Structural Selectivity of Topological Indexes in Alkane Series

M. RAZINGER*

Boris Kidrič Institute of Chemistry, YU-61115 Ljubljana, Yugoslavia

J. R. CHRÉTIEN and J. E. DUBOIS

Institut de Topologie et de Dynamique des Systèmes, Université Paris VII, associé au C.N.R.S., 75005 Paris, France

A study of the structural selectivity of six topological indexes is presented. The structural population on which the index values are calculated is the set of all alkane isomers up to dodecane and all possible monocyclic and bicyclic saturated hydrocarbons with four to eight carbon atoms (663 and 376 structures, respectively). The selectivities of indexes for the given population are quantified, compared, and discussed. The connection between selectivity and structure–property correlation performance is discussed also.

INTRODUCTION

The steady growth of the use of computers in the manipulation of chemical structural information during the past years caused the revival of interest of chemists in graph theory. The need for the alphanumerical representation of chemical structures that makes possible their computer handling resulted in the development of various fragmentation codes and linear notations.[1] Although the early structural codes were developed rather pragmatically for documentation and structure–property correlation purposes, subsequent and more evolved alphanumerical representations of chemical structure were firmly based on graph theory.[2,3]

In this paper we shall limit ourselves to the study and comparison of a subgroup of all topological descriptors, namely, the topological indexes in which the topological information is most compressed. Though limited by their condensed form and the resulting loss of information, there is no doubt that these indexes can be very useful for many purposes.[4,5] The task of defining an index that could have different, unique but structurally significant values for different structures seems to be very difficult or even impossible to accomplish.[6] Nevertheless, high structural selectivity of an index is a very sought-for property, being a prerequisite for its successful use in structure–property correlation studies. Of course, high selectivity by itself is not sufficient to obtain good results in such studies if the values of indexes do not reflect structural information properly. A good example of such a case are CAS Registry Numbers, which are unique representations but have no structural meaning whatever and, hence, can be used, according to their limitation, only for documentation and not for correlation purposes.

Several review papers on topological indexes were published in past years, and some of them deal also with selectivity.[4–8] Because we developed a computer program for the automatic simultaneous generation of several topological descriptors that is described in detail elsewhere,[9] we were in the position to make a comparative study of the selectivity of various indexes on a rather large structural data base of alkanes and cycloalkanes. We have compared the following indexes, named in this paper after their first authors for the sake of clarity (the complete list of authors being given under References and Notes): Wiener's number W,[10] Gutman's M_1,[11] Hosoya's Z,[12] Randić's connectivity index $^1\chi$,[13] Bonchev's information on distance I_D,[14] and Balaban's averaged distance sum connectivity J.[15] Although none of the above indexes can distinguish by its value all the structures in the studied population, the number of the occurrences of the same value for different structures differs widely between indexes.

The generated data base of topological indexes was further evaluated by a program that sorts the indexes according to their increasing value, counts the number of identical values of each index for different structures, and displays them on the screen of a graphic videoterminal. The output data of this program were used to define and calculate an objective measure of structural selectivity of indexes, which is presented in the form of tables and diagrams.

GENERATION OF TOPOLOGICAL INDEXES

The method of calculation of various topological descriptors is based on the adjacency matrix A of the graph, representing a given chemical structures, and higher powers of A, i.e., A^2, A^3, ..., A^N (N being the number of points in the graph). Since the possibilities and limitations of the program TOPIND developed for the automatic generation of topological descriptors are described in detail elsewhere,[9] we shall present here only an example of generation of topological descriptors of a hexane isomer, shown in the form of an interactive dialogue in Figure 1.

After arbitrarily labeling the points in the graph representing the structure whose descriptors he wants to calculate, the user enters the upper off-diagonal elements of the adjacency matrix, but only until the last non-zero element in each row (because the matrix is initialized as a $N \times N$ dimensional zero matrix beforehand). The program enables the user to store the matrix or read the previously entered data. This option makes possible the use of some large structural data bases as the source of input data via the conversion of its structural code (DEL in DARC, connection table in CAS, WLN in NIH-EPA information system) to adjacency matrix. The chemical name of the structure in question can be entered either as a commonly used name, or as an abbreviated modification of the IUPAC name for later identification of indexes. The output data of the program are topological codes and partitions (i.e., sequences of numbers) and topological indexes (i.e., simple integer or real numbers) representing the input structures. The programs are written in interactive DEC-FORTRAN IV language and are implemented on VAX 780 and PDP 11/34 installations, running under VMS and RSX-11/M operating systems, respectively.

RESULTS AND DISCUSSION

The structural selectivity of indexes can be evaluated and compared visually by using the graphic output option of the sorting program. The index values for a given set of structures are displayed as a sort of histogram (Figure 2). The distribution profile (i.e., the spacing and height of lines in the

Table I. Number of Distinct Values[a] and Corresponding Selectivity Coefficients of Six Topological Indexes for 663 Alkane Isomers

	alkane [N(str)]															
	pentane (3)		hexane (5)		heptane (9)		octane (18)		nonane (35)		decane (75)		undecane (159)		dodecane (355)	
index	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)
Gutman	3	1.0	4	0.80	5	0.56	6	0.33	7	0.20	8	0.11	9	0.06	10	0.03
Wiener	3	1.0	5	1.0	7	0.78	16	0.89	16	0.46	40	0.53	37	0.23	87	0.25
Hosoya	3	1.0	5	1.0	9	1.0	14	0.78	23	0.66	40	0.53	69	0.43	118	0.33
Randić	3	1.0	5	1.0	9	1.0	16	0.89	28	0.80	49	0.65	83	0.52	137	0.39
Bonchev	3	1.0	5	1.0	9	1.0	18	1.0	33	0.94	71	0.95	138	0.87	304	0.86
Balaban	3	1.0	5	1.0	9	1.0	18	1.0	35	1.0	74	0.99	159	1.0	347	0.98

[a] Real number type indexes were calculated to four decimal places.

```
>RUN TOPIND
Do you want to store or to read the input data? [Y/N]
N
Enter the number of graph points and the highest
wanted power of the adjacency matrix!
6,6
Name the structure (up to 40 characters)!
2M C5
Enter the upper off-diagonal elements
of the matrix row after row!
1
1,1
0
1
1
        0 1 0 0 0 0
        1 0 1 1 0 0
        0 1 0 0 0 0
        0 1 0 0 1 0
        0 0 0 1 0 1
        0 0 0 0 1 0
Is the matrix O.K.? [Y/N]
Y
Select one of three options of output file:
    1. All the powers of A matrix, codes and indices
    2. A and D matrices, codes, indices
    3. Topological codes and indices only
Enter the number of option!
2
Do you want to input another structure? [Y/N]
N
117  —  STOP Results:INDEX.NUM,KODA.NUM
>
```

Figure 1. Interactive dialogue showing the use of the program TOPIND for the calculation of topological descriptors of 2-methylpentane. All entries, made by the user, are underlined.

histogram) gives a clear picture of the selectivity of the corresponding index. The height of each line is equal to the number of structures having the same index value, represented by the position of the line. Thus, the graphic option of the program makes possible a quick evaluation of the selectivity of a given index for the structural population under consideration, without a detailed scrutiny of the numerical results of the sorting program. The values of six indexes for all 35 nonane isomers are shown in this way in Figure 2.

To have a quantitative measure of structural selectivity of topological indexes, we defined the selectivity coefficient C(s) of an index for a given set of structures:

$$C(s) = N(\text{val})/N(\text{str})$$

The selectivity coefficient is equal to the number of distinct values N(val), assigned by an index to a family of structures, divided by the number of structures N(str) in this family. The coefficient C(s) is closely related to the "mean isomer degeneracy i" of Bonchev et al.[6] In the study of the selectivity of indexes in isomeric families, the value of C(s) is exactly the inverse of i, but it is better suited for the diagrammatic representation of selectivity since the value of C(s) varies between 1 and $1/N$(str) and that of i between 1 and N(str).

Tables I and II summarize the data on the selectivity of six indexes for all alkane isomers up to dodecane and all mono and bicyclic saturated hydrocarbons with four to eight carbon atoms, respectively. The number of distinct values and the corresponding selectivity coefficients are given for each family
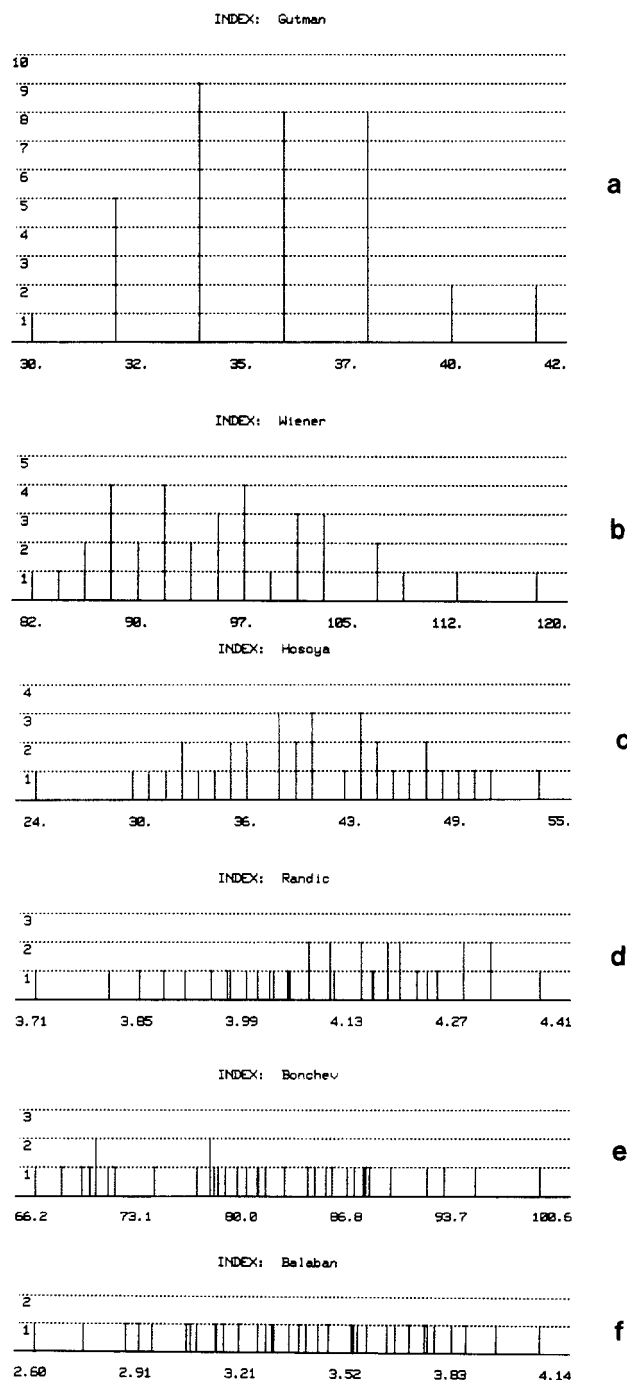
Figure 2. Example of the graphic output option of the sorting program. On the screen of a graphic videoterminal are shown the distribution profiles of the values of six indexes for 35 nonane isomers: (a) Gutman, (b) Wiener, (c) Hosoya, (d) Randić, (e) Bonchev, and (f) Balaban.

of structures and each of the indexes. The dependence of the selectivity coefficient C(s) of six indexes on the number of

STRUCTURAL SELECTIVITY OF TOPOLOGICAL INDEXES

J. Chem. Inf. Comput. Sci., Vol. 25, No. 1, 1985 **25**

Table II. Number of Distinct Values[a] and Corresponding Selectivity Coefficients of Three Most Selective Indexes for 376 Mono- and Bicyclic Structures

	no. of carbon atoms [N(str)]									
	4 (3)		5 (10)		6 (29)		7 (84)		8 (250)	
index	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)	N(val)	C(s)
Bonchev	2	0.67	5	0.50	12	0.41	30	0.36	70	0.28
Randić	3	1.0	10	1.0	24	0.83	60	0.71	144	0.58
Balaban	3	1.0	10	1.0	29	1.0	84	1.0	248	0.99

[a] Indexes were calculated to eight decimal places.
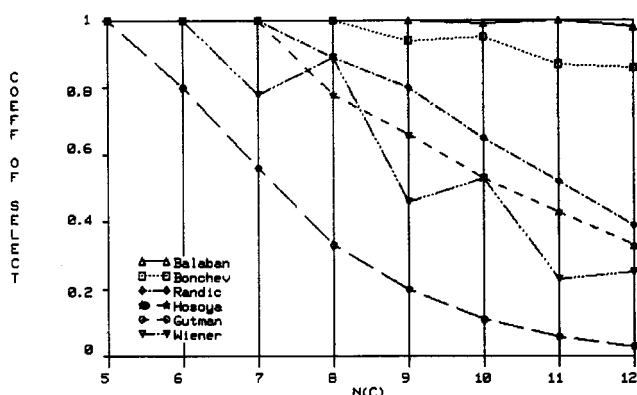
Figure 3. Selectivity coefficient of six topological indexes for alkanes.
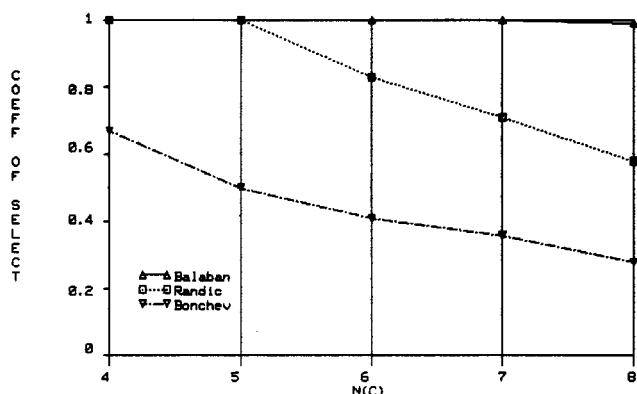
Figure 4. Selectivity coefficient of three real number type indexes for cyclic structures.

Table III. Dependence of Structural Selectivity of Three Real Number Type Indexes on Precision of Calculation (Number of Distinct Values for Higher Alkanes)

alkane	N(str)	no. of decimal places	Randić	Bonchev	Balaban
octane	18	4	16	18	18
		8	18	18	18
nonane	35	4	28	33	35
		8	31	34	35
decane	75	4	49	71	74
		8	55	71	75
undecane	159	4	83	138	159
		8	92	140	159
dodecane	355	4	137	304	347
		8	179	314	349
		17	179	314	349

carbon atoms in structures is presented also diagrammatically in Figures 3 and 4.

The most apparent feature in the diagram is the decrease of selectivity of indexes with increasing number of carbon atoms in alkanes. This effect is known and can be accounted for by the fact that the structural differences between two "most similar" isomers become less pronounced with increasing number of isomers in alkanes. The structural dissimilarity, for example, between isopentane and neopentane is far greater than between 2,2,3,4,5-pentamethylheptane and 2,2,3,3,5-pentamethylheptane. Only those indexes based upon the distance distribution in structures show an exception to the rule of monotonous decrease of selectivity with increasing number of carbon atoms. The evident saw-tooth profiles of Wiener's and to a lesser extent of Bonchev's curves are the consequence of different distance distributions in the isomers with odd and even number of carbon atoms. The maximal possible distance in odd close-packed structures is lower than that in the even structures for the same number of neighbor spheres around the center;[6] this is reflected in the lowered selectivity of distance indexes for odd alkanes.

From the data in Tables I and II and curves in Figures 3 and 4, it is obvious that there is a great difference in the selectivity of various indexes. The least selective index of the six studied (Gutman's) gives only $N - 2$ different values for

all isomers of an alkane with N carbon atoms; this is the reason why its selectivity decreases quickly with growing N, C(s) falling below 0.10 for undecanes and dodecanes. From the point of view of selectivity, the best index defined until now, at least for the tree structures, is Balaban's. It discriminates all isomers up to nonane and has C(s) still above 0.98 even for dodecane. The selectivity of the remaining four indexes is situated between these two extremes.

In the case of real number type indexes (Randić's, Bonchev's, and Balaban's), the selectivity depends upon the precision of their calculation, i.e., on the number of decimal places taken into account during the comparison of the index values. The data in Table III describe this dependence for the three real number type indexes on the structural population of higher alkanes. It shows clearly that the selectivity of indexes calculated to eight decimal places is superior to the selectivity of the same indexes calculated only to four decimal places (this is usually the precision of the data reported in the literature). For the dodecane isomers where the difference in selectivity is most pronounced, we calculated the values of indexes with 17 decimal places (this being the maximal possible precision of a 16-bit computer with double precision variables). The number of distinct index values N(val) was the same for 8- and 17-place precision. We can conclude that the usually reported data on index values with the precision of only four decimal places may give flase or insufficient results in the determination of structural selectivity.

In structure–property correlation studies, one is tempted to conclude that the most selective indexes also will show the best correlation. This is unfortunately not always the case, as can be seen from the data in Table IV.

Since in this paper our attention is focused on the selectivity of indexes, we shall present only briefly a comparison of the performances of the six indexes in the correlation between various properties and structure of alkane isomers (for more data, see, for example, references 5 and 17). Accordingly, only the data on correlation between physicochemical properties and index values of some selected populations of structures will be presented. The data in Table IV suggest that for some properties and populations there exists a fairly good agreement between the coefficient of selectivity C(s) and the coefficient of correlation of linear regression R (populations 1, 4, and 5),

Table IV. Data on Selectivity and Structure–Property Correlation Coefficients for Six Topological Indexes of Some Selected Alkane Isomer Populations[a]

		index					
		Gutman	Wiener	Hosoya	Randić	Bonchev	Balaban
population 1, C2–C7, N(str) = 21	N(val)	13	19	19	21	21	21
	C(s)	0.62	0.90	0.90	1.0	1.0	1.0
	$R(T_b)$	0.90	0.94	0.91	0.99	0.94	0.79
population 2, C7, N(str) = 9	N(val)	5	7	9	9	9	9
	C(s)	0.56	0.78	1.0	1.0	1.0	1.0
	$R(T_b)$	0.86	0.73	0.96	0.92	0.71	0.71
population 3, C8, N(str) = 18	N(val)	6	16	14	16	18	18
	C(s)	0.33	0.89	0.78	0.89	1.0	1.0
	$R(T_b)$	0.72	0.54	0.89	0.82	0.54	0.48
population 4, C2–C6, N(str) = 12	N(val)	10	12	11	12	12	12
	C(s)	0.83	1.0	0.91	1.0	1.0	1.0
	R(CSA)	0.91	0.95	0.92	0.99	0.93	0.85
population 5, C5–C7, N(str) = 17	N(val)	9	15	16	17	17	17
	C(s)	0.53	0.88	0.94	1.0	1.0	1.0
	$R(d_4^{20})$	0.71	0.89	0.88	0.93	0.87	0.45

[a] Data sources: T_b (boiling point temperature);[5,10] CSA (cavity surface area);[16] d_4^{20} (density of liquid phase).[17]

while in other cases a less selective index can correlate better with some property than a more selective one (populations 2 and 3).

The exceptions to the rule of agreement between C(s) and R are Gutman's index on one side and Bonchev's and Balaban's indexes on the other side. In the cases of populations 2 (all heptane isomers) and 3 (all octane isomers), there occurs a curious inversion when the least selective index correlates better with boiling point temperature than the most selective ones. In these two populations also, Hosoya's index has a higher R than Randić's, the latter being superior in other populations and also generally known to give excellent correlation with various properties.[5] Populations 2 and 3 are the only ones amongst the five presented in Table IV in which the number of carbon atoms is equal for all members. This fact leads us to the conclusion that some indexes, even if of feeble selectivity, are better suited for correlation with properties within isomer series, while other indexes are better for the correlation with structural sets in which also the size of structures (not only the branching pattern) is changing. Another conclusion that can be drawn from the values of R in Table IV is the fact that indexes, based upon distance distribution like Wiener's, Bonchev's, and Balaban's, are inferior to other indexes in correlations with properties within isomeric families (populations 2 and 3).

One of the causes of all these effects is the basic choice of different graph invariants used for the description of structures by different indexes (distance, degree, etc.). Taking into account only one graph invariant inevitably emphasizes the effect of the corresponding structural feature on the index value. Because of this, the quest for an index that could be used equally well for correlations with a host of different properties, depending on different structural features, looks very unpromising. To overcome this problem, there are two possible but basically different solutions.

The first possibility is to select several graph invariants for the definition of a new descriptor. In this way one can hope to get a numerical value that will reflect various properties of a given structure equally well. The work in this direction was published recently;[6] the result of the research was not an index but a "superindex", i.e., a sequence of numbers rather than a simple number. It shows excellent selectivity, but it is not in the condensed form of a topological index, nor was it tested in structure–property correlations.

The second possibility to get a better index–property correlation is to choose a graph invariant, to define an algorithm for its transformation into a numerical descriptor, and to see

which properties correlate best with the thus-defined index. There are several publications reporting on successful work in this direction.[5,17] The drawback of this approach lies in the fact that the deliberate limitation of the structural description to a selected invariant must inevitably result in the limited scope of structure–property correlations with such an index.

The choice of alkanes for the structural population on which the comparisons were made was not arbitrary: because of the rapidly increasing number of isomers in alkane series, the structural differences between isomers with the same number of carbon atoms become more and more subtle. This is the reason why the discrimination of structures by index value becomes a demanding task and why, at least for the test of selectivity, the choice of structural set was adequate.

The study of structure–property correlation calls for different structural populations. Such correlations were calculated most often just on alkanes, but this was done more to test new descriptors than to get results of some practical value. Additionally, the data on various properties of alkanes become scarce with the growing number of isomers so that there is no sense in pushing correlations on this population too far.

CONCLUSION

In this paper we exploited the possibilities of the program for the automatic generation of topological descriptors[9] to generate the values of six topological indexes for a large population of isomers of aliphatic and cyclic alkanes. The selectivity of all six indexes was studied, quantified, and compared. The influence of selectivity on performance in structure–property correlation studies was observed and interpreted. A similar study of topological codes (sequences of numbers representing chemical structures) is reported elsewhere.[18] Here, the large effect of the method of compression of topological information upon selectivity can be mentioned briefly because it affects often the selectivity of indexes. Three topological descriptors (Wiener's and Bonchev's indexes and Altenburg's polynomial) are all based directly on the distance matrix of the graph representing the structure, but their selectivity is very different. Bonchev's method of information compression is clearly superior to Wiener's, resulting in a far more selective index; but the Altenburg polynomial, in which the compression is minimal (i.e., only the summation of equal distances), is more selective than any of the two. There is no doubt that topological codes could be exploited better for structure–property correlations also. Up to now, the only attempt in this direction is the work of Randić and Wilkins[19,20]

on molecular code (i.e., sequence of the number of graph paths of different length).

Finally, an index that would be highly selective and could at the same time be successful in correlating many different properties is not yet discovered. Because of this, one must choose for correlation studies an index according to the property and, perhaps, use also other statistical methods besides the standard linear regression.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E. "Computer Handling of Chemical Structure Information"; Macdonald: London; American Elsevier: New York, 1971; p 12.
(2) Morgan, H. L. "Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at CAS". *J. Chem. Doc.* **1965**, *5*, 107–113.
(3) Dubois, J. E. In "Chemical Applications of Graph Theory"; Balaban, A. T., Ed.; Academic Press: New York, 1976; Chapter 11, p 333.
(4) Balaban, A. T.; Chiriac, A.; Motoc, I.; Simon, Z. "Steric Fit in QSAR". *Lect. Notes Chem.* **1980**, *15*, 22–39.
(5) Kier, L. B.; Hall, L. H. "Molecular Connectivity in Chemistry and Drug Research"; Academic Press: New York, 1976.
(6) Bonchev, D.; Mekenyan, Ov.; Trinajstić, N. "Isomer Discrimination by Topological Information Approach". *J. Comput. Chem.* **1981**, *2*, 127–148.
(7) Bonchev, D.; Knop, J. V.; Trinajstić, N. "Mathematical Models of Branching". *Math. Chem.* **1979**, *No. 7*, 21–42.
(8) Balaban, A. T. "Chemical Graphs. XXXIV. Five New Topological Indices for the Branching of Tree-Like Graphs". *Theor. Chim. Acta* **1979**, *53*, 355–375.
(9) Razinger, M.; Chrétien, J. R.; Dubois, J. E. "Graphes Chimiques: Génération Automatique des Descripteurs Topologiques". *Vestn. Slov. Kem. Drus.* **1984**, *31*, 211–227.
(10) Wiener, H. "Structural Determination of Paraffin Boiling Points". *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
(11) Gutman, I.; Ruščić, B.; Trinajstić, N.; Wilcox, C. F., Jr. "Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes". *J. Chem. Phys.* **1975**, *62*, 3399–3405.
(12) Hosoya, H. "Graphical Enumeration of the Coefficients of the Secular Polynomials of the Hueckel Molecular Orbitals". *Theor. Chim. Acta* **1972**, *25*, 215–222.
(13) Randić, M. "On Characterization of Molecular Branching". *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
(14) Bonchev, D.; Trinajstić, N. "Information Theory, Distance Matrix, and Molecular Branching". *J. Chem. Phys.* **1977**, *67*, 4517–4533.
(15) Balaban, A. T. "Highly Discriminating Distance-Based Topological Index". *Chem. Phys. Lett.* **1982**, *89*, 399–404.
(16) Hermann, R. B. "Theory of Hydrophobic Bonding. II. The Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area". *J. Phys. Chem.* **1972**, *76*, 2754–2759.
(17) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. "Molecular Connectivity I–III". *J. Pharm. Sci.* **1975**, *64*, 1971–1981; "Molecular Connectivity V". *J. Pharm. Sci.* **1976**, *65*, 1226–1230.
(18) Razinger, M.; Mičović, S. "Structural Selectivity of Topological Codes in Alkane Series". *Vestn. Slov. Kem. Drus.* **1983**, *30*, 199–212.
(19) Randić, M.; Wilkins, C. L. "Graph Theoretical Analysis of Molecular Properties. Isomeric Variations in Nonanes". *Int. J. Quant. Chem.* **1980**, *18*, 1005–1027.
(20) Wilkins, C. L.; Randić, M. "A Graph Theoretical Approach to Structure-Property and Structure-Activity Correlations". *Theor. Chim. Acta* **1980**, *58*, 45–68.

Optimization of a Similarity Metric for Library Searching of Highly Compressed Vapor-Phase Infrared Spectra

MICHAEL F. DELANEY,* JOHN R. HALLOWELL, JR., and F. VINCENT WARREN, JR.[†]

Department of Chemistry, Boston University, Boston, Massachusetts 02215

Compound identification by library searching of experimental spectra using instruments based on small computer systems is becoming increasingly common. For successful searching performance, large libraries are needed. Compressed spectra are typically used to increase both the number of spectra that can be stored on a small computer and the search speed. In this study, a similarity metric for matching highly compressed binary intensity vapor-phase infrared spectra is optimized with three distinct approaches. Two of the approaches are generally applicable for library searching performance evaluation. The results of all three approaches are in excellent mutual agreement.

INTRODUCTION

Library searching (LS) remains the method of choice for rapid, on-line computerized compound identification by chemical spectra.[1] Libraries of various types of spectra have been employed, most notably mass,[2] infrared,[3] and nuclear magnetic resonance[4] spectra. In many cases, the only way to accomodate the large numbers of reference compounds needed for a useful library is to greatly compress the library spectra. The size of the library, and its effect on searching speed, can be especially critical when microcomputers are used for LS.[5]

Among highly compressed spectral representations for LS, the most popular has been the use of binary intensity spectra.[6] In this situation, a peak in a spectrum is encoded as a "1", and the absence of a peak as a "0". This approach has been used to advantage in mass spectrometry (MS). For vapor-phase infrared (VPIR) spectrometry, we demonstrated[7] how the

information content of the library and the LS performance could be improved by incorporating a measured amount of peak-width information into a binary spectrum.

The comparison metric by which similarity between an unknown spectrum and the reference library members is quantitated is also an important criterion in designing a spectral search system. Various metrics have been used with binary intensity representations of different types of spectrometry. The metrics studied, which of necessity are based on Boolean functions and combinations thereof, have included AND,[8] XOR (exclusive OR),[6,9] AND/OR,[10] EXNOR (exclusive NOR),[11] and XOR/OR.[7] An additional metric, originally introduced by Grotch[8,12,13] and the subject of this paper, is a composite metric: $XOR - \mu$ AND. In this metric, μ is a variable factor that controls the relative contribution of the XOR and AND functions.

We recently developed a procedure for the quantitative evaluation of library searching performance.[14] In this approach, any type of spectrometry, library representation, or