Pharmaceutical Research, Vol. 6, No. 9, 1989

Report

A Graph-Theoretical Approach to the Prediction of Physical Properties of Alkanes Based on the Distance Matrix

James F. Caputo[1,3] and Kathleen J. Cook[2]

Received August 22, 1988; accepted April 3, 1989

A new topological index, the largest eigenvalue of the distance matrix (*DI*), is presented as a measure of molecular branching. The *DI* and Balaban's *J* index are used to predict the densities of a series of alkanes. The statistical correlations obtained are excellent and give a correlation coefficient of 0.961.

KEY WORDS: topological index; distance matrix; graph theory; eigenvalues; density; alkanes; *J* index.

INTRODUCTION

The prediction of physical properties of molecules based on their chemical structure has long been a goal of the chemist. During the last 10 years there has been a resurgence of interest in the use of graph-theoretical methods. The majority of research in this area has been focused on the development of topological indices derived from the hydrogen-suppressed graph representation of the chemical structure of the molecular species of interest. The most successful topological indices used, in this regard, have been derived from the adjacency matrix representation of the chemical graph, such as Randic's connectivity index (1), and from Kier's and Hall's extended connectivity indices (2,3). The Kier and Hall extended connectivity indices have been extensively used in the development of quantitative structure–property relationship (QSPR) and in the development of quantitative structure–activity relationships (QSAR) (4–8).

Topological indices derived from the distance matrix of the chemical graph are not as well known as those derived from the adjacency matrix, with the exception of Wiener's path number (9). Distance-based topological indices include Balaban's *J* index (10), Hosoya's distance polynomial, the Altenburg polynomial, and Bonchev's and Polansky's distance sum index (11). Of these indices, Balaban's *J* index is of particular interest because it is one of the least degenerate indices known. In attempting to develop quantitative relationships between chemical structure (as represented by its hydrogen-suppressed graph) and its physical properties, it is advantageous to have a unique index. In addition, it is preferable to have an index that quantifies a known aspect of the topological structure. Balanban's *J* index is known to be related to the branching in a structure (10). In the present work a new topological index is being introduced. This new index is defined as the largest eigenvalue of the distance matrix and has been named the distance index, *DI*. The *DI* and the *J* index, both of which reflect molecular branching, are correlated with the densities of a series of alkanes. Density was chosen as a model property because it is sensitive to molecular branching.

The *DI* is being proposed as a new topological index which also reflects the degree of branching in alkanes. This new index has the advantage in that it relates directly to both the amount and the kind of branching and can be specified in a structural context. Thus, it is anticipated that the *J* index and the *DI* should be ideal topological indices to develop correlations with those physical properties of alkanes that are reflected by their molecular branching.

THEORETICAL

Both Balaban's *J* and the *DI* are derived from the distance matrix of the hydrogen-suppressed graph. The distance matrix can be defined in terms of its diagonal elements $d_{i,i}$ and its off-diagonal elements $d_{i,j}$. The $d_{i,i}$ elements are equal to zero and the $d_{i,j}$'s are equal to the number of edges between the *i*th and the *j*th vertices in the shortest path between them. In the chemical context edges are bonds and vertices are atoms. Thus, Balaban's *J* index can be defined by

$$J = \frac{q}{(\mu + 1)} \Sigma(S_i S_j)^{-1/2} \quad (1)$$

where μ is the cyclomatic number (the number of rings in the structure), q is equal to the number of edges, and the S_i's represent the sum of all the elements in the *i*th row of the distance matrix. The summation extends over all adjacent pairs of vertices.

The distance matrix was calculated from the adjacency matrix by making use of Dijkstra's algorithm (12) and implemented as a Fortran program. The eigenvalues of the distance matrix were determined with the aid of a SAS com-

---

[1] Reproduction and Growth Physiology, The Upjohn Company, 301 Henrietta Street, Kalamazoo, Michigan 49001.

[2] Drug Metabolism Research, The Upjohn Company, Kalamazoo, Michigan 49001.

[3] To whom correspondence should be addressed.

Table I. The *DI* and *J* Indices

Name	*DI*	DI^{-1}	*J* index
Pentane	8.2882	0.1207	2.9923
2-Methylpentane	7.4593	0.1341	2.5395
2,2-Dimethylpropane	6.1764	0.1619	3.0237
2-Methylpentane	11.0588	0.0904	2.6272
3-Methylpentane	10.7424	0.0931	2.7542
2,2-Dimethlyhexane	9.6702	0.1034	3.1685
2,3-Dimethylbutane	9.6088	0.1041	3.5491
2,4-Dimethylhexane	13.8216	0.0724	2.9532
2,2-Dimethylpentane	13.6353	0.0733	3.1545
2-Methylhexane	15.4048	0.0649	2.6783
Heptane	16.6254	0.0601	2.4475
3-Methylhexane	14.8636	0.0673	2.8318
2,2,3-Trimethylbutane	12.0449	0.0830	3.5412
3,3-Dimethylpentane	13.0698	0.0765	3.3604
2,3-Dimethylpentane	13.6346	0.0733	3.1442
3-Ethylpentane	14.2969	0.0699	2.7542
Octane	21.8364	0.0458	2.5301
2-Methylheptane	20.4792	0.0488	2.7158
3-Methylheptane	19.7628	0.0506	2.8621
4-Methylheptane	19.5420	0.0512	2.9196
2,2-Dimethylhexane	18.4133	0.0543	3.1118
3,3-Dimethylhexane	17.4426	0.0573	3.3734
2,3-Dimethylhexane	18.1815	0.0550	3.1708
2,4-Dimethylhexane	18.3964	0.0544	3.0988
2,2,3-Trimethylpentane	16.3152	0.0613	3.6233
2,2,3-Trimethylpentane	16.7054	0.0599	3.3889
2,3,4-Trimethylpentane	16.4963	0.0606	3.4642
2,3,3-Trimethylpentane	16.0683	0.0622	3.7083
2-Methyl-3-ethylpentane	17.4187	0.0574	3.3549
3,4-Dimethylexane	17.6759	0.0566	3.2925
3-Methyl-3-ethylpentane	16.6705	0.0600	3.5832
3-Ethylhexane	18.7788	0.0533	3.0744
Nonane	27.7422	0.0360	2.5951
2-Methyloctane	26.2722	0.0381	2.7467
3-Methyloctane	25.4119	0.0394	2.8766
2,2-Dimethylheptane	23.9634	0.0417	3.0730
2,3-Dimethylheptane	23.5541	0.0425	3.1553
2,4-Dimethylheptane	23.5441	0.0425	3.1512
2,2,3-Trimethylhexane	21.2250	0.0471	3.5887
2,3,4-Trimethylhexane	21.1970	0.0472	3.5758
3-Ethylheptane	24.0988	0.0415	3.0922
3,3-Dimethylheptane	22.6772	0.0441	3.3301
2,2,6-Trimethylhexane	28.6670	0.0349	3.2055
2,2,5-Tirmethylhexane	22.4662	0.0445	3.2807
2,2,4-Trimethylhexane	21.6063	0.0463	3.4673
2,3,5-Trimethylhexane	22.0627	0.0453	3.3766
2,2,3,4-Tetramethylpentane	19.7257	0.0507	3.8776
2-Methyl-4-ethylhexane	22.6204	0.0442	3.3074
2,2-Dimethyl-3-ethylpentane	20.2981	0.0493	3.7929
4,4-Dimethylheptane	22.2705	0.0449	3.4311
2,6-Dimethylheptane	24.7896	0.0403	2.9147
2,5-Dimethylheptane	23.9292	0.0418	3.0608
3,4-Dimethylheptane	22.6789	0.0441	3.3248
2,4-Dimethyl-3-ethylpentane	20.7438	0.0482	3.6776
2,2,5-Trimethylhexane	22.4662	0.0445	3.2807
2,3,3-Trimethylhexane	20.7945	0.0481	3.7021
2-Methyl-3-ethylhexane	22.2198	0.0450	3.4101
2,3,4-Trimethylhexane	21.1970	0.0472	3.5758
2,2,3,3-Tetramethylpentane	18.8440	0.0531	4.1447
3-Methyl-4-ethylhexane	21.7527	0.0460	3.4995
3,3,4-Trimethylhexane	20.3172	0.0492	3.8024
4-Ethylheptane	23.6799	0.0422	3.1753
2,3,3,4-Tetramethylpentane	19.3005	0.0518	4.0137
2,3-Dimethyl-3-ethylpentane	18.9563	0.0504	3.9192
4-Methyloctane	25.0207	0.0400	2.9548
2,4,6-Trimethylheptane	27.9451	0.0358	3.3374
3,3-Diethylpentane	20.3923	0.0490	3.8247
2,2-Dimethyl-4-ethylhexane	26.1010	0.0383	3.6308
3,4-Dimethyl-3-ethylhexane	21.3349	0.0469	3.6174
3,4-Dimethylheptane	22.6789	0.0441	3.3248
3-Methyl-3-ethylhexane	21.3349	0.0469	3.6174
2,2,5-Trimethylheptane	27.6823	0.0361	3.3555
2,2-Dimethyloctane	30.2636	0.0330	3.0438
2,4-Dimethyloctane	29.5290	0.0339	3.1600
2,7-Dimethyloctane	31.1984	0.0321	2.9095
4-Isopropylheptane	27.3408	0.0366	3.4999
4-Propylheptane	29.4624	0.0339	3.2055
5-Methylnonane	31.0969	0.0322	2.9984
Undecane	41.6384	0.0240	2.6909
Dodecane	49.6287	0.0201	2.7272

puter program using the SAS Interactive Matrix Language (13).

RESULTS

A two-term regression model, plus the intercept, resulted in an excellent correlation of alkane densities with *J* and *DI*. The regression model is

$$D = -1.338DI^{-1} + 0.327J + 0.664$$
$$r = 0.9608, \quad s = 0.0086, \quad n = 77 \qquad (2)$$

The choice to include a term in the model was based on the *F* test for each term. In this case, the *F* value was significant at a *P* value of 0.0001 for both *J* and DI^{-1}. Table I gives the values for both *J* and *DI*. Table II gives the results of the density correlation based on the model given by Eq. (2).

Correlations of density with just DI^{-1} were not very impressive. Equation (3) gives the results:

$$D = -1.1338DI^{-1} + 0.7743$$
$$r = 0.867, \quad s = 0.154, \quad n = 7 \qquad (3)$$

The correlation of density with *J* or with a J^{-1} term resulted in correlations with an *r* value of approximately 0.5. Consequently, both *J* and DI^{-1} are important.

DISCUSSION

In the search to derive new topological indices, it is most useful to attempt to find those indices that can be clearly interpreted in terms of the structural elements of the molecule. To this end, the *DI* is well suited to interpretation. The *DI* exhibits a rather direct relationship to the branching pattern in the alkanes. This can best be seen by comparing the eight-carbon alkanes. These are represented by their hydrogen-suppressed graphs as shown in Fig. 1. As a point of

Table II. The Observed Versus Calculated Densities[a]

Name	Observed	Calculated	Residual
Pentane	0.6262	0.6337	−0.0075
2-Methylbutane	0.6197	0.6042	0.0155
2,2-Dimethylpropane	0.5910	0.5906	0.0004
2-Methylpentane	0.6532	0.6539	−0.0007
3-Methylpentane	0.6643	0.6553	0.0090
2,2-Dimethylbutane	0.6492	0.6581	−0.0089
2,3-Dimethylbutane	0.6616	0.6701	−0.0085
2,4-Dimethylpentane	0.6727	0.6842	−0.0115
2,2-Dimethylpentane	0.6738	0.6899	−0.0161
2-Methylhexane	0.6786	0.6830	−0.0044
Heptane	0.6838	0.6803	0.0035
3-Methylhexane	0.6871	0.6856	0.0015
2,2,3-Trimethylbutane	0.6901	0.6924	−0.0023
3,3-Dimethylpentane	0.6933	0.6933	0.0000
2,3-Dimethylpentane	0.6951	0.6895	0.0056
3-Ethylpentane	0.6982	0.6801	0.0181
Octane	0.7025	0.6985	0.0040
2-Methylheptane	0.6979	0.7015	−0.0036
3-Methylheptane	0.7058	0.7044	0.0014
4-Methylheptane	0.7046	0.7058	−0.0012
2,2-Dimethylhexane	0.6953	0.7088	−0.0135
3,3-Dimethylhexane	0.7100	0.7143	−0.0043
2,3-Dimethylhexane	0.7171	0.7101	0.0070
2,4-Dimethylhexane	0.7004	0.7083	−0.0079
2,2,3-Trimethylpentane	0.7160	0.7184	−0.0024
2,2,3-Trimethylpentane	0.6919	0.7121	−0.0202
2,3,4-Trimethylpentane	0.7191	0.7138	0.0053
2,3,3-Trimethylpentane	0.7262	0.7203	0.0059
2-Methyl-3-ethylpentane	0.7193	0.7136	0.0057
3,4-Dimethylexane	0.7192	0.7124	0.0068
3-Methyl-3-ethylpentane	0.7274	0.7185	0.0089
3-Ethylhexane	0.7136	0.7087	0.0049
Nonane	0.7176	0.7111	0.0065
2-Methyloctane	0.7134	0.7140	−0.0006
3-Methyloctane	0.7207	0.7170	0.0037
2,2-Dimethylheptane	0.7105	0.7210	−0.0105
2,3-Dimethylheptane	0.7260	0.7230	0.0030
2,4-Dimethylheptane	0.7160	0.7228	−0.0068
2,2,3-Trimethylhexane	0.7292	0.7325	−0.0033
3-Ethylheptane	0.7270	0.7219	0.0051
3,3-Dimethylheptane	0.7250	0.7271	−0.0021
Decane	0.7301	0.7203	0.0098
2,2,6-Trimethylheptane	0.7195	0.7328	−0.0133
2,2,5-Trimethylhexane	0.7072	0.7250	−0.0178
2,2,4-Trimethylhexane	0.7156	0.7293	−0.0137
2,3,4-Trimethylhexane	0.7238	0.7320	−0.0082
2,3,5-Trimethylhexane	0.7219	0.7273	−0.0054
2,2,3,4-Tetramethylpentane	0.7390	0.7383	0.0007
2-Methyl-4-ethylhexane	0.7230	0.7262	−0.0032
2,2-Dimethyl-3-ethylpentane	0.7348	0.7370	−0.0022
4,4-Dimethylheptane	0.7250	0.7296	−0.0046
2,6-Dimethylheptane	0.7089	0.7172	−0.0083
2,5-Dimethylheptane	0.7150	0.7205	−0.0055
3,4-Dimethylheptane	0.7230	0.7269	−0.0039
2,4-Dimethyl-3-ethylpentane	0.7379	0.7343	0.0036
2,2,5-Trimethylhexane	0.7188	0.7250	−0.0062
2,3,3-Trimethylhexane	0.7380	0.7352	0.0028
2-Methyl-3-ethylhexane	0.7310	0.7288	0.0022
2,3,4-Trimethylhexane	0.7392	0.7320	0.0072
2,2,3,3-Tetramethylpentane	0.7567	0.7447	0.0120
3-Methyl-4-ethylhexane	0.7420	0.7307	0.0113
3,3,4-Trimethylhexane	0.7454	0.7374	0.0080
4-Ethylheptane	0.7300	0.7239	0.0061
2,3,3,4-Tetramethylpentane	0.7547	0.7417	0.0130
2,3-Dimethyl-3-ethylpentane	0.7540	0.7401	0.0139
4-Methyloctane	0.7199	0.7189	0.0010
2,4,6-Trimethylheptane	0.7225	0.7362	−0.0137
2,2-Dimethyl-4-ethylhexane	0.7330	0.7433	−0.0103
3,4-Dimethylheptane	0.7314	0.7269	0.0045
3-Methyl-3-ethylhexane	0.7410	0.7337	0.0073
2,2,5-Trimethylheptane	0.7260	0.7365	−0.0105
2,2-Dimethyloctane	0.7245	0.7293	−0.0048
2,4-Dimethyloctane	0.7259	0.7323	−0.0064
2,7-Dimethyloctane	0.7242	0.7259	−0.0017
4-Isopropylheptane	0.7410	0.7408	0.0002
4-Propylheptane	0.7364	0.7338	0.0026
5-Methylnonane	0.7326	0.7288	0.0038
Undecane	0.7402	0.7272	0.0130
Dodecane	0.7487	0.7326	0.0161

[a] Values of the densities were taken from Ref. 5.

reference *n*-octane is used. Graphs which are simple paths will exhibit the highest value of the *DI* for a given number of vertices. Then as the branching moves closer to the graph center the value of the *DI* decreases until one reaches the center. Multiple branching takes precedent over single branching in lowering the value of the *DI*.

In comparing multiple branch points, one must consider nearness to the graph center and nearness of the branch points to each other. As an example, compare 2,4-dimethylhexane with 2,3,-dimethylhexane. The former has a *DI* equal to 18.3964, and the latter has a value of 18.1815. 2,3-Dimethylhexane has one branch point at a central vertex

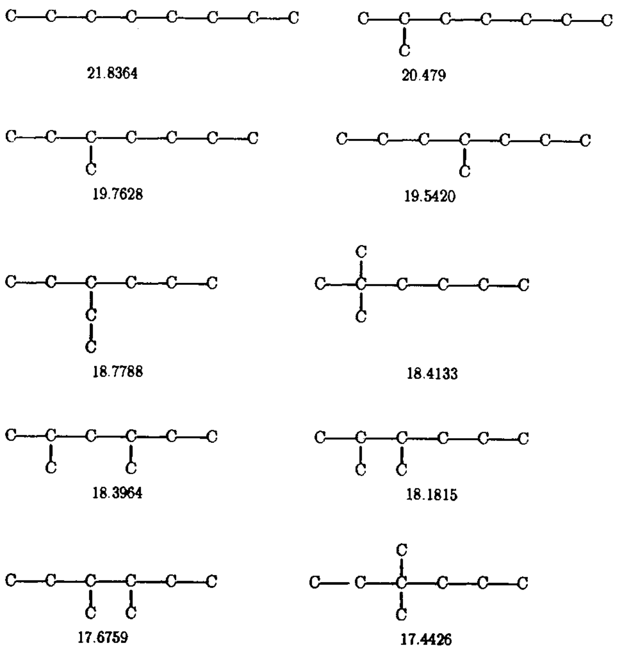

Fig. 1. Some of the eight-carbon hydrogen-suppressed graphs and their corresponding *DI* values.

Fig. 2. The smallest pair of isospectral trees.

and the other is adjacent to it, whereas 2,4-dimethylhexane has one branch point at a central vertex, but the second branch point is not adjacent to it. Thus, 2,3-dimethylhexane has the lower value for its *DI*. Therefore, the *DI* is quantifying both the amount of branching in the structure and the proximity of the branches in multiply branched species. Clearly, the *DI* is encoding information on where structural branches occur and how close together the branch points are in the structure. In addition, it is optimal to have a unique index. It was a lack of uniqueness that prompted us to consider the largest eigenvalue of the distance matrix rather than the largest eigenvalue of the adjacency matrix. It is well known that the two eight-vertex graphs (14) shown in Fig. 2 have the same set of eigenvalues (they are cospectral graphs). This pair of graphs is the smallest of cospectral trees. A tree is an acyclic connected graph. Thus from this definition the acyclic alkanes are tree graphs. With regard to eigenvalues of the distance matrix of trees, this degeneracy in eigenvalues does not occur until one reaches 17 vertices (15). Consequently, the *DI* is better suited as a topological index to characterize molecular branching.

Densities of the alkanes have also been correlated with the Kier and Hall connectivity indices (5). In their work the best correlations were obtained using ${}^3\chi$ and $1/{}^1\chi$. These indices are also reflective of molecular branching. The *DI* is highly correlated with ${}^1\chi$.

CONCLUSION

The *DI* appears to be an excellent topological index to express molecular branching of the alkanes. In combination with Balaban's *J* index it produced an excellent correlation with the density of the alkanes. In addition, the *DI* is easily extended to multiply bonded systems and to heteroatomic systems by making use of an edge weighted and vertex weighted graph to compute the weighted distance matrix, from which one obtains the needed eigenvalues. Finally, the *DI* should produce good correlations with physical properties of compounds that are dependent on molecular branching.

REFERENCES

1. M. Randic. *J. Am. Chem. Soc.* 97:6609–6615 (1975).
2. L. B. Kier, L. H. Hall, W. J. Murray, and M. Randic. *J. Pharm. Sci.* 64:1971–1974 (1975).
3. L. H. Hall, L. B. Kier, and W. J. Murray. *J. Pharm. Sci.* 64:1974–1977 (1975).
4. W. J. Murray, L. H. Hall, and L. B. Kier. *J. Pharm. Sci.* 64:1978–1981 (1975).
5. L. B. Kier, W. J. Murray, M. Randic, and L. H. Hall. *J. Pharm. Sci.* 65:1226–1230 (1976).
6. T. Di Paolo, L. B. Kier, and L. H. Hall. *Mol. Pharmacol.* 13:31–37 (1976).
7. L. H. Hall and L. B. Kier. *J. Pharm. Sci.* 66:642–644 (1977).
8. L. B. Kier and L. H. Hall. *Eur. J. Med. Chem.* 12:307–312 (1977).
9. H. Wiener. *J. Am. Chem. Soc.* 69:17–20 (1947).
10. A. T. Balaban. *Chem. Phys. Lett.* 89:399–404 (1982).
11. D. H. Rouvray. In R. B. King (ed.), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983, pp. 156–177.
12. G. Chartrand and L. Lesniak. *Graphs and Digraphs*, Wadsworth & Brooks/Cole Advanced Books & Software, Calif., 1986, pp. 31–34.
13. SAS Institute Inc. *SAS/IMLTM User's Guide, Version 5 Edition*, SAS Institute, Cary, N.C., 1985.
14. G. A. Baker, Jr. *J. Math. Phys.* 7:2238–2242 (1966).
15. D. B. McKay. *Ars. Combinator.* 3:219–232 (1977).