characterization of Atomic Environments. *J. Comput. Chem.* **1980**, *1*, 386–399. (b) Randić, M.; Woodworth, W. L.; Graovac, A. Unusual Random Walks. *Int. J. Quantum Chem.* **1983**, *24*, 435–452.
- Shalabi, A. S. Random Walks: Computations and Applications to Chemistry. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 483–491.
- Rücker, G.; Rücker, Ch. Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relationships among Atoms into Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 187–191. Rücker, G.; Rücker, Ch. On Using the Adjacency Matrix Power Method for Perception of Symmetry and for Isomorphism Testing of Highly Intricate Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 123–126. Rücker, G.; Rücker, Ch. Isocodal and Isospectral Points, Edges, and Pairs and How to Cope with Them in Computerized Symmetry Recognition. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 422–427.
- Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969. Gould, R. *Graph Theory*; Benjamin/Cummings: Menlo Park, CA, 1988.
- Rücker, Ch.; Rücker, G. Understanding the Properties of Isospectral Points and Pairs in Graphs: The Concept of Orthogonal Relation. *J. Math. Chem.* **1992**, *9*, 207–238.
- Randić, M.; Wilkins, C. L. Graph Theoretical Ordering of Structures as a Basis for Systematic Searches for Regularities in Molecular Data. *J. Phys. Chem.* **1979**, *83*, 1526–1540. Randić, M. Survey of Structural Regularities in Molecular Properties. I. Carbon-13 Chemical Shifts in Alkanes. *Int. J. Quantum Chem.* **1983**, *23*, 1707–1722.
- Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. An Algorithm for Construction of the Molecular Distance Matrix. *J. Comput. Chem.* **1987**, *8*, 170–173. For a review see: Mihalić, Z.; Veljan, D.; Amić, D.; Nikolić, S.; Plavšić, D.; Trinajstić, N. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223–258.
- (a) Winiker, R.; Beckhaus, H.-D.; Richardt, C. Thermal Stability, Strain Enthalpy, and Structure of Symmetrical Hexaalkyl-substituted Ethanes. *Chem. Ber.* **1980**, *113*, 3456–3476. (b) Hellmann, G.; Hellmann, S.

- Beckhaus, H.-D.; Rüchardt, C. Thermal Stability, Strain Enthalpy, and Structure of Symmetrical Tetrasubstituted Ethanes. *Chem. Ber.* **1982**, *115*, 3364–3383. (c) Beckhaus, H.-D.; Hellmann, G.; Rüchardt, C. Tetra-*tert*-butylethane. *Chem. Ber.* **1978**, *111*, 72–83. (d) Hellmann, S.; Beckhaus, H.-D.; Rüchardt, C. Syntheses, Spectra, Structures, and Strain of Highly Branched Pentanes. *Chem. Ber.* **1983**, *116*, 2219–2237. (e) Beckhaus, H.-D.; Rüchardt, C.; Anderson, J. E. Conformational Isomerism and Dynamic Gearing in Ethanes with Many Alkyl Substituents. *Tetrahedron* **1982**, *38*, 2299–2309. (f) Flamm-ter Meer, M. Ph.D. Thesis, Universität Freiburg, 1984.
- (32) Randić, M. Representation of Molecular Graphs by Basic Graphs. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 57–69.
- (33) 4 and 5 are trivially derived from 2 by attaching a single vertex at one or the other of the endospectral vertices. Interestingly, this same pair of trees was earlier found to have several other vertex and graph invariants degenerate.¹⁶
- (34) Vertices are isospectral/endospectral if they are related by at least one orthogonal matrix.²⁸ The following can be proven: Isospectral/endospectral vertices have identical walk count sequences if and only if in at least one of these matrices each sum over a row (or over a column) is equal to 1.
- (35) Knop, J. V.; Szymanski, K.; Trinajstić, N.; Kleiner, A. F.; Randić, M. On Irreducible Endospectral Graphs. *J. Math. Phys.* **1986**, *27*, 2601–2612.
- (36) Ivanciuc, O.; Balaban, A. T. Nonisomorphic Graphs with Identical Atomic Counts of Self-Returning Walks: Isocodal Graphs. *J. Math. Chem.* **1992**, *11*, 155–167.
- (37) Rouvray, D. H. The Challenge of Characterizing Branching in Molecular Species. *Discrete Appl. Math.* **1988**, *19*, 317–338.
- (38) Bertz, S. H. Branching in Graphs and Molecules. *Discrete Appl. Math.* **1988**, *19*, 65–83.
- (39) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix, and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.
- (40) Randić, M. On Unique Numbering of Atoms and Unique Codes for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 105–108. Rücker, Ch.; Rücker, G., Mathematical Relation between Extended Connectivity and Eigenvector Coefficients. *J. Math. Chem.*, in press.
- (41) Mekenyan, O.; Bonchev, D.; Balaban, A. T. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). New Topological Indices, Ordering of Graphs, and Recognition of Graph Similarity. *J. Comput. Chem.* **1984**, *5*, 629–639. Balaban, A. T.; Mekenyan, O.; Bonchev, D. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538–551.
- (42) Barysz, M.; Trinajstić, N. A Novel Approach to the Characterization of Chemical Structures. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1984**, *18*, 661–673.
- (43) It can generally be proven that each mwc_k is an even integer, and so is their sum.
- (44) Twc was also tested on all mono- and bicycloalkane graphs of 4 to 8 vertices, as given in ref 20, and found to exhibit low degeneracy: the mean isomer degeneracy in these cyclic graphs for $n = 4$ is 1.00, $n = 5$ 1.00, $n = 6$ 1.036, $n = 7$ 1.037, and $n = 8$ 1.082.
- (45) Cvetković, D. M.; Gutman, I. Note on Branching. *Croat. Chem. Acta* **1977**, *49*, 115–121.
- (46) Randić, M. Chemical Shift Sums. *J. Magn. Reson.* **1980**, *39*, 431–436. Randić, M.; Trinajstić, N. Composition as a Method for Data Reduction: Application to Carbon-13 NMR Chemical Shifts. *Theor. Chim. Acta* **1988**, *73*, 233–246.
- (47) Miyashita, Y.; Okuyama, T.; Ohsako, H.; Sasaki, S. Graph Theoretical Approach to Carbon-13 Chemical Shift Sum in Alkanes. *J. Am. Chem. Soc.* **1989**, *111*, 3469–3470. Miyashita, Y.; Ohsako, H.; Okuyama, T.; Sasaki, S.; Randić, M. Computer-Assisted Studies of Structure-Property Relationships Using Graph Invariants. *Magn. Reson. Chem.* **1991**, *29*, 362–365.
- (48) Lindeman, L. P.; Adams, J. Q. Carbon-13 Nuclear Magnetic Resonance Spectrometry. Chemical Shifts for the Paraffins through C_9 . *Anal. Chem.* **1971**, *43*, 1245–1252.
- (49) Grant, D. M.; Paul, E. G. Carbon-13 Magnetic Resonance. II. Chemical Shift Data for the Alkanes. *J. Am. Chem. Soc.* **1964**, *86*, 2984–2990.
- (50) An alternative classification by awc_1 , awc_2 , awc_3 , sum_2 gives even better results for 2° and 3° C atoms (maximal span 2.4 and 1.7 ppm, respectively), but not for 1° C atoms, where sum_2 is obviously not an independent variable, nor for 4° C atoms.
- (51) In Figure 12 the abscissa can be calibrated in p_2 also, since for trees $p_2 = awc_2 - awc_1$.
- (52) Note that the r^2 values given in the left halves of Tables IV and VI are those of the best possible L/A type correlations, not those of the original L/A scheme.
- (53) Lomas, J. S. Tri(1-adamantyl)methane and Its Thermolysis. *J. Chem. Soc., Perkin Trans. 2* **1992**, 1531–1532. Wroczynski, R. J.; Mislow, K. NMR Evidence for C_3 Ground-State Conformations and Multiple Site Exchange Processes in Systems of the Type $t-Bu_3MX$. *J. Am. Chem. Soc.* **1979**, *101*, 3980–3981.