277

Distance/Distance Matrices

Milan Randić,* Alexander F. Kleiner, and Luz M. DeAlba

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

We introduce novel matrices for graphs embedded on two- and three-dimensional grids. The matrices are defined in terms of geometrical and topological distances in such graphs. We report on some properties of these distance/distance matrices and have listed several structural invariants derived from distance/distance matrices. The normalized Perron root (the first eigenvalue) of such matrices, λ/n, for path graphs apparently is an index of molecular folding. The ratio $\phi = \lambda/n$ is 1 for (geometrically) linear structures, while it approaches 0 as the path graph is repeatedly folded.

Graphs have found considerable use in chemistry,[1] particularly in modeling molecular structure.[2] In applications of graphs to the study of structure–property relationships, molecules are represented by selected molecular descriptors, often referred to as topological indices.[3] These topological indices, which often have a direct structural interpretation, are defined in terms of selected structural parts and hopefully should help one in building molecular models for structure–property relationships. Among hundreds of possible descriptors a few have emerged again and again as the most useful for characterization of molecules.[4] These include the following: the Hosoya Z topological index, the first deliberate single number representation of molecules. It is a global molecular quantity, defined in terms of the count of disjoint edges in a graph.[5] The Wiener number W, was introduced already in the late 1940s[6] as one of two descriptors for regressions of physicochemical properties (such as the boiling points) of molecules. It illustrates a bond additive quantity. For each bond in a molecule one multiplies the number of vertices (atoms) on each side of an edge (bond), and adds all such contributions. Finally, the molecular connectivity index,[7] χ, also a bond additive quantity, discriminates bonds according to their type. Bonds having different numbers of nearest neighbors for its terminal atoms make different contributions. These, and several other indices, have been found very useful in performing regression analysis of several molecular properties and for different families of molecules.

Over the years criticisms have been heard that graphs are deficient as molecular models. They apparently fail to represent heteroatoms and the three-dimensional structure of molecules. These criticisms usually came from adherents of the traditional approaches to structure–activity studies, who use one set of molecular properties to "explain" other, more complex, properties. Such criticisms are, however, only in part justified. Representations of heteroatoms and representation of three-dimensional structures by graphs embedded in space are currently only in their infancy.

It has been recognized, particularly by Kier and Hall,[8] that heteroatoms require *additional* molecular descriptors. Kier and Hall achieved this by introducing their valence δ parameters into the connectivity indices. Recently a systematic search for optimal descriptors for heteroatoms was started.[9] It will take some time before we find optimal descriptors for all heteroatoms of interest. It may take even more time before one finds possible regularities among such descriptors that may relate to the electronic structure of heteroatoms involved.

By embedding graphs on regular networks (such as a graphite network in two-dimensional space and a diamond grid for three-dimensional space), one gets matrices which are geometry sensitive.[10] Structural invariants derived from such geometry-sensitive matrices will be geometry dependent. Hence, in this way one can discriminate between *cis* and *trans* isomers, or between *anti* and *gauche* conformers.

Both these important developments, which make modeling of molecules by graphs more realistic, are in their infancy. One of the reasons is that the number of researchers interested in this topic has been small. However, the subject has been continually receiving more attention, and we hope that in the near future we will witness a rapid development of topological indices for heteroatoms and for three-dimensional structures.

In this paper we will describe a novel approach to structures embedded in two-dimensional space that goes beyond the use of simple geometry matrices. The approach can be extended to structures embedded in three-dimensional space.

DISTANCE MATRICES

Any matrix, the elements of which satisfy the axioms of distance, can be referred to as a distance matrix. The axioms of distance require that (1) distance is a positive quantity, $D_{i,j}$, assigned to a pair of elements (points in an n-dimensional space), with the provision that $D_{i,i}$ is zero; (2) distance does not depend on the direction of measurement, i.e., $D_{i,j} = D_{j,i}$; and (3) distance satisfies the triangular inequality, i.e., the direct distance between i and j is the shortest distance, $D_{i,j} < D_{i,k} + D_{k,j}$. Besides the familiar Euclidean distance of everyday three-dimensional geometry, and its generalization to metric tensors of "curved" spaces in physics, we may mention a generalization of the distance measure due to Minkowski.[11] The special case of Minkowski distance is the so-called city-block distance (based on absolute values of the differences in the coordinates, rather than using various powers of such differences). In addition there are the Hausdorf distance,[12] Hamming distance used in coding theory[13], Mahalanobis distance of cluster analysis,[14] and so on. It is interesting to see that Kirchhoff electricity resistance laws also satisfy distance axioms. This observation has led recently to suggest the resistance distance matrices for cyclic graphs. Here the edges of a graph are viewed as resistors. Such resistance-distance matrices represent a "natural" generalization of topological distance matrix (*vide infra*) to cyclic graphs.[15]

We will briefly mention only two distance matrices, one based on the topological distance for vertices within a graph, and the other based on geometric distances of vertices in a graph.
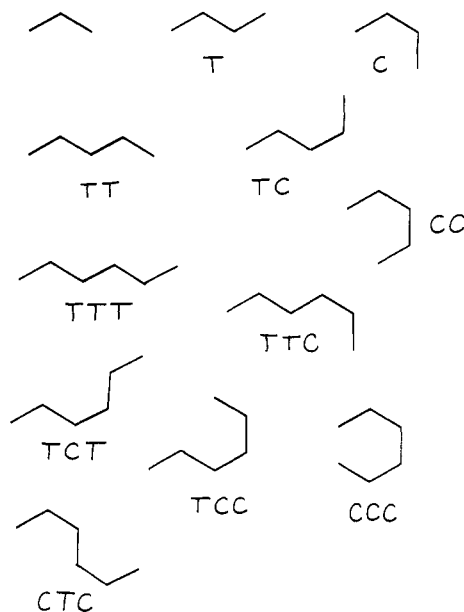

Figure 1. Possible conformations of smaller chains superimposed on a graphite lattice.

Topological Distance Matrix. The matrix elements in the (topological) distance matrix of a graph represent the distances between vertices in a graph measured by the number of edges between the vertices. In trees vertices are separated by a unique path, the length of which gives the sought distance. In cyclic graphs one has to select the shortest (of several or many) possible paths. The distance matrix of graphs has received some attention in the mathematical literature.[16] It also served as a source for construction of several topological indices. Hosoya[5] observed that the Wiener index W can be obtained from the distance matrix of a graph (tree) by summing all entries above the main diagonal. Balaban[17] considered row sums in a distance matrix from which he constructed his J index by applying the reciprocal root algorithm (in analogy with the connectivity index) to pairs of adjacent vertices. The J index was one of early topological indices showing considerable discriminatory power. The smallest trees having the same J values were found among graphs with $n = 12$ vertices.[18]

The graph distance matrix for a tree contains the same information that is in the adjacency matrix. Its use in applications in chemistry is, perhaps, limited because the entries in such a matrix for more distance vertices increase while interactions of distant atoms decrease. One can therefore modify the distance matrix, by considering reciprocal distance $1/D$ (as in potential), reciprocal $1/D^2$ (as in force), and other similar constructions, like $1/D^6 - 1/D^{12}$ (as in the Lennard-Jones long-range potential). Thus modified matrices may better simulate interactions between atoms at greater distances and may produce more suitable mathematical descriptors. Recently the reciprocal distance matrix has been tested in several applications.[19] We will consider one such modification, but of quite a different type, in the next section.

Geometry Matrices. Geometry matrices are based on the actual molecular geometry of a structure in three-dimensional space, or on geometry-based standardized bond lengths and bond angles. The latter can be derived by embedding a graph on a regular two-dimensional or three-dimensional grid. In the latter case we often speak of topographic matrices. In Figure 1 we illustrate embedding of smaller path graphs on a honeycomb (graphite) lattice, and in Table 1 we show the topographic matrix for the *all-trans* isomer of decane. The entries in such matrices depend on the embedding, and different

Table 1. Topographic Matrix for *all-trans* Chain of Ten Carbon Atoms Superimposed on a Graphite Lattice

0	1	√3	√7	2√3	√19	3√3	√37	4√3	√61
1	0	1	√3	√7	2√3	√19	3√3	√37	4√3
√3	1	0	1	√3	√7	2√3	√19	3√3	√37
√7	√3	1	0	1	√3	√7	2√3	√19	3√3
2√3	√7	√3	1	0	1	√3	√7	2√3	√19
√19	2√3	√7	√3	1	0	1	√3	√7	2√3
3√3	√19	2√3	√7	√3	1	0	1	√3	√7
√37	3√3	√19	2√3	√7	√3	1	0	1	√3
4√3	√37	3√3	√19	2√3	√7	√3	1	0	1
√61	4√3	√37	3√3	√19	2√3	√7	√3	1	0

embeddings of a same graph result in different matrices. Hence, structural invariants derived from such matrices will depend on the assumed conformation of a graph.[10] Geometry-dependent matrices appear to be useful tools for characterization of three-dimensional structures. There is, however, one disadvantage: Such matrices do not carry explicit information on bonding in a structure. For example, the topographic matrices for a six-membered ring and *all-cis* path P_6 (its spanning tree)

are identical. The reason is that the distances between corresponding pair of vertices in both structures are the same, including also the terminal vertices of path P_6:

$$\begin{pmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ 1 & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} \\ \sqrt{3} & 1 & 0 & 1 & \sqrt{3} & 2 \\ 2 & \sqrt{3} & 1 & 0 & 1 & \sqrt{3} \\ \sqrt{3} & 2 & \sqrt{3} & 1 & 0 & 1 \\ 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 0 \end{pmatrix}$$

In most applications to molecules one can deduce the bonding pattern by focusing attention to the shortest distances, but in some molecules and clusters the bonding pattern need not be clear. Observe that topographic matrices, as graph distance matrices, have considerable redundant information. For complete reconstruction of a molecule it suffices to know the precise geometrical disposition of four successive atoms, that is, the relative position of three successive bonds. Each four successive atoms defines a tetrahedron in space (or trapezoid in a plane) and thus fully determine the local geometry of a structure, and consequently the global geometry.

Topographic matrices have the same disadvantage already mentioned for the graph distance matrix: The atoms at greater separation are associated with large entries in a matrix. Our motivation for a novel distance matrix in part originate with this clear undesirable aspect of the distance matrices—increase of the matrix elements with interatomic separation.

DISTANCE/DISTANCE MATRICES

To overcome such limitations, we consider a novel distance-based matrix for graphs. Our approach applies to embedded arbitrary graphs (trees, cyclic, and polycyclic structures). However, we will in this introductory paper restrict attention primarily to path graphs. In particular, we will consider embedding of path graphs on the honeycomb lattice (regular hexagonal graphite grid).

Consider *trans-* and *cis*-butadiene, which have the same graph distance matrix but different topographic matrices:

butadiene graph   *trans*-butadiene   *cis*-butadiene
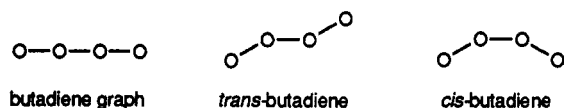
Can we combine the information on topographic distances and the information on graph distances in a single matrix representation?

One way to achieve this is to take the Hadamard (entrywise) product[20] of two matrices as follows: construct the matrix $\mathbf{D}'$ whose diagonal entries are 1 and whose off-diagonal entries are the reciprocals of the entries in the graph distance matrix. Let $\mathbf{DD} = \mathbf{D}'\cdot\mathbf{T}$, Hadamard product of $\mathbf{D}'$ with $\mathbf{T}$. Since the geometric distance equals or is smaller than the graph distance, then the entries in $\mathbf{DD}$ are at most 1. This simple procedure thus eliminates one of the undesirable features of the distance matrix, the size growth. For *trans*- and *cis*-butadiene we then obtain novel distance matrices:

trans-butadiene
$$\begin{pmatrix} 0 & 1 & \sqrt{3}/2 & \sqrt{7}/3 \\ 1 & 0 & 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 & 0 & 1 \\ \sqrt{7}/3 & \sqrt{3}/2 & 1 & 0 \end{pmatrix}$$

cis-butadiene
$$\begin{pmatrix} 0 & 1 & \sqrt{3}/2 & 2/3 \\ 1 & 0 & 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 & 0 & 1 \\ 2/3 & \sqrt{3}/2 & 1 & 0 \end{pmatrix}$$

In Table 2 we give the novel matrices for graphs of Figure 1. We will refer to these matrices as distance/distance matrices, or $\mathbf{DD}$ matrices, as a reminder that in their construction information from two different distance matrices is fused.

Reconstruction. The question can immediately be raised: Is there a loss of information that followed contraction of two distinct matrices into a single matrix? The answer would be no if from a given $\mathbf{DD}$ matrix we could always reconstruct the original topographic and graph distance matrix. We will show that indeed this is possible for topographic matrices, but it is not true for general geometry matrices.

Let us illustrate the above claims on two examples. Consider first the matrix of *trans*-butadiene (shown already). If we set all entries in the $\mathbf{DD}$ matrix zero except those that have value 1 (when the graph distance equals the topographic distance) we get the adjacency matrix $\mathbf{A}$ for the graph. From an adjacency matrix a graph can be reconstructed (because of a one-to-one correspondence) and hence the graph distance matrix $\mathbf{D}$ can be derived. By having the $\mathbf{D}$ matrix and $\mathbf{DD}$ matrix, it is simple to obtain the topographic matrix $\mathbf{T}$ (by taking the Hadamard product of the two matrices). Having matrices $\mathbf{A}$, $\mathbf{D}$, $\mathbf{T}$, and $\mathbf{DD}$, embedding of the graph on the underlying lattice grid is straightforward.

The above reconstruction was possible because we restricted the two-dimensional structure by requiring its embedding on a grid (here hexagonal regular lattice). The same will be true for graphs embedded on three-dimensional grids. Thus all entries in $\mathbf{DD}$ matrices, except for adjacent vertices, are less than 1, and the adjacency can be simply discerned. In a general case, however, this will not be the case. Consider graph K_3 and path graph P_3 (of length 2):
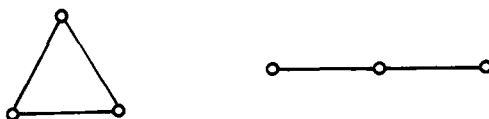
The $\mathbf{DD}$ matrices for both graphs are the same:
$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Hence neither graphs K_3 and P_3 nor their $\mathbf{D}$ and $\mathbf{T}$ matrices can be reconstructed. Similarly, the complete graph K_4 and linearly embedded path P_4 have the same $\mathbf{DD}$ matrix:
$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

These considerations extend to larger complete graphs K_n and larger strictly linear embedded paths P_n. These are the only cases of such degeneracy for $\mathbf{DD}$ matrices, because the strictly linear paths and the complete graphs are the only graphs for which topological distance equals geometrical distance (in n-dimensional space).

INVARIANTS OF DD MATRICES

$\mathbf{DD}$ matrices are source of novel structural invariants that can serve as molecular descriptors in structure–property-activity studies. We will consider here a few invariants of $\mathbf{DD}$ matrices.

Eigenvalues. Eigenvalues of $\mathbf{DD}$, and in particular the Perron root[21] (the first eigenvalue of the entrywise nonnegative matrix) of the $\mathbf{DD}$ matrix, can be expected, in analogy with the first eigenvalue of the adjacency matrix, to reflect some structural characteristics of embedded graphs. The first eigenvalue of adjacency matrix was suggested by Lovasz and Pelikan[22] as an index of molecular branching. Does the first eigenvalue of the $\mathbf{DD}$ matrix have simple structural interpretation? In Table 3 we list the eigenvalues and the coefficients of the characteristic polynomials for the six embeddings of path P_6 (path of length 5) illustrated at the lower part of Figure 1. We ordered graphs according to the value of the Perron root. The ordering suggests that the Perron root λ_1 parallels, at least qualitatively, the "degree of folding" of path P_6. The same appears to be the case also with P_7 (illustrated in Figure 2).

Path Numbers. In Table 4 we collected information on path numbers $p_1, p_2, p_3, \ldots$ in $\mathbf{DD}$ matrices for graphs of Figure 1. In Table 5 we give the same information for the nine conformers of P_7. The entries in each row first encode the conformation, which is followed by paths of length 1, paths of length 2, etc. The final entry, the path ID number, represents the summation of all path numbers. Conformations are labeled by recording the relative orientations of successive triple of bonds, encoded as *trans* or *cis*, T and C, respectively. On a regular hexagonal lattice adjacent vertices are separated by one bond, and the next nearest neighbors are at the distance $\sqrt{3}/2$. Hence, the leading members of the path sequences are same for isomers, and depend only on the size of the graph. The remaining members show differences for different embedding of a path graph.

Atomic ID Numbers. By considering row sums of the entries in $\mathbf{DD}$ matrices we can define the atomic ID numbers. In Table 6 we collected such information for all atomic environments for paths of length 6 (P_7). Some regularities in atomic ID numbers are apparent. As we move from the periphery toward the center of a molecule, the atomic ID numbers increase. This is not surprising. For *all-trans* isomers the $\mathbf{DD}$ matrices are Toeplitz matrices[23] (entries on each diagonal are same). Since off-diagonal entries in the first row correspond to a sequence of decreasing numbers, each

Table 2. Distance/Distance Matrices for Embedded Graphs of Figure 1

$$
\begin{pmatrix}
0 & 1.000 & 0.866 \\
1.000 & 0 & 1.000 \\
0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
T = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 \\
1.000 & 0 & 1.000 & 0.866 \\
0.866 & 1.000 & 0 & 1.000 \\
0.882 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
C = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.667 \\
1.000 & 0 & 1.000 & 0.866 \\
0.866 & 1.000 & 0 & 1.000 \\
0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TT = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.866 \\
1.000 & 0 & 1.000 & 0.866 & 0.882 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 \\
0.866 & 0.882 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.750 \\
1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 \\
0.750 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
CC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.667 & 0.433 \\
1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.667 & 0.866 & 1.000 & 0 & 1.000 \\
0.433 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TTT = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.866 & 0.872 \\
1.000 & 0 & 1.000 & 0.866 & 0.882 & 0.866 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.882 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.866 & 0.882 & 0.866 & 1.000 & 0 & 1.000 \\
0.972 & 0.866 & 0.882 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TTC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.866 & 0.721 \\
1.000 & 0 & 1.000 & 0.866 & 0.882 & 0.750 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.866 & 0.882 & 0.866 & 1.000 & 0 & 1.000 \\
0.721 & 0.750 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TCT = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.750 & 0.800 \\
1.000 & 0 & 1.000 & 0.866 & 0.667 & 0.750 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.882 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.750 & 0.667 & 0.866 & 1.000 & 0 & 1.000 \\
0.800 & 0.750 & 0.882 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
TCC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.882 & 0.750 & 0.529 \\
1.000 & 0 & 1.000 & 0.866 & 0.667 & 0.433 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.882 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.750 & 0.667 & 0.866 & 1.000 & 0 & 1.000 \\
0.529 & 0.433 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
CCC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.667 & 0.750 & 0.721 \\
1.000 & 0 & 1.000 & 0.866 & 0.882 & 0.750 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.667 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.750 & 0.882 & 0.866 & 1.000 & 0 & 1.000 \\
0.721 & 0.750 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

$$
CTC = \begin{pmatrix}
0 & 1.000 & 0.866 & 0.667 & 0.433 & 0.200 \\
1.000 & 0 & 1.000 & 0.866 & 0.667 & 0.433 \\
0.866 & 1.000 & 0 & 1.000 & 0.866 & 0.667 \\
0.667 & 0.866 & 1.000 & 0 & 1.000 & 0.866 \\
0.433 & 0.667 & 0.866 & 1.000 & 0 & 1.000 \\
0.200 & 0.433 & 0.667 & 0.866 & 1.000 & 0
\end{pmatrix}
$$

Table 3. Eigenvalues of Distance/Distance Matrices for All Possible Configurations of a Chain of Six Carbon Atoms and Coefficients of the Corresponding Characteristic Polynomials

	eignevalues					
	TTT	TTC	TCT	TCC	CCC.	CTC
λ_1	4.572	4.419	4.403	4.156	4.309	3.897
λ_2	-0.716	-0.503	-0.451	-0.115	-0.437	+0.227
λ_3	-0.806	-0.803	-0.851	-0.875	-0.676	-0.984
λ_4	-0.916	-0.954	-0.911	-1.026	-1.030	-1.046
λ_5	-1.013	-1.047	-1.085	-1.047	-1.062	-1.047
λ_{-1}	-1.122	-1.112	-.098	-1.092	-1.104	-1.047

power	characteristic polynomial coefficients					
x^6	1	1	1	1	1	1
x^5	0	0	0	0	0	0
x^4	-12.593	-11.833	-11.765	-10.697	-11.312	-9.748
x^3	-30.585	-27.415	-27.088	-27.525	-25.327	-18.273
x^2	-30.936	-26.315	-25.740	-18.449	-23.244	-11.304
x	-14.788	-11.715	-11.277	-6.112	-9.766	-0.781
1	-2.743	-1.983	-1.859	-0.492	-1.537	+0.998

successive row has one smaller entry in a summation replaced by a larger entry, thus increasing the atomic ID number.

The atomic ID numbers for interior vertices are not unique; the same atomic ID number occurs in more than one conformer. For example, atomic ID = 5.4857 occurs in TTTT for vertex 2 (and 6, due to the symmetry), and in TTTC it occurs for vertex 6. The two conformers differ in the local
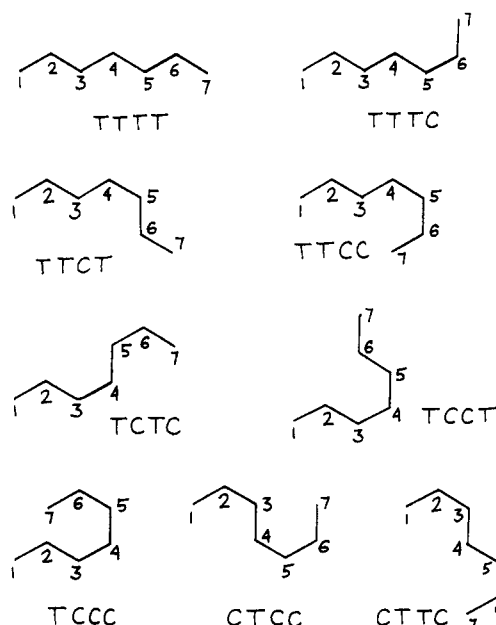

Figure 2. Possible conformations of chains having seven vertices (carbon atoms) superimposed on a graphite lattice.

environment of vertex 6. In TTTT the two fragments joined at vertex 6 are fused in the *trans* conformation while in TTTC

DISTANCE/DISTANCE MATRICES

J. Chem. Inf. Comput. Sci., Vol. 34, No. 2, 1994 **281**

Table 4. Path Numbers and ID Numbers for All Graphs of Figure 1

		T	C
p_1	2.000 00	3.000 00	3.000 00
p_2	0.866 03	1.732 05	1.732 05
p_3		0.881 02	0.666 67
ID no.	2.866 03	5.613 97	5.398 72

	TT	TC	CC
p_1	4.000 00	4.000 00	4.000 00
p_2	2.598 08	2.598 08	2.598 08
p_3	1.763 83	1.548 58	1.333 33
p_4	0.866 03	0.750 00	0.433 02
ID no.	9.227 94	8.896 66	8.364 42

	TTT	TTC	TCT
p_1	5.000 00	5.000 00	5.000 00
p_2	3.464 10	3.464 10	3.464 10
p_3	2.645 75	2.430 50	2.430 50
p_4	1.732 05	1.616 03	1.500 00
p_5	0.871 78	0.721 11	0.800 00
ID no.	13.713 68	13.231 74	13.194 60

	TCC	CTC	CCC
p_1	5.000 00	5.000 00	5.000 00
p_2	3.464 10	3.464 10	3.464 10
p_3	2.215 25	2.215 25	2.000 00
p_4	1.183 01	1.500 00	0.866 03
p_5	0.529 15	0.721 11	0.200 00
ID no.	12.391 52	12.900 46	11.520 13

Table 5. Path Numbers for All Graphs of Figure 2

	TTTT	TTTC	TTCT
p_1	6.0000	6.0000	6.0000
p_2	4.3301	4.3301	4.3301
p_3	3.5277	3.3124	3.3124
p_4	2.5981	2.4821	2.3660
p_5	1.7436	1.5929	1.5211
p_6	0.8660	0.7638	0.7676

	TTCC	TCTC	TCCT
p_1	6.0000	6.0000	6.0000
p_2	4.3301	4.3301	4.3301
p_3	3.0972	3.0972	3.0972
p_4	2.0490	2.2500	1.9320
p_5	1.2502	1.5211	1.0583
p_6	0.5000	0.7638	0.5775

	TTTT	TTTC	TTCT
p_1	6.0000	6.0000	6.0000
p_2	4.3301	4.3301	4.3301
p_3	2.8819	2.8819	3.0972
p_4	1.6160	1.9330	2.3660
p_5	0.7292	1.2503	1.4422
p_6	0.2887	0.5774	0.5774

Table 6. Atomic ID Numbers and Molecular ID Numbers for All Vertices of Embedded Paths of Figure 2

	TTTT	TTTC	TTCT
1	5.3518	5.2495	5.1027
2	5.4857	5.3351	5.2979
3	5.4800	5.3640	5.1487
4	5.4959	5.2806	5.4959
5		5.4800	5.4800
6		5.4857	5.0038
7		4.7676	5.0655
ID no.	38.1309	36.9625	36.5945

	TTCC	TCTC	TCCT
1	4.8351	5.0617	4.6044
2	5.0271	5.0038	4.4949
3	4.8317	5.3640	5.1487
4	5.2806	5.2806	5.4959
5	5.4800	5.1487	
6	5.0038	5.2979	
7	3.9949	4.7676	
ID no.	34.4532	35.9243	33.9919

	TCCC	CTCC	CTTC
1	4.3158	4.5812	4.5812
2	4.1657	5.0271	5.3351
3	4.8317	4.8317	5.3640
4	5.2806	5.0654	5.0654
5	5.1487	5.3640	
6	4.4949	5.0038	
7	3.4544	4.0722	
ID no.	31.6618	28.9182	35.6258

they are fused in the *cis* conformation. The same occurs again for other pairs of vertices which partition a conformer into fragments that are fused in alternative fashion. The terminal vertices appear for the most part to be unique, except in TTTC and TCTC when vertices 7 in both case have the same atomic ID. This exception can be understood when one examines the distances between the terminal vertex 7 in both graphs: The corresponding distances in the two graphs are equal, since the two conformers differ in reflection about a plane passing vertices 3, 4, and 7.

SOME MATHEMATICAL PROPERTIES

A close look at the characteristic polynomial $P(\lambda) = (-1)^n$ det$(\mathbf{DD} - \lambda\mathbf{I})$ of the distance/distance matrices shows some parallelism: As the eigenvalues decrease along a series of conformers, so decrease the coefficients of the characteristic

polynomial. In general, for graphs (or weighted graphs) there is no simple (i.e., linear) relationship betwee the coefficients of the polynomial and its first eigenvalue. Here for **DD** matrices of embedded paths we find a strong correlation between the Perron root λ_1 and the coefficients of $P(\lambda)$. In Figure 3 we illustrate the correlations of λ_1 with the coefficients of the powers x^4, x^3, and x^2 of the characteristic polynomials for the six conformers of P_6. The corresponding statistical parameters are the coefficient of regression $R = 0.9949$, the standard error of estimate $S = 0.149$, and the F ratio is 3900. In Figure 4 we illustrate the correlations of the coefficients of the characteristic polynomials among themselves. Here for the same six conformers of P_6 we show first the regression of the coefficients of x^4 against the coefficients of x^3, then x^3 against x^2, and so on. Now the coefficients of the regressions are even larger. For the regression of x^4 against x^3 we have $R = 0.9986$ (the standard error $S = 0.1595$ and F ratio is 3641).

This interesting and intriguing property of **DD** matrices deserves further study. If this proves to be a general property of such matrices, at least for the matrices of embedded graphs on a graphite network, it would offer a simple relationship between paths derived by different embeddings of P_n. It would allow one to characterize each such family of embedded paths by a single member. Then one could assign to each other member in a family a single scaling factor given by the quotient of the first eigenvalues of the embedded graphs.

STRUCTURE–PROPERTY RELATIONSHIP

There is a limited number of properties that one can correlate with novel invariants. Most thermodynamic physicochemical data belong to a mixture of conformations, not to a single conformation. Hence, such data are not relevant here. One can *compute* several properties of embedded graphs. These include various moments, including the moments of inertia (second moments), the end-to-end distances, and the molecular surface. Instead, however, we will examine the degree of molecular similarity as derived form the path numbers.
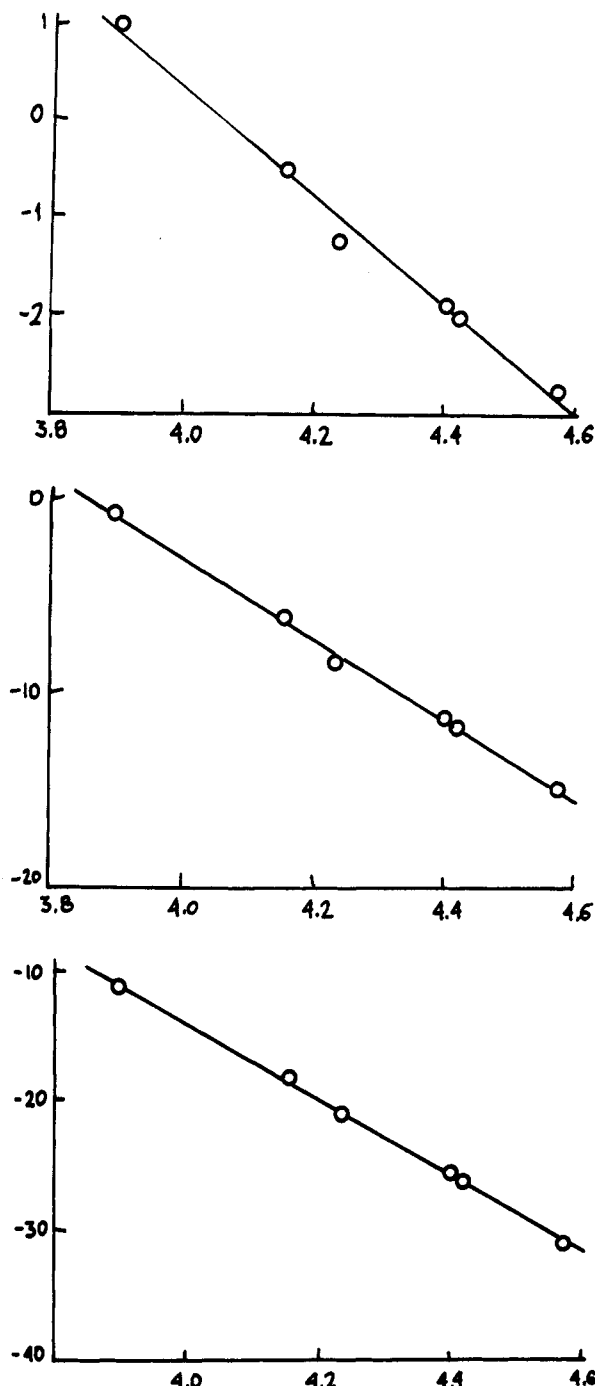
282 *J. Chem. Inf. Comput. Sci., Vol. 34, No. 2, 1994*

RANDIC ET AL.

Figure 3. Correlation between (top) the first eigenvalue and the constant term in the characteristic polynomials of the **DD** matrices for linear chains of six vertices (embedded on a graphite lattice); (middle) the first eigenvalue and the coefficient of the linear term; (bottom) the first eigenvalue and the coefficient of the quadratic term.

Similarity among Conformers. Because similar molecules are expected to have similar properties,[21] which underlines the Emil Fisher "lock and key" model of drug–receptor interaction, the notion of molecular similarity is of paramount importance. In Figure 2 we have shown all nine conformations for P_7. We would like to find the degree of similarity between pairs of conformers. Just by mere inspection of Figure 2 it is not clear how to determine which pair of structures are the most similar. In Table 7 we give the numerical values for the similarity/dissimilarity table for the nine configurations of P_7. The table was based on Euclidean metrics and use of the path numbers for the characterization of individual molecules. It is not surprising to see that the largest entry in the table,

1.656, belongs to the pair TTTT and TCCC, indicating them as the least similar. For other less similar pair of structures we find the following: TTTC–TCCC (1.381); TTCT–TCCC (1.267); TCTC–TCCC (1.141), and so on. Figure 5 confirms visually that the results agree with out intuitive notions of similarity. Although the lack of similarity among structures is generally of lesser interest, it helps to illustrate that the model considered is plausible. It is more important to find structures that are most similar. Among the most similar pairs structures we find the following: TTTC–TTCT (0.136); TCTC–CTTC (0.233); TTCC–TCCT (0.237), TTCT–TCTC (0.245), and TTCC–CTCC (0.256). Visual inspection of the corresponding molecular diagrams (Figure 6) shows that the results are very plausible.

FOLDING OF LONG PATHS

Let us examine closer the eigenvalues of **DD** matrices. The examples (Table 3) show that in most cases all eigenvalues, except the Perron root, are negative, the property known to hold for the complete graphs K_n. This, again, is not unexpected, since most of off-diagonal entries in **DD** matrices (for the cases illustrated) have values close to 1. The exceptions to the above regularities are highly "bent" or "folded" graphs, like all-cis-P_6 and the closely related graphs TCCT and TCCC of P_7. This suggests that the characteristic polynomials may be indicative of the "degree of folding" of the embedded path graphs. Because of the parallelism between the Perron root λ_1 and the coefficients of the characteristic polynomials, we may expect that λ_1 as a measure of the folding of the embedded path.

There is a distant analogy here with the interpretation of first eigenvalue of the adjacency matrix. Lovasz and Pelikan[20] proposed the first eigenvalue λ_1 of the adjacency matrix to be an index of molecular branching. To arrive, however, at a useful index for the degree of molecular folding (of a long path), we have to eliminate the size dependence of λ_1 on n. One would like larger λ_1 to indicate less folded graph. However, as illustrated on all-trans-P_5 and all-cis-P_6, we find that a smaller paths can have a smaller first eigenvalue than a larger "folded" path later, $\lambda_1 = 3.692$ and $\lambda_1 = 3.897$, respectively. This is because we are comparing graphs of *different* length. To get a satisfactory index, we have to normalize the eigenvalue λ_1 by dividing it by n, the number of vertices, to get an index $\phi = \lambda_1/n$. If we now compare the normalized eigenvalues as a measure of folding to the above mentioned two graphs, we find that all-trans-P_5 (TT) and all-cis-P_6 (CCC) have $\phi = 0.738$ and $\phi = 0.650$, respectively. This is satisfactory since now the more "linear" isomer (regardless of its size) has a larger index ϕ than the folded structure, even if of larger size.

One would like that the extreme $\phi = 1$ represents linear (unfolded) chain and, if possible, that $\phi = 0$, represents "the most folded" chain (if such can be defined). This would indicate that the proposed index is a "good" measure of molecular folding. In Table 8 we list ϕ values, the index of folding, for the structures of Figure 1 and Figure 2. The results support the notion that ϕ is a valid index of the degree of folding of a chain. The numerical values in Table 8 are intermediate, i.e., close neither to 1 nor to 0, the theoretical extreme values. Nevertheless the ϕ values give plausible results. Thus we arrived at quantifying yet another structural molecular feature, molecular folding.

The parameter ϕ can be of interest not only as a global measure, but also, when applied to a local molecular fragment, as a local index of structural folding. It may be of interest
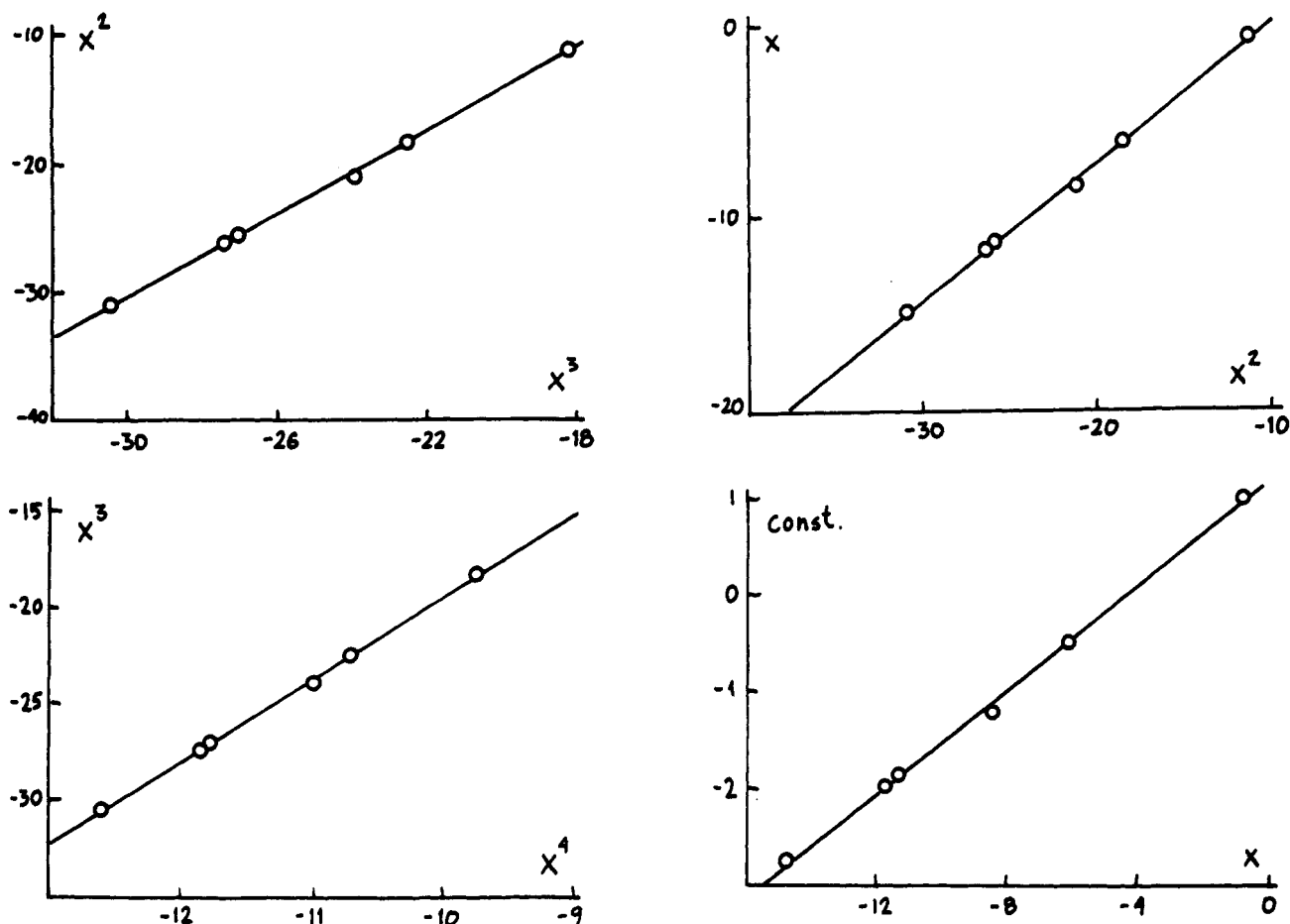
DISTANCE/DISTANCE MATRICES

J. Chem. Inf. Comput. Sci., Vol. 34, No. 2, 1994 **283**

Figure 4. Correlations between coefficients of characteristic polynomials for linear chains of six vertices (embedded on a graphite lattice). The corresponding powers are indicated on the coordinate axes.

Table 7. Similarity/Dissimilarity Matrix among Configurations of Figure 2 (The Smaller the Entry in the Table the More Similar Are the Corresponding Structures)

	TTTT	TTTC	TTCT	TTCC	TCTC	TCCT	TCCC	CTCC	CTTC
TTTT	0	0.305	0.399	0.930	0.605	1.087	1.656	1.059	0.643
TTTC		0	0.136	0.649	0.325	0.817	1.381	0.799	0.342
TTCT			0	0.540	0.245	0.696	1.267	0.696	0.298
TTCC				0	0.428	0.237	0.742	0.256	0.379
TCTC					0	0.591	1.141	0.505	0.233
TCCT						0	0.582	0.288	0.579
TCCC							0	0.675	1.096
CTCC								0	0.520
CTTC									0

for characterization of local molecular environments, such as in clover leaf folding of t-RNA, in the folding of proteins, or in peptides, for characterization of polymers, and for characterization of other chainlike structures in chemistry and biochemistry. Therefore it is important that we further investigate the properties of the novel index of folding. In particular, it is of interest to find which structures will lead to the extreme values of $\phi = 1$ and $\phi = 0$.

Extreme Case $\phi = 1$. It is not difficult to see that a strictly linear (embedded) chain will have the extreme value of $\phi = 1$. We have seen that the **DD** matrix for a strictly linear chain of length 2 (P_3) is the same as that of the complete graph K_3. This parallelism extends to paths of any size P_n and complete graph K_n. Since the Perron root of such matrices is $\lambda_1 = n - 1$, the folding index $\phi = 1$. Hence, a strictly linear structure will have the folding index 1, showing no deviation from linearity.

Extreme Case $\phi = 0$. One may expect the folding index 0 for structures that most depart from linear geometry. It is,

however, less clear how such structures may look. Comparison of all *cis* structures of increasing size shows that indeed the index decreases as the folding increases. This is numerically demonstrated in Table 9. Here, we permit a structure to fold upon itself. Thus, by continuing the folding process we create a structure that can be viewed as the projection of a helical structure. If we extend the structures along the z-axis (perpendicular to the honeycomb lattice plane), we get a three-dimensional helical structure. As we see from Table 9, the values of ϕ decrease with the increase in the length of the helix. This again suggests that the Perron root is a useful measure of the degree of departure from linearity, hence a measure of the degree of folding. The convergence of ϕ to zero is slow, so doubts may be raised if we can ever reach $\phi = 0$.

An alternative approach in a search for the elusive "most nonlinear structure", i.e., a structure that would depart most, or be perceived to depart the most, from a strictly linear structure is to consider folding of a path on itself. For example,
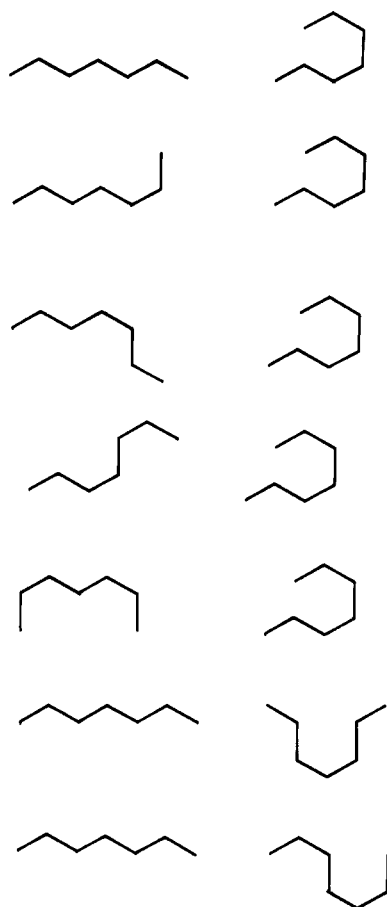
284 *J. Chem. Inf. Comput. Sci., Vol. 34, No. 2, 1994*

RANDIC ET AL.

Figure 5. The least similar conformations among chains having seven vertices superimposed on a graphite lattice.

consider P_7 and fold it completely so that it overlaps itself. The corresponding **DD** matrix then is

$$
\begin{pmatrix}
0 & 1 & 0 & 1/3 & 0 & 1/5 & 0 & 1/7 \\
1 & 0 & 1 & 0 & 1/3 & 0 & 1/5 & 0 \\
0 & 1 & 0 & 1 & 0 & 1/3 & 0 & 1/5 \\
1/3 & 0 & 1 & 0 & 1 & 0 & 1/3 & 0 \\
0 & 1/3 & 0 & 1 & 0 & 1 & 0 & 1/3 \\
1/5 & 0 & 1/3 & 0 & 1 & 0 & 1 & 0 \\
0 & 1/5 & 0 & 1/3 & 0 & 1 & 0 & 1 \\
1/7 & 0 & 1/5 & 0 & 1/3 & 0 & 1 & 0
\end{pmatrix}
$$

With $\lambda_1 = 2.2850$ this gives $\phi = 0.326$. This is (for the size of the path) a relatively small value of ϕ. In Table 10 we list the values for the higher members of this highly folded hypothetical structure. Now one has some basis to expect that in the limit we may get $\phi = 0$. This would support our interpretation that ϕ represents a good index of the molecular folding.

DISCUSSION

By establishing the extreme ϕ-structures, the strictly linear chain and fully overlapping folded chain, we only ensured that the extreme values of the index agree with our notions for a measure of the structural folding. We also have to examine intermediate cases so to see how well ϕ characterizes the molecular folding. Could, for example, ϕ be an index of molecular shape rather than folding? Such questions are not easy to answer, as we deal with concepts, the folding of a molecule and the molecular shape, that themselves are not yet fully defined. In fact, the consideration of the shape and the folding indices hopefully may lead to a rigorous mathematical
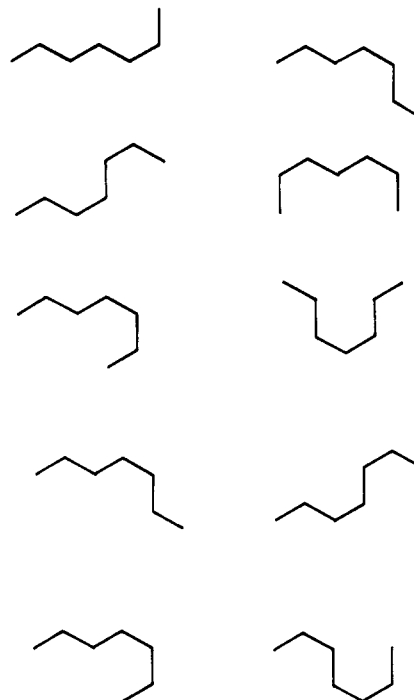

Figure 6. The most similar conformations among chains having seven vertices superimposed on a graphite lattice.

Table 8. Folding Index ϕ for All Structures of Figures 1 and 2

conformer	ϕ	conformer	ϕ
T	0.7019	CCC	0.6496
C	0.6767	TTTT	0.7783
TT	0.7384	TTTC	0.7555
TC	0.7130	TTCT	0.7476
CC	0.6738	TTCC	0.7079
TTT	0.7260	TCTC	0.7340
TTC	0.7365	TCCT	0.6971
TCT	0.7338	TCCC	0.6558
TCC	0.6927	CTCC	0.6965
CTC	0.7182	CTTC	0.7297

Table 9. First Eigenvalue and Normalized λ/n (folding index ϕ) for *all-trans* and *all-cis* Conformations of Chains of Different Length Superimposed on a Graphite Lattice

| path	all-trans		path	all-cis	
	λ	ϕ		λ	ϕ
T_4	2.8079	0.7019	C_4	2.7068	0.6767
T_5	3.6920	0.7384	C_5	3.3690	0.6738
T_6	4.5719	0.7620	C_6	3.8974	0.6496
T_7	5.4479	0.7783	C_7	4.3020	0.6146
T_8	6.3220	0.7903	C_8	4.6388	0.5798
T_9	7.1994	0.7994	C_9	4.9350	0.5483
T_{10}	8.0659	0.8066	C_{10}	5.2046	0.5204
T_{11}	8.9363	0.8124	C_{11}	5.4520	0.4956
T_{12}	9.8063	0.8172	C_{12}	5.6783	0.4732
T_{13}	10.6756	0.8212	C_{13}	5.8806	0.4524

definition of these important structural concepts. One should be aware that different authors have different notions about what some not yet fully defined structural concepts represent. This is already illustrated by the lack of agreement, for example, of what the attributes like benzenoid, aromatic, etc. represent. For some, benzenoid represents structures made by formal fusion of benzene rings;[25] others prefer that such structures have resemblance to benzene, hence have Kekule valence structures.[26] Some researchers based their definition of aromaticity on HMO (Hueckel molecular orbital model);[27] others prefer the conjugated circuit model.[28] The situation is somewhat similar to the notion of shape, which recently has

DISTANCE/DISTANCE MATRICES

J. Chem. Inf. Comput. Sci., Vol. 34, No. 2, 1994 **285**

Table 10. Folding Index ϕ for Chains of Different Length Repeatedly Folded over a Single Link

length	ϕ	length	ϕ
2	0.4714	14	0.2031
4	0.3887	16	0.1947
6	0.3264	18	0.1797
8	0.2820	20	0.1670
10	0.2491	22	0.1561
12	0.2237	24	0.1468

attracted some attention. Kier[29] speaks of the shape of graphs. One may refer to such as the topological shape (analogous to the notion of the topological distance, which is legitimate). Here shape reflects some measure of variations of graph topological distances within different graphs. Mezey[30] considers continuous objects and the topography of surfaces of such objects. We are inclined to reserve the term "shape" (in the case of discrete objects, such as molecular structure) to graphs *embedded* on two-dimensional or three-dimensional grids (and in the general case to graphs embedded in space, regardless of the presence of a regular underlying grid). Under these circumstances does ϕ measure molecular shape or measure molecular folding?

To see the difference, consider a comparison between the structures shown below:
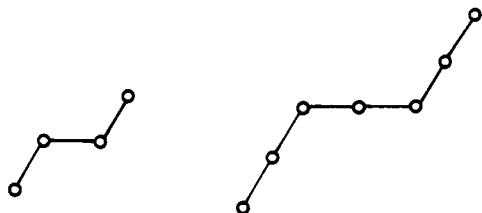

One may argue that the two (embedded) graphs have the same shape but differ only by the scaling. Such viewing is correct if we ignore the "inside" vertices, i.e., if we only see the outside contour of the structure. After constructing the **DD** matrix for the above structures and finding the Perron root, we obtain the following ϕ values respectively for the smaller (*trans*-butadiene) and the larger structures:

trans-butadiene: $\quad \lambda_1 = 2.8079 \quad \phi = 0.702$

larger structure: $\quad \lambda_1 = 5.5816 \quad \phi = 0.797$

Hence, surely the two structures (of the same shape) have different values of ϕ. Thus ϕ does not encode molecular shape (at least not in the above sense, when inner structure is ignored). The larger structure has larger ϕ, and it is interpreted to have a smaller deviation form linearity. One can argue convincingly that indeed this is the case, since relatively speaking fewer of its atoms depart from linearly fused local conformation.

To further support the interpretation of ϕ as an index of molecular folding, consider three structures, each having a single "kink", derived by different embedding of P_7, as depicted in Figure 7. We want again to show that ϕ does not measure "kinkiness" of a structure; the three structures all have a single such structural element. However, one can speak of different degree of "departure" of a structure from the linear form even when structures have a single "kink". To see that this makes sense, view the vertices in such structures as data points and use statistical analysis to find, using the least-squares method, a line that is most close to the data points. The three cases then give different answers. The smallest standard error, hence, the smallest departure from strict linearity or the smallest amount of folding, is expected for the first structure
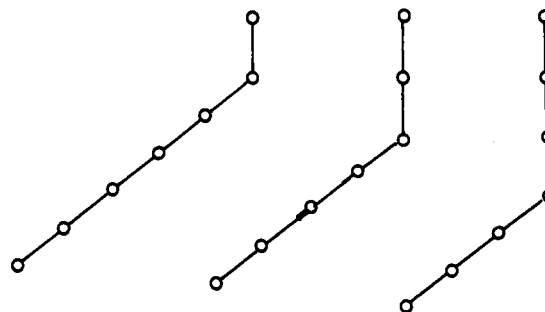
Figure 7.

which has the kink at the end of the molecule. After constructing **DD** matrices for the three cases, we find the following λ_1 and ϕ values:

$$\lambda_1 = 5.8588 \qquad \lambda_1 = 5.7370 \qquad \lambda_1 = 5.6905$$

$$\phi = 0.837 \qquad \phi = 0.820 \qquad \phi = 0.813$$

The above supports our notion that ϕ is a measure of molecular folding. It agrees with out expectations and further illustrates that the novel index ϕ not only measures molecular folding, i.e., the departure of molecular structure from strictly linear chain, but is a quite sensitive index of such folding.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Balaban, A. T. Ed. *Chemical applications of graph theory*; Academic Press: London, 1976. King, R. B., Ed. *Chemical applications of topology and graph theory*; Studies in Physical and Theoretical *Chemistry* 28; Elsevier: Amsterdam, 1983. Balasubramanian, K. Applications of combinatorics and graph theory to spectroscopy and quantum chemistry. *Chem. Rev.* 1985, *85*, 599–618. Trinajstić, N. *Mathematics and computational concepts in chemistry*; Ellis Horwood: Chichester, 1986. King, R. B., Rouvray, D. H., Ed. *Graph theory and topology in chemistry*; Studies in *Physical and Theoretical Chemistry* (Elsevier: Amsterdam, 1987. Lacher, L. C., Ed. MATH/CHEM/COMP 1987. Studies in *Physical and Theoretical Chemistry* 54; Elsevier: Amsterdam, 1988. Rouvray, D. H., Ed. *Computational graph theory*; Nova Sci. Publ.: New York, 1990. Klein, D. J.; Randić, M., Ed. *Mathematical Chemistry*, VCH: Weinheim, 1990; Also appeared as *J. Math. Chem.* 1990, *4*. Graovac, A., Ed. MATH/CHEM/COMP 1988. Studies in *Physical and Theoretical Chemistry* 63; Elsevier: Amsterdam 1989.

(2) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992.

(3) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, Ov. Topological indices for structure-activity correlations. *Top. Curr. Chem.* 1983, *114*, 21–55. Randić, M. In search of structural invariants. *J. Math. Chem.* 1992, *9*, 97–146. Randić, M.; Trinajstić, N. In search of graph invariants of chemical interest. *J. Mol. Struct. (Theochem)*, in press.

(4) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical and combined molecular descriptors in QSAR/QSPR research, *J. Chem. Inf. Comput. Sci.* 1993, *33*, 835–857. Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R.; Veith, G. D. Topological indices: Their nature, mutual relatedness and applications. *Math. Modell.* 1986, *8*, 300–308. Randić, M. Comparative regression analysis. Regressions based on a single descriptor. *Croat. Chem. Acta* 1993, *66*, 289–312. Randić, M.; Trinajstić, N. Viewpoint 4-Comparative structure property studies: the connectivity basis. *J. Mol. Struct. (Theochem)* 1993, *284*, 209–221.

(5) Hosoya, H. Topological Index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* 1971, *44*, 2332–2339.

(6) Wiener, H. Structural determination of the paraffin boiling points. *J. Am. Chem. Soc.* 1947, *69*, 17–20.

(7) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* 1975, *97*, 6609–6619.

(8) Kier, L. B.; Hall, L. H. Molecular connectivity VII. Specific treatment of heteroatoms. *J. Pharm. Chem.* **1976**, *65*, 1806–1809. Kier, L. B.; Hall, L. H. *Molecular connectivity in chemistry and drug research*; Academic Press: New York, 1976. Kier, L. B.; Hall, L. H. *Molecular connectivity in structure-activity analysis*; Research Studies Press: Letchworth, 1986.

(9) Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure-activity relationship. *Chemometrics Intel. Lab. Syst.* **1991**, *10*, 213–227. Randić, M. On computation of optimal parameters for multivariate analysis of structure-property relationship. *J. Comput. Chem.* **1991**, *12*, 970–980. Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen containing molecules. *J. Math. Chem.* Submitted for publication. Randić, M.; Basak, S. Search for empirical descriptors for heteroatom in QSAR. Work in progress.

(10) Randić, M. Molecular topographic descriptors. *Stud. Phys. Theor. Chem.* **1988**, *54*, 101–108. Randić, M. On the characterization of three-dimensional structures. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1988**, *15*, 201–208. Randić, M.; Jerman-Blažič, B.; Trinajstić, N. Development of 3-dimensional molecular descriptors. *Comput. Chem.* **1990**, *14*, 237–246. Mihalić, Z.; Trinajstić, N. *J. Mol. Struct. (Theochem)* **1991**, *232*, 65. Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the three-dimensional Wiener number. *J. Math. Chem.* **1983**, *3*, 299. Bogdanov, B.; Nikolić, S.; Trinajstic, N. On the three-dimensional Wiener number. A comment. *J. Math. Chem.* **1990**, *5*, 305–306. Nikolić, S.; Trinajstić, N.; Mihalić, M.; Carter, S. *Chem. Phys. Lett.* **1991**, *179*, 21. Balasubramanian, K. Geometry dependent characteristic polynomial of molecular structures. *Chem. Phys. Lett.* **1990**, *169*, 224–230.

(11) Hardy, G. H.; Littlewood, J. E.; Polya, G. *Inequalities*; Cambridge University Press: Cambridge, 1964.

(12) Hausdorf, F. *Set Theory*; Chelsea: New York, 1962.

(13) Hamming, R. W. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147–160.

(14) Mahalanobis, P. C. On the generalized distance in statistics, *Proc. Natl. Inst. Sci., Calcutta* **1936**, *12*, 501–514.

(15) Klein, D. J.; Randić, M. Resistance distance. *J. Math. Chem.* **1993**, *12*, 81–95.

(16) Buckley, F.; Harary, F. *Distance in Graphs*; Addison Wesley: Reading, MA, 1990.

(17) Balaban, A. T. Highly discriminant distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404. Balaban, A. T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206.

(18) Balaban, A. T.; Quintas, L. V. The smallest graphs, trees and 4-trees with degenerate topological index. *MATCH* **1983**, *14*, 213–233.

(19) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235–250. Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Design of topological indices. Part 4, Reciprocal matrix, related local index invariants and topological indices. *J. Math. Chem.* **1993**, *12*, 309–318.

(20) Horn, R. A.; Johnson, C. R. *Matrix analysis*; Cambridge Univ. Press: Cambridge, 1985.

(21) Horn, R. A.; Johnson, C. R. *Topics in Matrix Analysis*; Cambridge Univ. Press: Cambridge, 1991.

(22) Lovasz, L.; Pelikan, J. *Period. Math. Hung.* **1973**, *3*, 175.

(23) Widom, H. Toeplitz Matrices. *Studies in Real and Complex Analysis*; Math. Assoc. Am. (distributed by Prentice-Hall Inc.), 1965; pp 179–209.

(24) Randić, M. Design of Molecules with Desired Properties, A Molecular Similarity Approach to Property Optimization. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G., Eds.; Wiley: New York, 1990. Randić, M. The nature of the chemical structure. *J. Math. Chem.* **1990**, *4*, 157–184. Jerman-Blažič, B.; Fabič-Petrač, I.; Randić, M. Evaluation of the molecular similarity and property prediction for QSAR purposes. *Chemometrics Intel. Lab. Syst.* **1989**, *6*, 49–63. Jerman-Blazic, B.; Randic, M. Similarity measures for set of strings and application in chemical classification. *J. Math. Chem.* **1990**, *4*, 217–225.

(25) Gutman, I.; Cyvin, S. J. *Introduction to the theory of benzenoid hydrocarbons*; Springer Verlag: Berlin, 1989.

(26) Randić, M.; Nikolić, S.; Trinajstić, N. On the benzenoid character of polycyclic conjugated hydrocarbons. *Gazz. Chim. Ital.* **1987**, *117*, 69–73.

(27) Hueckel, E. Quantentheoretische Beitrage zum Benzolproblem. I. Die Elektronenkonfigurationen des Benzols und verwandter Verbindungen. *Z. Phys.* **1931**, *70*, 204.

(28) Randić, M. Aromaticity and conjugation. *J. Am. Chem. Soc.* **1977**, *99*, 444–450.

(29) Kier, L. B. Indexes of molecular shape from chemical graphs. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Publ.: Commacks, NY, 1990.

(30) Mezey, P. G. Non-visual molecular shape analysis: Shape changes in electronic excitations and chemical reactions. In *Computational Advances in Organic Chemistry (Molecular structure and reactivity)*; Ogretir, C., Csizmadia, I. G., Eds.; Kluwer: Dordrecht, 1991.

Razinger, M.; Chretien, J. R.; Dubois, J. E. Structural selectivity of topological indices in alkane series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23–27.