Walk Counts, Labyrinthicity, and Complexity of Acyclic and Cyclic Graphs and Molecules

Gerta Rücker and Christoph Rücker*

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

Received June 22, 1999

It is demonstrated how the complexity of a (molecular) graph can be quantified in terms of the walk counts, extremely easily obtained graph invariants that depend on size, branching, cyclicity, and edge and vertex weights (unsaturation, heteroatoms). The influence of symmetry is easily accounted for. The term *labyrinthicity* is proposed for what is measured by walk counts alone, neglecting symmetry. The total walk count and recently advanced measures of labyrinthicity or complexity are compared with respect to the ordering of structures and to the computational effort required to obtain numerical values.

INTRODUCTION

The intuitive concept of molecular complexity gained interest in recent years for evaluation of strategies of synthesis.[1−3] However, like many useful concepts in chemistry, molecular complexity is not easy to define or to quantify. Most authors agree that the complexity of a molecule increases with increasing size, increasing branching, and increasing cyclicity for acyclic and cyclic structures, respectively, with increasing unsaturation (presence of multiple bonds), with increasing presence of heteroatoms, and with decreasing symmetry. Quantification of molecular complexity therefore requires quantification of the above concepts. This was attempted several times in graph-theoretic and information-theoretic terms, as reviewed recently in detail by Bonchev[4] and briefly by Randić.[5]

Unfortunately, a physical property of molecules that could serve as a measure of molecular complexity is not known, so there is some freedom for personal taste in the construction of a theoretical measure. However, it is clear that any measure should be able to take into account the features given above, that it should be firmly based on mathematics, and that it should order as many molecular species as possible in a manner that is self-consistent and logical and appears reasonable to chemists.

In 1975 Randić introduced as "the branching index" the quantity now known as the connectivity index χ, an edge-additive number that decreases with increasing branching.[6] Bonchev and Trinajstić in the late 1970s observed similar behavior of the Wiener number W (the sum of all distances in a graph). They explicitly formulated rules for branching in acyclic saturated structures, which essentially are rules on the behavior of W for particular structural changes, where a decrease in W was taken as an increase in branching in obvious as well as in less obvious situations.[7a] Similarly, for cyclic saturated isomeric structures, rules on cyclicity were given, in which a decrease in W was identified with an increase in cyclicity.[8a−c] For increasing size, however, W and χ increase (as do most other TIs), and therefore W or χ cannot be used as a general inverse measure of molecular complexity. Moreover, like many other graph-theoretical indices, W and χ are rather degenerate, that is, many nonisomorphic graphs (corresponding, e.g., to isomers) exhibit the same numerical value. The rules on branching[7b,9] and on cyclicity[8d,10] were subsequently revised and partially modified on the basis of other graph-theoretical invariants.

In 1977 Cvetković and Gutman showed that an earlier proposed graph-theoretical invariant, the first (largest) eigenvalue λ_1 of a graph's adjacency matrix, is a measure of branching. These authors were even able to demonstrate a connection between λ_1 and the walks in a graph (see below).[11]

The number of all paths in a graph or molecule was suggested as a complexity measure by Randić in 1979.[12] The number of paths of length 2 ($p2 = \eta =$ number of "connections") was first used in the 1980s by Bertz as an easily obtained graph invariant, which nevertheless is sensitive to branching and modifiable for the presence of unsaturation.[1]

The number of spanning trees, a special class of subgraphs, was proposed as a measure of a (cyclic) graph's complexity in 1983[13] and recently again.[14]

One of us in 1991 suggested a very rough but extremely easily obtained size-independent measure for the complexity of a polycyclic molecular skeleton, the ratio κ of the number of cycles and the number of atoms.[15]

We in 1990, in the context of detecting symmetry and isomorphism in graphs, profited from the increase of walk counts with increasing edge weight and with increasing number of loops in a graph, corresponding to increasing bond multiplicity and the presence of heteroatoms in a molecule, respectively.[16] In 1993 we studied the structure dependence of walks in a graph or molecule and noticed that the total number of walks, the total walk count "twc can be looked upon as a measure of a tree's complexity, increasing both with increasing size and increasing branching."[17] We also stated that the differences in twc among isomeric alkanes completely reflect Bonchev and Trinajstić's intuitive rules

10.1021/ci9900579 CCC: $19.00
Published on Web 12/18/1999

of branching,[7a] though we gave few details at that time. We demonstrated, however, the high discrimination power (low degeneracy) of twc within acyclic as well as mono- and bicyclic alkanes.[17] In a recent paper we published the twc values of more than 500 acyclic and mono- through polycyclic saturated hydrocarbons and provided a file containing walk count data of these and many other saturated cyclic structures, so that the interested reader can perform extensive comparisons and may convince himself of the walk counts' merits.[18]

In 1986 Bertz and Herndon first proposed as a complexity measure the number of all connected subgraphs.[2,19] Bonchev in 1997 derived from the numbers of subgraphs quantities designated ^{e}TC ("*e*th-order topological complexities"), which are sums of the total adjacencies of all connected subgraphs of size *e* edges, and finally TC, the sum of the ^{e}TCs.[4d,e]

Bone and Villar in 1997 in an attempt to quantify the diversity of a compound library considered the number of all connected subsets of atoms in a molecular structure.[20]

Randić's most recent proposal for measuring molecular complexity is related to the use of *W* in that the basic graph invariants used are the distances.[5b] It is based on the molecular augmented valence recently defined by the same author, a vertex-additive quantity, where each vertex contributes the sum of the degrees of vertices at all distances, weighted by one of several possible weighting functions that decrease with increasing distance.[21]

Bertz and Zamfirescu recently considered as new complexity indices the numbers of particular subgraphs (cliques, bicliques) required to cover all the graph's edges.[22]

It is the purpose of the present paper to give a concise summary of the behavior of walk counts with respect to structural variations. It will thus be demonstrated that these quantities in a simple and quite natural manner account for all the basic complexity-enhancing features and result in reasonable ordering of structures.

DEFINITION OF ATOMIC AND MOLECULAR WALK COUNTS

A walk in a (molecular) graph is an alternating sequence of vertices and edges beginning and ending with a vertex, in which each edge is incident with the two vertices immediately preceding and following it. The number of edges along the walk is called its length. In a walk (in contrast to a path), repetition of individual vertices and edges is allowed. Therefore, there is no length limit for walks in a finite graph of *n* vertices.

It has been known for a long time that for a graph the number of walks of length *k* beginning in vertex *i* and ending in vertex *j* is given by the element $a_{ij}^{(k)}$ in the *k*th power of the adjacency matrix, $\mathbf{A}^k$.[23] The number of all walks of length *k* starting at atom *i* (and ending in any atom, including *i* itself), the atomic walk count of length *k* of atom *i*, $\mathrm{awc}_k(i)$, is therefore obtained by summing the elements in the *i*th row (or column) of matrix $\mathbf{A}^k$.[17]

Summation of the $\mathrm{awc}_k(i)$ over the lengths $k = 1$ to $n - 1$ gives what we call the atomic walk count sum for atom *i*, awcs(*i*). Finally, summing the awcs(*i*) of all atoms followed by division by 2 gives the total walk count of a (molecular) graph, twc. Thus, twc is half the sum of all elements in the adjacency matrix powers from exponent 1 to $n - 1$, where *n* is the number of the graph's vertices.[24]

$$\mathrm{twc} = {}^1/_2 \sum_{i=1}^{n} \mathrm{awcs}(i) = {}^1/_2 \sum_{k=1}^{n-1} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{(k)}$$

Alternatively, summation of the $\mathrm{awc}_k(i)$ over all atoms gives the molecular walk count of length *k*, mwc_k. Each mwc_k thus is the sum of the entries in $\mathbf{A}^k$. Summation of all mwc_k for length $k = 1$ to $n - 1$ followed by division by 2 again gives the total walk count, twc.[17]

Thus, the quantities $\mathrm{awc}_k(i)$ and awcs(*i*) characterizing the structural situation of vertex (atom) *i* in the (molecular) graph, and the quantities mwc_k and twc characterizing the graph as a whole, are conveniently obtained once the walk counts $a_{ij}^{(k)}$ are known.

It was therefore important to find that the atomic walk counts $\mathrm{awc}_k(i)$ can be obtained in an even easier way than by matrix multiplication. In fact $\mathrm{awc}_k(i)$ is simply the sum of the awc_{k-1} of all neighbor atoms of atom *i*. Since $\mathrm{awc}_1(i)$ is the degree of atom *i* (the number of its neighbors), $\mathrm{awc}_2(i)$ is the sum of the degrees of all neighbors of atom *i*. Values of $\mathrm{awc}_k(i)$ are thus obtained extremely easily by repeated summation over the contributions of the neighbor atoms of *i*, starting with the degrees, and this procedure is the long-known Morgan algorithm. That is, $\mathrm{awc}_k(i)$ is the extended connectivity $\mathrm{EC}_k(i)$. All the above was described and illustrated in our earlier publication.[17] The equivalence of walk counting and the Morgan algorithm was independently found by Figueras.[25]

The connection between the fundamental graph-combinatorial concept walk counts and the fundamental graph-spectral quantity λ_1 was shown by Cvetković and Gutman in 1977.[11] In our notation

$$\lim_{k\to\infty} (\mathrm{mwc}_k/n)^{1/k} = \lambda_1$$

That is, the principal eigenvalue can be understood as an averaged extended degree of the graph's vertices (averaged over all vertices and extended to infinitely long walks).

We later showed a corresponding connection to exist between atomic walk counts within a graph and the coefficients in the eigenvector corresponding to λ_1.[26]

DEPENDENCE OF WALK COUNTS ON STRUCTURAL FEATURES

Dependence on Size. The walk counts mwc_k increase with increasing graph size *n*, that is, with increasing number of vertices, since each newly introduced vertex is connected to at least one vertex and thus contributes at least two new entries "1" in the adjacency matrix. The total walk count twc increases even more rapidly than a mwc_k, since for an increase in *n* by 1 another higher power of **A** is included in the summation. In Table 1 mwc_k and twc values are given for the graphs corresponding to the straight-chain alkanes of $n = 1-10$ and for the parent monocycloalkane graphs of $n = 3-10$.

For the unbranched alkane of *n* carbon atoms (the chain graph of *n* vertices) and for $k \leq n - 1$, the following recursion formulas can be derived:

Table 1. Values of mwc_k and twc for Unbranched Alkanes and Cycloalkanes

name	mwc_1	mwc_2	mwc_3	mwc_4	mwc_5	mwc_6	mwc_7	mwc_8	twc
methane	0	0	0	0	0	0	0	0	0
ethane	2	2	2	2	2	2	2	2	1
propane	4	6	8	12	16	24	32	48	5
butane	6	10	16	26	42	68	110	178	16
pentane	8	14	24	42	72	126	216	378	44
hexane	10	18	32	58	104	188	338	610	111
heptane	12	22	40	74	136	252	464	860	268
octane	14	26	48	90	168	316	592	1114	627
nonane	16	30	56	106	200	380	720	1370	1439
decane	18	34	64	122	232	444	848	1626	3250
cyclopropane	6	12	24	48	96	192	384	768	9
cyclobutane	8	16	32	64	128	256	512	1024	28
cyclopentane	10	20	40	80	160	320	640	1280	75
cyclohexane	12	24	48	96	192	384	768	1536	186
cycloheptane	14	28	56	112	224	448	896	1792	441
cyclooctane	16	32	64	128	256	512	1024	2048	1016
cyclononane	18	36	72	144	288	576	1152	2304	2295
cyclodecane	20	40	80	160	320	640	1280	2560	5110

$$mwc_1 = 2n - 2$$

$$mwc_{k+1} = \begin{cases} 2mwc_k - 2\binom{k}{k/2} & \text{if } k \text{ is even} \\ 2mwc_k - 2\binom{k}{(k+1)/2} & \text{if } k \text{ is odd} \end{cases}$$

For the parent monocycloalkane of n carbon atoms (cycle of n vertices), the following can be shown:

$$mwc_k = n \cdot 2^k$$

$$twc = n \cdot (2^{n-1} - 1)$$

For regular graphs of degree d these equations can be generalized:

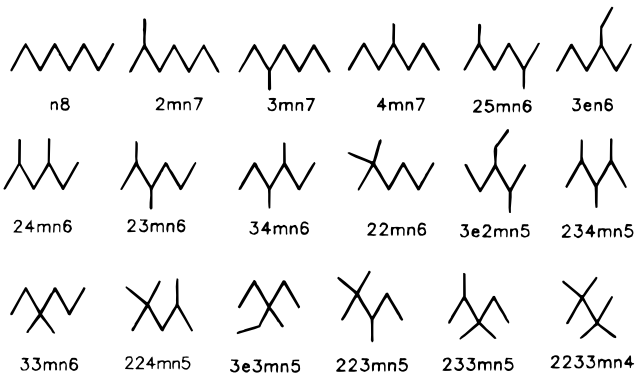
$$mwc_k = n \cdot d^k$$

$$twc = n \cdot d(d^{n-1} - 1)/2(d - 1)$$

The rapid (exponential) increase in twc for increasing n is responsible for twc being highly discriminating despite being an integer number graph invariant.[17,18,26]

Dependence on Branching. Walk counts mwc_k and twc increase with increasing branching in acyclic graphs. For constant n the lowest mwc_k and twc are always associated with the chain graph (n-alkane), the highest with the star graph ($n - 1$ peripheral vertices connected to a central vertex, not possible for alkanes except for 2-methylpropane and 2,2-dimethylpropane).[26] Chart 1 shows for the 18 acyclic octane graphs the ordering given by twc along with the values of a few graph invariants that were used earlier to order these isomers by complexity. As seen from Chart 1, index twc is sensitive not only to the number of branches but also to more subtle structural details, increasing with increasing centrality of branch attachment ("branch centrality"[4d]), with increasing length of a branch, and with decreasing distance between branching points ("branch clustering" or "multiple vertex branching"[4d]). Another way to express the same facts is to state that adding a new vertex to a given graph results in a higher twc for attachment at a vertex of higher degree or at a vertex closer to a higher degree vertex. To see this, compare those octanes that can be derived from a common precursor heptane by attaching a methyl group in various positions, e.g., n8, 2mn7, 3mn7, 4mn7 (from n7); or 2mn7, 3mn7, 25mn6, 24mn6, 23mn6, 22mn6 (from 2mn6); or 3mn7, 4mn7, 3en6, 24mn6, 23mn6, 34mn6, 33mn6 (from 3mn6).

Chart 1. The 18 Acyclic Octane Graphs and Some of Their TI Values



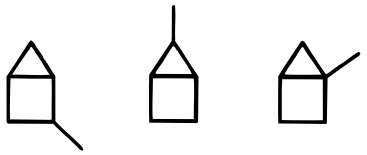
name	twc	W	TC[a]	Bertz' ranking[b]	λ_1
n8	627	84	224	1	1.87939
2mn7	764	79	279	2	1.94986
3mn7	838	76	312	3	1.98904
4mn7	856	75	323	4	2.00000
25mn6	911	74	348	6	2.00000
3en6	928	72	356	5	2.02852
24mn6	997	71	393	7	2.04208
23mn6	1068	70	414	8	2.07431
34mn6	1136	68	448	9	2.09529
22mn6	1142	71	411	12	2.11199
3e2mn5	1152	67	459	10	2.10100
234mn5	1296	65	532	11	2.13578
33mn6	1301	67	477	13	2.15664
224mn5	1317	66	519	15	2.14896
3e3mn5	1441	64	532	14	2.18890
223mn5	1536	63	597	16	2.20595
233mn5	1609	62	618	17	2.22158
2233mn4	2047	58	798	18	2.30278

[a] Reference 4d. [b] Reference 9.

The same features are described also by W,[7a] by Bonchev's TC,[4d,e] by λ_1, or by Bertz' sequences of line counts in the iterated line graphs.[9] Not surprisingly, twc, W, TC, λ_1, or Bertz' measure result in a very similar ordering of the octane isomers. Thus all these quantities seem to measure roughly the same thing, branching, at least for this sample. Nevertheless, there are a few discrepancies, resulting from different weights being attributed to the competing complexity features by these measures. Thus 22mn6 is ascribed higher branching than 23mn6 or 34mn6 by twc, λ_1, and Bertz' measure, owing to a preponderance of multiple vertex branching over branch centrality, while W and TC set reverse priorities and result in a reverse order for these particular isomers.

Dependence on Cyclicity. Walk counts mwc_k and twc increase with increasing number of cycles in a graph. This is again trivial, since each ring closure corresponds to two new entries "1" in **A** instead of two "0". Comparison of the two series of compounds in Table 1 shows that the increase in twc caused by a ring closure (the ratio twc(ring):twc(chain)) is more pronounced for smaller than for larger n.

Chart 2. Methylbicyclo[3.2.1]pentanes and Some of Their TI Values



1 2 3

#	twc	W	Randic AV[a]	Randic complexity[a]	λ_1	wcx
1	475	25	39.625	39.625	2.53948	475.0
2	491	25	39.750	25.875	2.56155	314.5
3	527	24	40.625	40.625	2.59907	527.0

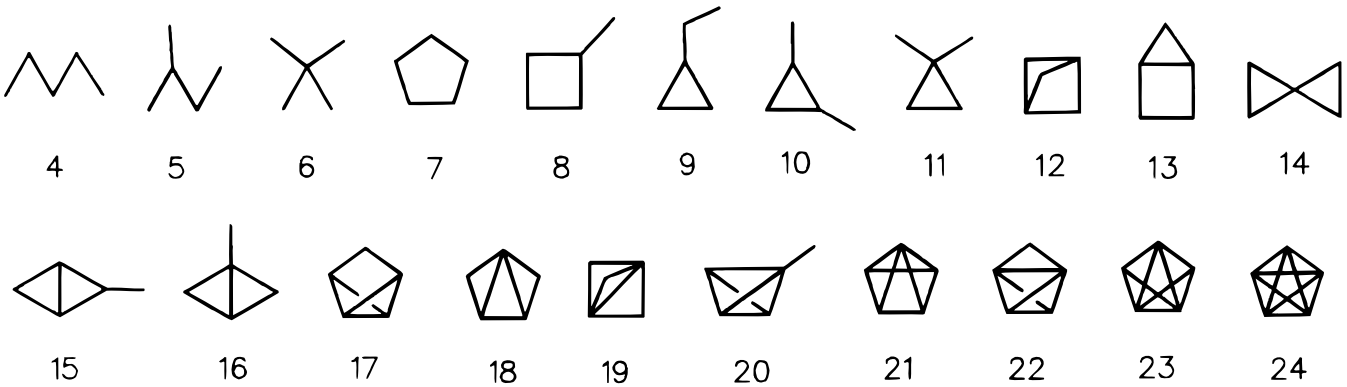
[a] Reference 5b.

This corresponds to the chemical experience that a small cycloalkane behaves distinctly different from the corresponding *n*-alkane, while the behavior of a large cycloalkane is similar to that of a long *n*-alkane.

For cyclic graphs of constant *n*, the lowest mwc_k and twc are associated with the *n*-membered monocycle, the highest with the complete graph K_n. Otherwise twc and the mwc_k increase according to the same principles as in acyclic graphs. Thus if a substituent can be attached to different positions of a cyclic skeleton, twc (and mwc_k) is higher for attachment at a point of higher degree, or for attachment at a point closer to a vertex of higher degree. This is shown in Chart 2 for the methylbicyclo[2.1.0]pentanes **1**−**3**, where additionally the values for Randić's recent AV and complexity measures are given.[5b]

A set of more varied structures is given in Chart 3, the 21 simple connected graphs of $n = 5$ and vertex degree ≤ 4, corresponding to the acyclic and (poly)cyclic pentanes. The structures are given in twc order. In Chart 3 values of the same descriptors are given as in Chart 2, and additionally the values of path count p2, of Randić's cyclicity index γ,[10] of the rough complexity measure κ, and of Bonchev's TC,[4d,e] as far as the data are available. Note that there are a few discrepancies. Thus while twc places **7**, **8**, **9** in this logical

Chart 3. The 21 Acyclic−Polycyclic Pentane Graphs and Some of Their TI Values



4 5 6 7 8 9 10 11 12 13 14

15 16 17 18 19 20 21 22 23 24

#	twc	W	p2	γ[a]	κ	TC[b]	Randic AV[c]	Randic Complexity[c]	λ_1	wcx
4	44	20	3		0	60	18.375	11.438	1.73205	28.0
5	53	18	4		0	76	19.500	16.125	1.84776	44.0
6	70	16	6		0	100	21.000	9.750	2.00000	32.5
7	75	15	5	0	0.2	160	25.000	5.000	2.00000	15.0
8	89	16	6		0.2	190	25.125	19.875	2.13578	69.5
9	93	17	6		0.2	172	24.750	19.625	2.21432	71.5
10	107	16	7		0.2	212	27.375	15.625	2.30278	66.0
11	116	15	8		0.2	230	26.500	16.750	2.34292	74.5
12	147	14	9		0.4	504	31.500	12.750	2.44949	59.5
13	150	14	9	0.32	0.4	482	31.500	18.750	2.48119	89.5
14	166	14	10		0.4	492	32.000	14.000	2.56155	76.0
15	175	15	10		0.4	522	31.625	24.375	2.64119	131.0
16	188	14	11		0.4	566	32.500	26.250	2.68554	151.5
17	247	13	13	0.56	0.6	1278	38.000	22.250	2.85577	141.5
18	269	13	14	0.56	0.6	1316	38.500	23.750	2.93543	168.5
19	289	13	15		0.6	1394	39.000	16.000	3.00000	120.0
20	304	13	15		0.6	1396	39.000	22.500	3.08613	167.0
21	390	12	18	0.76	0.8	3216	45.000	18.750	3.23607	166.5
22	421	12	19	0.72	0.8	3290	45.500	26.500	3.32340	240.0
23	600	11	24	0.88	1.0	7806	52.500	31.750	3.64575	235.5
24	850	10	30	1.00	1.2	18180	60.000	12.000	4.00000	170.0

[a] Reference 10. [b] Reference 4d,e. [c] Reference 5b.

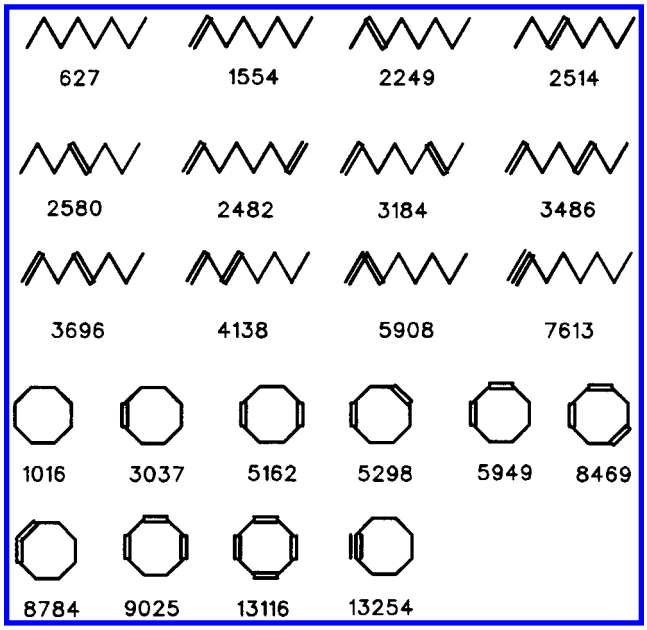


Figure 1. Unbranched chain and eight-membered ring $n = 8$ hydrocarbons and their twc values.

order, TC reverses **8** and **9**, AV rates **9** lower than **7** and **8**, and *W* reverses the order completely. The given relative ordering of **10** and **11** is reasonable by the principles discussed above, yet this order is reversed by AV. We note in passing that the cyclicity index γ is unfortunately 0 by definition for monocyclics such as **7** and that otherwise, curiously enough, there is some parallelism between γ and κ.

Dependence on Edge Weights (Unsaturation). Walk counts mwc_k and twc increase with increasing unsaturation, that is, with increasing edge weights. In the simplest case, introduction of a double bond instead of a single bond between atoms *i* and *j* results in adjacency matrix entries $a_{ij} = a_{ji} = 2$ instead of $a_{ij} = a_{ji} = 1$, so trivially the sums of entries in the powers of **A** will increase. In the alternative picture of walks, a double bond gives rise to new walks, since it may be traversed via one or the other of its constituent lines.

Figure 1 shows straight-chain and eight-membered ring $n = 8$ hydrocarbons of various degree of unsaturation together with their twc values. As with branching, the effect of unsaturation is the more pronounced the more central the position of the multiple bond. The order found for dienes is in accord with chemical experience, in that a second double bond has a similar effect as the first one if the two are far apart and that on the other hand two conjugated or cumulated double bonds constitute a new distinct feature.

Dependence on Vertex Weights (Heteroatoms). Walk counts mwc_k and twc increase with increasing vertex weights, that is, with increasing presence of heteroatoms. A heteroatom in position *i* is represented in the adjacency matrix by an entry $a_{ii} > 0$ and thus will result in an increased sum of matrix entries. In the alternative picture of walks, $a_{ii} = 1$ means one loop on atom *i*, which gives rise to additional walks such as the walk of length 1 from *i* to *i*.

Figure 2 shows the influence of heteroatoms on twc. Arbitrarily here oxygen atoms were used and represented by $a_{ii} = 1$. A heteroatom influences twc more the more central its position within the graph. A second heteroatom

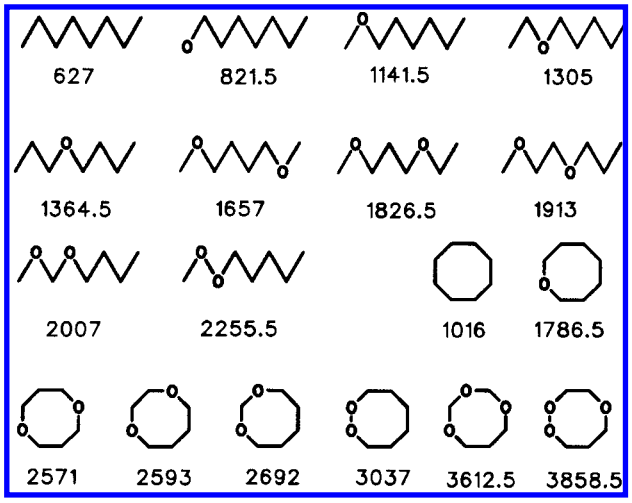
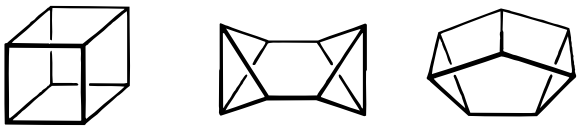


Figure 2. Some octanes containing heteroatoms and their twc values.

Chart 4. The Pentacyclic Octane Graphs and Some of Their TI Values



cubane octabisvalene cuneane

name	twc	W	p2	Randic AV	Randic complexity	wcx
cubane	13116	48	24	81.0	10.125	1639.5
octabisvalene	13116	50	24	79.5	19.875	3279.0
cuneane	13116	46	24	82.5	31.125	4918.5

increases twc more if it is closer to the first heteroatom. This is in accord with chemical experience that, e.g., two oxygen atoms several bonds apart give rise merely to a bis-ether, while two closer oxygens result in a distinct functional group such as an acetal or a peroxide.

Accounting for Symmetry. Randić recently stressed that the intuitive notion of complexity does take into account the symmetry that may be present and that therefore a quantity claimed to measure complexity should do so, too.[5b] We therefore here propose the suggestive name "labyrinthicity" for what is measured by twc (or the mwc_k), that is, how many possibilities are given to walk or oscillate through the graph along its edges and loops. Complexity then is what is measured by a symmetry-modified twc. While such a modification can certainly be achieved by overlaying twc with a Shannon-type contribution, we here favor another very simple possibility. Index twc is a vertex-additive quantity, obtained by summing the awcs(*i*), and the modification consists of summing the contribution of one vertex only from each class of vertices equivalent by symmetry. We denote the resulting quantity by wcx, walk complexity. Walk complexity values are included in the last column in Charts 2 and 3. In Chart 4 wcx values are used to order the (regular) graphs of cubane, octabisvalene, and cuneane. Note that perceiving the symmetry correctly is not a trivial task.[5b,16,27]

ORDERING OF STRUCTURES

Obviously, in cases of competing complexity features, the question arises how to weight the influences of, e.g., skeletal

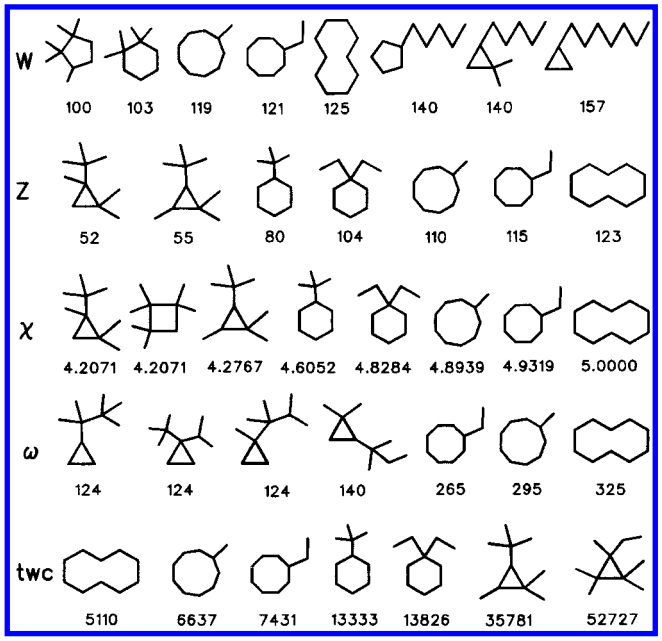


Figure 3. Ordering of monocyclic decanes by various TIs.

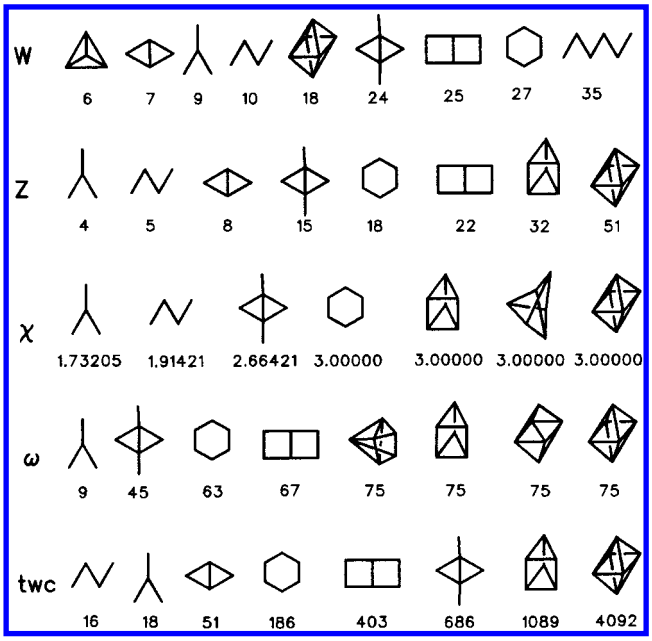


Figure 4. Ordering of acyclic−polycyclic butanes through hexanes by various TIs.

branching (or cyclicity), multiple bonds, and heteroatoms. The concept of walks has enough built-in flexibility to deal with this problem. Thus if one is not pleased with cyclooctene being rated more complex than oxacyclooctane, the edge weight of a double bond may be diminished (e.g., to 1.5 instead of 2) or the vertex weight of oxygen may be increased (e.g., to 2 instead of 1). Use of noninteger numbers as adjacency matrix entries of course renders the picture of walk counting more difficult to apply; it is, however, of no influence on the computational formalism.

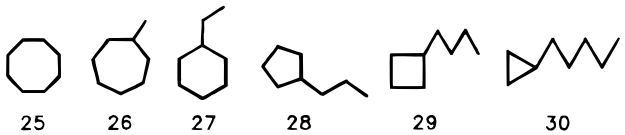
The use of walks, even for simple graphs (without multiple edges or vertex weights), may be modified. If the reader is not pleased with the exponential increase of twc with increasing n, the walks may be weighted using a function that drastically decreases with increasing k. For example, using the kth root of mwc_k instead of mwc_k would lead to a quantity similar to λ_1 instead of twc.[11] Alternatively, instead of twc an mwc_k may be used for comparison of structures with k appropriately chosen, and this latter possibility allows for a fairer comparison of structures of different size.

To compare the ordering of structures induced by twc and other TIs, we reconsidered two samples of structures, the monocyclic decanes (475 structures, sample 1) and the acyclic through polycyclic saturated hydrocarbon structures of $n = 4-6$ (105 structures, sample 2).[18] Unsaturated and heteroatom-containing structures were deliberately not considered here in order to avoid the element of arbitrariness in the numerical expression of theses structural features. In Figures 3 and 4 the ordering of structures by several important TIs (including Hosoya's Z and the detour index ω) is shown for sample 1 and 2, respectively. In each row of both figures all graphs exhibiting extremal values of the respective TI and a few with intermediate values are included. While W clearly orders the graphs by compactness (from the most compact to the most extended in both samples), one gets the impression that in sample 1 (variable ring size, number, and branching of side chains) all the other TIs measure more or less the same property ("branchedness"). However, in sample 2 (additionally variable graph size and variable number of rings), it is clearly seen that twc is the only TI (of those included here) ordering the graphs by what one would intuitively feel as the order of increasing labyrinthicity, that is, complexity without accounting for symmetry. See in particular the order of *tert*-butylcyclohexane and 1,1-diethylcyclohexane in sample 1 and of *n*-butane and 2-methylpropane in sample 2. Note that W, Z, χ do not give a logical order for cyclodecane, methylcyclononane, and ethylcyclooctane (Figure 3).

In 1986 Razinger published a paper entitled "Discrimination and Ordering of Chemical Structures by the Number of Walks."[28] In this work he constructed two new TIs based on the number of walks, one on the number of all walks of length 1 through n (for a structure of n C atoms), I(NW) and the other on the number of walks of length n only, I(LW). Since Razinger's primary objective was to obtain highly discriminating indices, he applied Randić's procedure (summation over edge contributions, where each edge contributes the reciprocal square root of the product of atom indices for its two end atoms[6]) using as atom index the number of walks starting at that atom. The procedure was successful: index I(LW), obtained in that manner, is even less degenerate than twc.[29] Razinger then (despite the work's title) did not order chemical structures by the number of walks, but by the index I(LW). For the acyclic heptanes the same order results from I(LW) as from twc or λ_1, while for the acyclic decanes a few discrepancies arise.

In Chart 5 the monocyclic octanes containing one unbranched side chain appear in the order induced in concordance by twc or mwc7 or mwc8 (or, in fact, by any lower walk count, not shown) or by λ_1. This is the logical order in which a side chain stepwise grows out of a shrinking ring. In the last column the I(LW) values are given, and Razinger observed that this index orders these structures in a quite different sequence, which to us seems rather chaotic. In fact I(LW) here behaves somewhat similar to W in that it gives a minimum somewhere along the logical sequence, as described for W first in ref 8c.[30]

Chart 5. Ordering of Some Monocyclic Octanes by Various TIs



#	twc	mwc7	mwc8	λ_1	W	I(LW) ·1000
25	1016	1024	2048	2.00000	64	31.25
26	1279	1338	2822	2.09118	61	22.39
27	1401	1496	3202	2.13578	64	20.03
28	1449	1560	3392	2.16484	67	20.45
29	1470	1592	3492	2.19399	74	22.58
30	1493	1628	3602	2.23499	78	25.32

Recall that for (cyclo)octanes mwc7 is the highest mwc included in twc and that the number of walks of length 8 starting at vertex *i* is the basic vertex invariant from which *I*(LW) for (cyclo)octanes was constructed. Thus the price Razinger had to pay for raising the discriminating power of an index to a very high level was the loss of the logical order of these compounds.

COMPUTATIONAL EFFORT

The various graph invariants discussed in this work as measures of molecular branching, cyclicity, and complexity require widely different efforts to obtain their numerical values. Though for most of them computer programs are available, the computational complexity is of some concern, and of two measures—other things being equal—the less computationally involved should be preferred.

The connectivity index χ requires information only on connectedness and degrees of vertices, both extremely easily taken from the adjacency matrix.

The Wiener number *W* and Randić's AV are obtained from the distance matrix, which itself is easily constructed from the adjacency matrix.

Eigenvalue λ_1 requires diagonalization of the adjacency matrix.

Path numbers are obtained by tracing the paths individually, which (except for very short paths or for tree graphs) is a notoriously demanding (NP-complete) endeavor. This detracts from the usefulness of, e.g., γ and ω.

The number of spanning trees is obtained from the Laplacian matrix, as the determinant of a submatrix.[14]

Finding all connected subgraphs is, to the best of our knowledge, a task increasing exponentially with the size of the graph.[2,19] Complexity measures such as Bonchev's TC and TC1 suffer from this fact.[4d,e,20] We recently made available a computer program for this task.[31]

In contrast, walk counts are obtained very easily by repeated addition of adjacency matrix entries, i.e., the Morgan algorithm, as shown above. In fact it is hard to imagine a procedure computationally simpler than a sequence of addition steps.

REFERENCES AND NOTES

(1) (a) Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599−3601. (b) Bertz, S. H. A Mathematical Model of Molecular Complexity. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 206−221. (c) Bertz, S. H.; Sommer, T. J. Applications of Graph Theory to Synthesis Planning: Complexity, Reflexivity, and Vulnerability. In *Organic Synthesis: Theory and Application*; Hudlicky, T., Ed.; JAI Press: Greenwich, CT, 1993; Vol. 2, pp 67−92.

(2) Bertz, S. H.; Sommer, T. J. Rigorous Mathematical Approaches to Strategic Bonds and Synthetic Analysis Based on Conceptually Simple New Complexity Indices. *Chem. Commun.* **1997**, 2409−2410.

(3) (a) Whitlock, H. W. On the Structure of Total Synthesis of Complex Natural Products. *J. Org. Chem.* **1998**, *63*, 7982−7989. (b) Chanon, M.; Barone, R.; Baralotto, C.; Juillard, M.; Hendrickson, J. B. Information Theory Description of Synthetic Strategies in the Polyquinane Series. The Holosynthon Concept. *Synthesis* **1998**, 1559−1583.

(4) (a) Bonchev, D.; Polansky, O. E. On the Topological Complexity of Chemical Systems. In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; pp 126−158. (b) Bonchev, D. The Problems of Computing Molecular Complexity. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; pp 34−63. (c) Bonchev, D.; Seitz, W. A. The Concept of Complexity in Chemistry. In *Concepts in Chemistry—a Contemporary Challenge*; Rouvray, D. H., Ed.; Research Studies Press: Taunton, 1997; pp 353−381. (d) Bonchev, D. Novel Indices for the Topological Complexity of Molecules. *SAR QSAR Environm. Res.* **1997**, *7*, 23−43. (e) Bonchev, D. Overall Connectivity and Topological Complexity: A New Tool for QSPR/QSAR. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach: Langhorne, PA, 1999; pp 361−401.

(5) (a) Randić, M. On Molecular Branching. *Acta Chim. Slov.* **1997**, *44*, 57−77. (b) Randić, M. On the Concept of Molecular Complexity. *Croat. Chem. Acta*., in press.

(6) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

(7) (a) Bonchev, D. Trinajstić, N. Information Theory, Distance Matrix, and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533. (b) Bonchev, D. Topological Order in Molecules. 1. Molecular Branching Revisited. *J. Mol. Struct.*: *THEOCHEM* **1995**, *336*, 137−156.

(8) (a) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Topological Characterization of Cyclic Structures. *Int. J. Quantum Chem.* **1980**, *17*, 845−893. (b) Mekenyan, O.; Bonchev, D.; Trinajstić, N. Algebraic Characterization of Bridged Polycyclic Compounds. *Int. J. Quantum Chem.* **1981**, *19*, 929−955. (c) Mekenyan, O.; Bonchev, D.; Trinajstić, N. Structural Complexity and Molecular Properties of Cyclic Molecules with Acyclic Branches. *Croat. Chem. Acta* **1983**, *56*, 237−261. (d) Bonchev, D.; Balaban, A. T.; Liu, X.; Klein, D. J. Molecular Cyclicity and Centricity of Polycyclic Graphs. I. Cyclicity Based on Resistance Distances or Reciprocal Distances. *Int. J. Quantum Chem*. **1994**, *50*, 1−20.

(9) Bertz, S. H. Branching in Graphs and Molecules. *Discr. Appl. Math.* **1988**, *19*, 65−83.

(10) Randić, M. On Characterization of Cyclic Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1063−1071.

(11) Cvetković, D. M.; Gutman, I. Note on Branching. *Croat. Chem. Acta* **1977**, *49*, 115−121. See also: Lovasz, L.; Pelikan, J. On the Eigenvalues of Trees. *Period. Math. Hung.* **1973**, *3*, 175−182.

(12) Randić, M.; Brissey, G. M.; Spencer, R. B.; Wilkins, C. L. Search for All Self-Avoiding Paths for Molecular Graphs. *Comput. Chem*. **1979**, *3*, 5−13.

(13) Gutman, I.; Mallion, R. B.; Essam, J. W. Counting the Spanning Trees of a Labelled Molecular Graph. *Mol. Phys.* **1983**, *50*, 859−877.

(14) Nikolić, S.; Trinajstić, N.; Jurić, A.; Mihalić, Z.; Krilov, G. Complexity of Some Interesting (Chemical) Graphs. *Croat. Chem. Acta* **1996**, *69*, 883−897.

(15) Rücker, C. Habilitation Thesis (in German), Universität Freiburg, 1991. A few κ values are as follows: cyclohexane 0.167, cyclopentane 0.200, cyclobutane 0.250, cyclopropane 0.333, bicyclo[1.1.1]pentane 0.400, bicyclo[1.1.0]butane 0.500, perhydrofullerene[60] 0.517, dodecahedrane 0.550, quadricyclane 0.571, hexaprismane 0.583, [1.1.1]propellane 0.600, cubane, cuneane, and octabisvalene 0.625 each, prismane 0.667, tetrahedrane 0.750.

(16) (a) Rücker, G.; Rücker, C. Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relations among Atoms into Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 187−191. (b) Rücker, G.; Rücker, C. 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. (c) Rücker, G.; Rücker, C. Isocodal and Isospectral Points, Edges, and Pairs in Graphs and How to Cope with Them in Computerized Symmetry Recognition. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 422−427.

(17) Rücker, G.; Rücker, C. Counts of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683−695.

(18) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points, and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788−802.
(19) Bertz, S. H.; Herndon, W. C. Similarity of Graphs and Molecules. In *Artificial Intelligence Applications in Chemistry*; Pierce, T. H., Hohne, B. A., Eds.; American Chemical Society: Washington, DC, 1986; pp 169−175.
(20) Bone, R. G. A.; Villar, H. O. Exhaustive Enumeration of Molecular Substructures. *J. Comput. Chem.* **1997**, *18*, 86−107.
(21) Randić, M. Augmented Valence−Novel Graph Theoretical Index with an Apparent Interpretation. *Int. J. Quantum Chem*., in press.
(22) Bertz, S. H.; Zamfirescu, C. New Complexity Indices Based on Edge Covers. *MATCH−Commun. Math. Comput. Chem.*, in press.
(23) (a) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969; Chapter 13. (b) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, FL, 1992; p 48.
(24) The Cayley-Hamilton theorem states that the higher powers of **A**, starting from $\mathbf{A}^n$, are linear combinations of lower powers and as such do not contribute new information.
(25) Figueras, J. Morgan Revisited. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 717−718.
(26) Rücker, C.; Rücker, G. Mathematical Relation between Extended Connectivity and Eigenvector Coefficients. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 534−538.
(27) Klin, M.; Rücker, C.; Rücker, G.; Tinhofer, G. Algebraic Combinatorics in Mathematical Chemistry. Methods and Algorithms. I. Permutation Groups and Coherent (Cellular) Algebras. Technical Report, Technical University of Munich, 1995.
(28) Razinger, M. Discrimination and Ordering of Chemical Structures by the Number of Walks. *Theor. Chim. Acta* **1986**, *70*, 365−378.
(29) We tested the degeneracy of twc in the sample of all possible mono- and bicyclic octanes in order to compare it with that of I(LW). Within this sample of $N = 255$ structures twc has 236 distinct values, while Razinger reported 245 distinct values of I(LW) in a corresponding (but incomplete) sample of $N = 250$.
(30) The number of subgraphs and Bonchev's TC1 and TC similarly have a maximum somewhere along the same sequence.
(31) Rücker, G.; Rücker, C. Automatic Enumeration of All Connected Subgraphs. *MATCH−Commun. Math. Comput. Chem.*, in press.

CI9900579