1 has no information on stereochemistry.

(22) The full list of hexagonal reaction graphs was summarized in ref 2.

(23) The node values are presumed to be 06 (carbon atoms).

(24) The enumeration of hexagonal reaction graphs was reported in ref 2.

(25) For the initial numbering of ITS 11 and the canonical coding, see the

(26) Fujita, S. *Yuki Gosei Kagaku Kyokaishi* **1986**, *44*, 354.

(27) ITS 14 has two reaction strings, i.e., 1–5+10–15–6+2–3+1 and 1–2+6–11+9–4+1. See ref 6.

(28) For the connection table of ITS, see ref 1.

preceding paper.[20]

Ring ID Numbers

MILAN RANDIĆ

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, and Ames Laboratory—DOE,[†] Iowa State University, Ames, Iowa 50011

There are advantages to a single-number representation of a molecule or molecular fragments, even though from the onset it is known that such characterizations necessarily are accompanied by certain loss of information. We consider construction of a single-number characterization of rings in polycyclic structures to serve as ring descriptors. The number, called a ring ID number, is based on the count of suitably weighted paths for atoms forming the ring. The approach is illustrated on various rings in trimethyltricycloheptanes. The proposed ring descriptor shows a high discrimination power. Moreover, ring ID numbers reflect some inherent structural features of rings. We find that rings which are apparently similar are represented by ID values that are numerically close.

INTRODUCTION

Present expansion of chemical data, including registration of new compounds, is associated with continuous needs for data retrieval. Searching modest data files, disregarding problems of fragment search, or searching for maximal common substructure can already be time consuming unless highly efficient searching is possible. Already, just finding compounds of interest, regardless of other possible tasks, may be quite difficult unless the correct name or code for the compound is known—something that frequently will not be the case. One strategy in efficient searching through a large data file is to introduce a reduced structure basis that will result in presorted samples. A natural choice is to select the number of atoms and number of rings as preliminary screening parameters; however, in practice such choices are not sufficiently discriminatory. For example, there are thousands of structures having a five-membered ring, yet we may be interested in a selected few having a particular structural feature, such as being fused to another ring or having specified substituents. It is therefore important to narrow down the search by including additional structural characteristics in a code, yet keeping the representation as simple as possible. There are no difficulties in developing a coding system that is comprehensive and *lengthy*, where, for example, a structure is represented uniquely and richly by a long list of specifications. The problem is to design a scheme in which a considerable amount of structural information is contained in a simple, *short* code. In addition, the approach should not be restricted to a special class of compounds. The shorter the code, the faster will be the search based on such codes. Hence, representations of a structure by a single number are highly desirable because they will considerably speed a searching process when compared to schemes using sequential data (e.g., connectivity tables). The issue is, Can a useful single-number condensation of structure, having sufficient discrimination, be developed?

It is useful to distinguish between the two conceptually distinct avenues for representing structures: (i) codes versus (ii) descriptors. Codes, as the name implies, presume prescribed *rules*, which when followed produce a name for the structure. Descriptors, as the name implies, use a selected property as the attribute on which characterization is based. Codes are not invariants; i.e., they require a particular (canonical) labeling of atoms. Descriptors are invariants, i.e., independent of how atoms are labeled. There are other important differences between codes and descriptors as will be outlined shortly.

Read[1] summarized desirable qualities for chemical codes (or names), which include line representation (based on common symbols), uniqueness, possibility for reconstruction (of a structure from a code), brevity, etc. In the case of trees (acyclic graphs) the *n*-tuple code of Knop and collaborators,[2] where for an *n*-atom tree a string of *n* numbers suffices, is an illustration of codes that satisfy the requirements. The *n*-tuple code can be extended to cyclic structures by adding ring-closure information, producing thus very compact codes.[3] The derived codes, *n*-tuple,[2] compact codes,[3] official IUPAC names,[4] or any of numerous alternative canonical labeling of atoms[5] all require the user to know *rules* that govern code construction, rules which differ in complexity, preferences, and arbitrariness. Typically, the simpler the coding rules, the more difficult it is to derive the code. But this is preferred to rules designed to anticipate and resolve all ambiguities and that thus become lengthy, complex, and cumbersome. While computational difficulties, if necessary, can be delegated to a computer, it is hard to foresee all structural features of unknown novel compounds. Hence, the frequent revisions of "official" nomenclature rules, such as made by IUPAC. An illustration of structural rules is canonical labelings of atoms, such as those based on the smallest binary code for the structures when the entries of the adjacency matrix are read from left to right and from top to bottom.[6]

Alternative to coding based on rules are approaches based on structural invariants, indices, as the basis for a design of molecular descriptors. Invariants are independent of assumed labeling and therefore allow a direct comparison of files in different laboratories. A disadvantage of invariants is that there is no guarantee that the representation for a compound is unique. Moreover, it is generally believed that no finite list of structural invariants suffices to specify a compound uniquely. Hence, there is some loss of structural information when structures are represented by invariants, and consequently such descriptors do not allow reconstruction. On the other hand, since invariants represent "mathematical" properties of structures, they have advantages in structure–property studies.

The use of structural invariants bypasses the conceptual dilemma in structure–property correlations concerning the disparity in comparison: Properties can be represented as numbers, and structures, which are clearly more complex mathematical objects, cannot be adequately represented as numbers. The important questions then are the following: How well can a structure be represented by a number? How useful are such representations in problems of chemical documentation? How useful are such descriptions in the study of structure–property correlations?

TOPOLOGICAL DESCRIPTORS

It has been surprising to find out how well apparently simplistic representations of chemical structures, in which a molecule is condensed to a single number, work out in both structure–property and structure–documentation areas. The first deliberate single-number representation of structures was considered in 1971 by Hosoya.[7] Hosoya introduced an index, called the topological index Z, based on counting all disjoint combinations of bonds in a structure. The name has been since generally used for other graph-based invariants used as molecular descriptors. Such numbers clearly represent a purely combinatorial and topological quantity, since they can be determined from the molecular graph, and more correctly should have been referred to as graph theoretical indices. The index Z not only serves as a sorting device for coding chemical structures but also is of interest in studies of isomeric variations of various properties of alkanes. Another quantity similarly used, although not initially intended as a single-number representation of a structure, is the Wiener number W. This quantity, initially introduced for discussion of structural variations in alkanes and other organic compounds, is one of two structural parameters found to represent various bond additive molecular properties exceedingly well.[8] The most successful single-number molecular representation for discussion of bond additivities, apparently, is the connectivity index χ, introduced by this author.[9] The success of this index,[10] amazing as it appeared, is not accidental. The index was designed to *parallel* molecular properties. The algorithm used in its construction assigns to bond types (m,n) the weights $1/(mn)^{1/2}$, which represent a solution of a system of inequalities induced by prior *ordering* of the structures (isomers). The ordering chosen parallels relative magnitudes for numerous physicochemical properties of paraffins. Thus, the index will necessarily reproduce the correct relative magnitudes (ordering of isomers), which in turn ensures a high correlation coefficient.[11]

The connectivity index has not been considered as a device for sorting or labeling structures because of considerable degeneracy that accompanies it. In Figure 1 we show structures having the same Z, W, and χ, which well illustrates limitations of these particular single-number descriptors for the purpose of chemical documentation. Because the above invariants are primarily used in structure–property correlations, the occurrence of structures having duplicate descriptors is not necessarily troublesome. Similar molecules may show similar properties! In fact, the expectation that similar compounds have similar properties follows from empirical observations that can be traced to the pioneering work on structure–activity as reflected in Emil Fischer's "lock and key model"[12] for interaction of drugs and enzymes (receptors). Recently this experience has been formulated as one of the fundamental postulates in structure–activity relationships (SAR).[13]

MOLECULAR ID NUMBERS

The question can be raised, Are there alternative single-number representations of structures that have a high discrimination power and that could, therefore, be of interest in
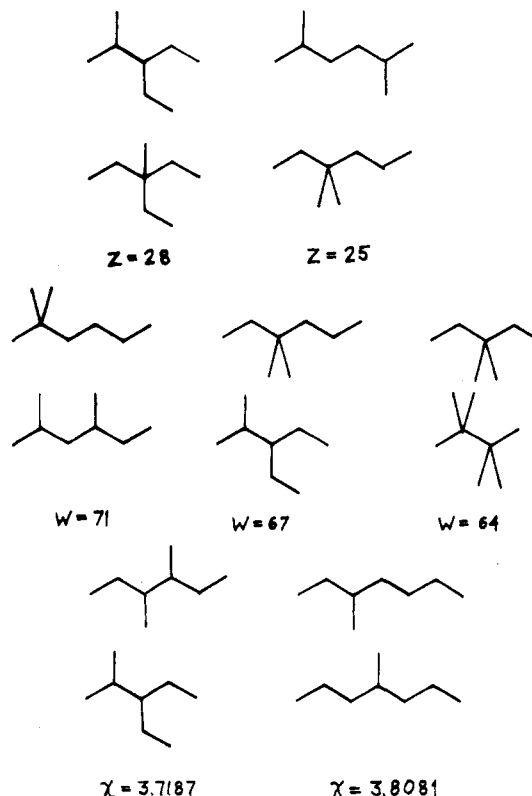

Figure 1. Smaller alkane skeletons showing degeneracy for the Hosoya topological index (Z), the Wiener number (W), and the connectivity index (χ).

chemical documentation? Balaban[14] was first to suggest one such construction, an index J, which is analogous to the connectivity χ index but instead of (m,n) bond types considers (i,j) graph distances. Because the distance matrix in comparison to the adjacency matrix has greater entry variations, one expects that an index based on distance matrix entries will be considerably less prone to accidental coincidences. Again duplicates occur but, as anticipated, fewer than in the case of the previously mentioned invariants. The smallest duplicates (i.e., a pair of structures having the same J index) have a dozen vertices, which, although not quite satisfactory, nevertheless points to capabilities of single-number representation as viable structure descriptors.

The first highly discriminatory single-number invariant is the so-called molecular ID number.[15] The number emerged as a byproduct of use of weighted paths to describe molecular structure.[16] The molecular ID number represents the sum of weighted paths in a structure and thus corresponds to the Wiener number, the sum of all paths in a structure, in unweighted structure. Because bonds are weighted, the "count" of paths is modified, and instead of integers, we have real numbers as the output. The approach allows some flexibility in the selection of bond weights. In past applications the selected weights were the same as that used in the definition of the connectivity index; i.e., each bond has weight $1/(mn)^{1/2}$, where m and n are the valencies of the end atoms of each bond. In the case of trees, the smallest duplicate structure with the same molecular ID number appears for trees with $n = 15$ vertices (Figure 2), i.e., one alkane pair among 4347 cases. The next smallest case is one pair when $n = 16$, one pair among 10 359 cases. Even in these cases of duplicate ID values, the structures are visibly different and can be recognized as nonisomorphic by examining other molecular parameters.[17]

Uniqueness of ID numbers was neither claimed nor expected. The ID numbers found use in discussions of structure–property relationships[18] and have shown promise in structural pattern recognition studies.[19] Their possible use in
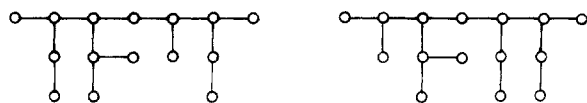
Figure 2. Smallest pair of nonisomorphic trees (having $n = 15$ vertices) having identical ID numbers.

Table I. ALL PATH Output of Weighted Path Numbers for 2-Methylpentane Carbon Skeleton[a]

	P(0)	P(1)	P(2)	P(3)	P(4)	ID
atom						
1	1.0000	0.5773	0.5690	0.1178	0.0833	2.3475
2	0.5773	1.5629	0.2041	0.1443		2.4887
3	0.7071	0.9082	0.8249			2.4403
4	0.7071	1.2071	0.2041	0.2357		2.3540
5	1.0000	0.7071	0.3535	0.1443	0.1666	2.3716
6	1.0000	0.5773	0.5690	0.1178	0.0833	2.3475
molecule	4.9915	2.7700	1.3620	0.3800	0.1666	9.6707

[a] The entries correspond to counts of paths of increasing length; the last entry for each atom is the atomic sum or atomic ID value. The last row corresponds to the molecule as a whole, the total number of (weighted) paths being the molecular ID number.

chemical documentation has yet to be explored. The finding that molecular ID numbers possess a large discriminatory power reopens the question, How discriminatory can a single-number representation of a molecule be? The name ID suggested itself after observed high discriminatory power of weighted path sums. As long as duplications are seldom, and thus do not cause frequent ambiguities in practical situations, the lack of uniqueness of "identification" numbers is not troublesome. The derived ID numbers, even though displaying unsurpassed discriminatory power, can be further improved to give the maximal possible diversity among ID numbers if one reconsiders the bond weight algorithm. One can redefine bond weighting factors to eliminate accidental coincidences that occur when various $1/(mn)^{1/2}$ factors are combined. It has been shown that when products of successive prime numbers are used in the weighting,[20] one eliminates numerous duplicate cases previously reported. The "prime number" weights lead to an astounding discrimination:[21] Among alkane graphs the first duplicate occurs when $n = 20$. This means one pair of structures among 366 319 isomers! Evaluation of paths (or weighted paths) in polycyclic compounds is, however, slow when the number of rings in a structure is increased. Knop, Trinajstić, and collaborators[22] considered, therefore, as an alternative the use of weighted walks, rather than paths. This allowed evaluation of modified ID numbers to be extended to very large polycyclic systems (e.g., benzenoid hydrocarbons). The single-number representations so constructed again displayed an outstanding discrimination and have the advantage of fast computations, although for some small cyclic systems duplicate ID values arose. We can summarize the situation as follows: Single-number representations of molecules appear promising for practical applications in chemical documentation. These introductory molecular ID numbers based on weighted paths and weighted walks, respectively, may undergo further modifications. It seems quite possible, even probable, that some such construction will satisfy needs that practice may indicate. It is, for example, very likely that ID quantities based on the distance matrix, rather than the adjacency matrix, will have the desired maximal discriminatory power. This one anticipates in analogy to the connectivity index and the J index of Balaban, where a considerable jump in the discrimination power was observed.[23] The subject and such speculations, however, are outside the scope of the present paper.

In Table I, we illustrate computer output of the ALL PATH program,[24] giving weighted path numbers for 2-methylpentane. In this case there are only the following bond types: (1,2), (1,3), (2,2), and (2,3). Therefore, the bond weight factors that
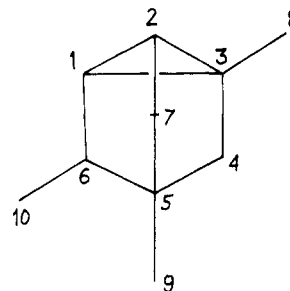

Figure 3. Trimethylcycloheptane and the labeling of carbon atoms assumed.

occur are respectively $1/2^{1/2}$, $1/3^{1/2}$, $1/4^{1/2}$, and $1/6^{1/2}$, i.e., 0.707 10, 0.577 35, 0.408 24, and 0.353 55. As seen from Table I, each atom is represented by a sequence of weighted paths, ordered in increasing path lengths. When all entries of an atomic sequence (i.e., all entries in a row) are added, one obtains an atomic ID value, shown in the last column. When atomic path numbers of different atoms in a molecule are added, keeping separate the results for paths of different length, we obtain the molecular path numbers or molecular path sequnce, shown in the last row of Table I. That is, instead of adding entries in each row, now we add entries in each column. The first entry in the derived sequence represents the sum of "atomic χ", the zero-order molecular connectivity index χ; the second entry similarly gives the sum of "bond χ", the first-order connectivity index χ. The other entries in the derived molecular path sequence are analogous (but not identical) to higher order connectivity indices χ.[25] They differ numerically somewhat, because of a repeating use of intermediate atoms in construction of weighted paths. Finally, the last entry in Table I represents the molecular ID number, which can be obtained by adding all molecular path numbers. Alternatively, they can be derived by using atomic ID numbers as follows: ID = $(1/2)$[atomic ID + atomic χ]. This is because path numbers, except for paths of length zero, are counted twice, once for each end atom of the path.
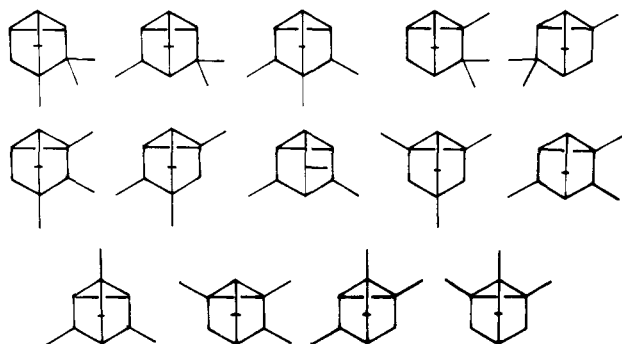
RING ID NUMBERS

We will use the $1/(mn)^{1/2}$ bond weights. Our interests here are molecular fragments and particularly we will focus attention on individual rings in molecules, and we will derive a single-number ring descriptor. As will be seen, when one restricts summation of path contributions only to ring atoms, one obtains a partial path sum that characterizes the selected molecular fragment (ring). It is of interest to investigate if such an approach leads to useful ring characterization. We will introduce the approach by considering an isomer of trimethyltricyclo[2.2.1.0^{2,6}]heptane shown in Figure 3, where arbitrary labeling of atoms is adopted. In Table II we listed the enumeration of weighted paths. In this example we have the following bond types (1,3), (1,4), (2,3), (2,4), (3,3), and (3,4), which call for the weight factors $1/3^{1/2}$, $1/4^{1/2}$, $1/6^{1/2}$, $1/8^{1/2}$, $1/9^{1/2}$, and $1/12^{1/2}$, respectively. As seen in Table II, each atom is represented by a sequence of weighted paths and a single atomic ID number. We would like now to extend the above characterization of atoms to selected molecular fragments. Observe that the above path characteristic pertained to both, to the smallest molecular fragment (the individual atoms) and the molecule as a whole. By confining attention only to ring atoms, we can derive ring path sequences and, as their total, ring ID numbers. In Table II we show the results for the three-membered and five-membered rings of the trimethyltricycloheptane of Figure 3. The results in Table II were obtained by using the modified ALL PATH program.[24] In the program structures are input by their neighbor lists while the bond weights are incorporated automatically; i.e., bond

RING ID NUMBERS

J. Chem. Inf. Comput. Sci., Vol. 28, No. 3, 1988 **145**

Table II. Path Count for Weighted Paths of Increasing Lengths for Individual Atoms of Trimethyltricycloheptane of Figure 3 (Upper Part)[a]

| atom | paths of length | | | | | | | | | ID |
	0	1	2	3	4	5	6	7	8	
1	0.5773	0.9553	0.8507	0.3165	0.1461	0.0662	0.0118			2.9241
2	0.5773	1.0302	0.6814	0.4070	0.1720	0.0551	0.0091	0.0010		2.9334
3	0.5000	1.4309	0.5315	0.3391	0.1736	0.0520	0.0059	0.0011		3.0344
4	0.7071	0.7071	0.7847	0.2877	0.1223	0.0725	0.0218	0.0016		2.7051
5	0.5000	1.4957	0.5322	0.2843	0.1824	0.0521	0.0046			3.0515
6	0.5773	1.1993	0.5557	0.2650	0.1315	0.0609	0.0154	0.0010		2.8065
7	0.7071	0.7618	0.6577	0.3563	0.1676	0.0647	0.0188	0.0024		2.7368
8	1.0000	0.5000	0.4654	0.2657	0.1695	0.0868	0.0260	0.0029	0.0005	2.5172
9	1.0000	0.5000	0.4978	0.2661	0.1421	0.0912	0.0260	0.0023		2.5257
10	1.0000	0.5773	0.3591	0.3208	0.1530	0.0759	0.0351	0.0089	0.0005	2.5310
molecule	7.1462	4.5789	2.9583	1.5544	0.7802	0.3389	0.0875	0.0107	0.0005	17.4561
ring atoms										
1, 2, 3	1.6547	1.7085	1.0318	0.5313	0.2459	0.0867	0.0134	0.0010		5.2733
1, 2, 5, 6, 7	2.8618	2.7212	1.6389	0.8146	0.3999	0.1496	0.0299	0.0022		8.6184
1, 3, 4, 5, 6	2.8618	2.8942	1.6275	0.7463	0.3780	0.1519	0.0298	0.0018		8.6917
2, 3, 4, 5, 7	2.8618	2.7129	1.5938	0.8373	0.4090	0.1483	0.0302	0.0031		8.5966

[a] Molecular path sequences and sequences for the individual rings are shown in the lower part of the table. The last column gives the corresponding atomic, molecular, and ring ID values.

Figure 4. Fourteen isomers of trimethylcycloheptane of Figure 3 ordered by increasing ID values.

Table III. Ring ID Values for the Rings in the 14 Trimethyltricycloheptanes (Shown in Figure 4)[a]

| molecule | three-membered ring, atoms 1, 2, 3 | five-membered rings | | |
		atoms 1, 3, 4, 5, 6	atoms 1, 2, 5, 6, 7	atoms 2, 3, 4, 5, 7
1	5.2501	8.5693*	8.3772	*
2	5.2611	8.6858	8.4999	8.5942
3	5.2602	8.6615	8.5661*	*
4	5.2756	8.5984	8.3328*	*
5	5.2740	8.7974	8.7242	8.6122
6	5.2738	8.5667*	8.3938	*
7	5.2733	8.6917	8.6184	8.5966
8	5.2691	8.6830*	*	*
9	5.2835	8.5921	8.5147*	*
10	5.2847	8.6857	8.5172	8.5949
11	5.2833	8.6065	8.5888*	*
12	5.2950	8.5904	8.4182	8.5133
13	5.2942	8.6377*	*	8.6243
14	5.3032	8.5401*	*	*

[a] The atoms constituting the individual rings are labeled as in Figure 3. Symmetry equivalent rings are marked by asterisks.

weights are computed internally. Once atoms of a molecular fragment are selected, the corresponding fragment ID numbers are derived. In particular, if atoms forming a ring are selected, a single number, a ring ID index, is obtained. Observe that the five-membered rings in the trimethylcycloheptane have similar, but different, ring ID values: 8.5966, 8.6184, and 8.6917. It remains to be seen if different rings, in general, will have different ring indices and if the derived numerical values for the ring indices preserve useful structural characteristics. In order to determine this, we consider the above questions in the next section with a collection of closely related compounds.

AN ILLUSTRATION

In Figure 4 we illustrate 14 trimethylcycloheptanes all having the same parent ring structure while differing in the pattern of methyl substitution. The compounds are ordered according to their respective molecular ID number. In Table III we collected the ring index values for the three-membered rings in these compounds. First, observe that the 14 different ring indices are obtained. The number of rings considered here is too small to conclude that, as a rule, we will always have different ring indices for different rings. However, the three-membered rings occur in very similar molecular environments: all molecules are of the same size and have the same parent ring structure. This is an indication that the ring ID index has quite a high resolution power. We also observe that three-membered rings without methyl substitution have the smallest ring ID values (from 5.2501 to 5.2691). Methyl substitution increases the ring index slightly (values from 5.2733 to 5.2847). Dimethyl-substituted rings have still larger ring ID values (5.2942 and 5.2950). Finally, the trimethyl-

substituted ring has the largest ring ID value (5.3032). It thus appears that the derived ring indices are not only different for different rings but that similar rings (i.e., an equal number of methyl substitutions) have similar ring ID values. Moreover, the *trend* among the ID ring values of rings of increasing substitution shows apparent structural regularity. These properties suggest that the procedure incorporates inherent local features which characterize rings rather well. The characterization, besides it application in chemical documentation, may have use in discussions of local molecular properties, because local structural similarity appears well preserved in the proposed ring ID index. From Figure 5 we see that even the minor variations in ring ID values for rings of the same type (e.g., unsubstituted rings) reflect special structural details (such as the number of next nearest neighbors!).

In Figure 8 we have summarized the results for five-membered rings in the 14 trimethyltricycloheptanes. As we see, again, the unsubstituted (five-membered) ring has the smallest ring ID (8.3328). Single methyl substituted five-membered rings have ID values from 8.3772 to 8.6122. The ring ID in rings substituted in two sites varies between 8.5133 and 8.6243. If methyls are on the same carbon, we have the maximal value of 8.7242. Finally, trisubstituted rings have ring ID values from 8.5676 to 8.7974. The values of ring indices overlap here, and it is somewhat difficult to disentangle the opposing influences of various structural factors because of the relatively
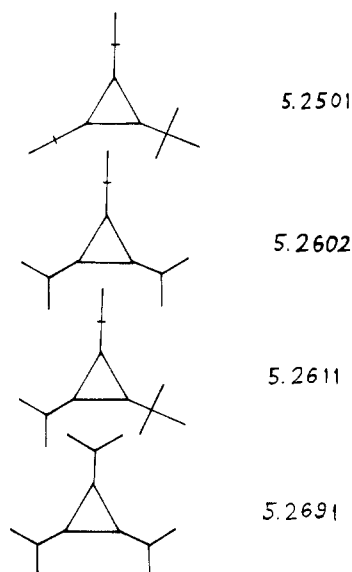
146 *J. Chem. Inf. Comput. Sci., Vol. 28, No. 3, 1988*

RANDIC

5.2501

5.2602

5.2611

5.2691

Figure 5. Next nearest neighbors for three-membered rings having no methyl substitution.

8.3328   8.3772   8.4999   8.4182   8.5401   8.5133
         8.3938   8.5172   8.6122   8.6243   8.5949

8.5147   8.5888   8.6065   8.5661   8.5942
8.5966   8.6377   8.6184   8.7242

8.5904   8.5976   8.5667   8.6857   8.6917

8.7974   8.5984   8.5693   8.6858   8.6615
                                   8.6850
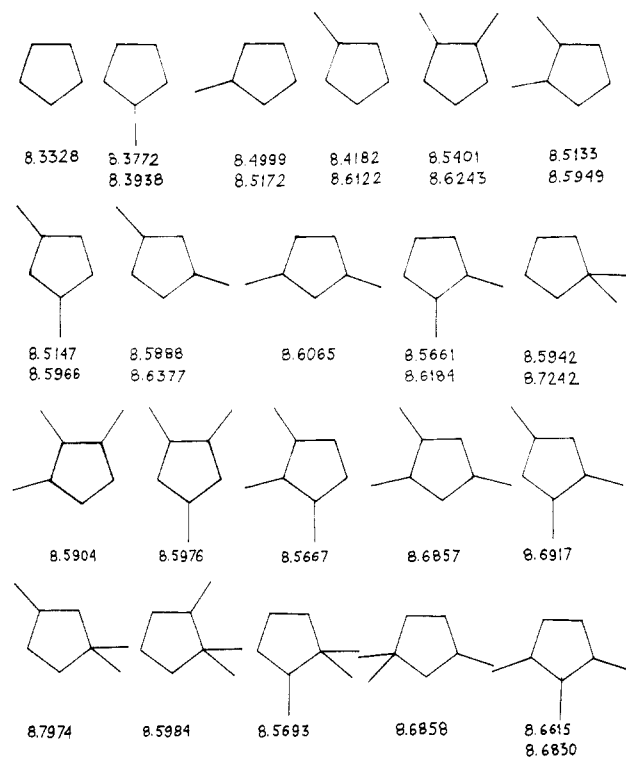
Figure 6. Distinct five-membered rings in the 14 trimethyltricyclo-heptanes of Figure 4. Atoms on same horizontal lines are symmetry equivalent in the unsubstituted rings.

STRUCTURAL FACTORS INFLUENCING CYCLICITY

In order to better understand the structural factors involved, we will consider additional illustrations. Topological characterization of cyclic structures with acyclic branches has been a subject of several investigations.[26] Mekenyan, Bonchev, and Trinajstić[27] examined how the Wiener number W, the sum of all paths, varies with variations in ring size and functionalities
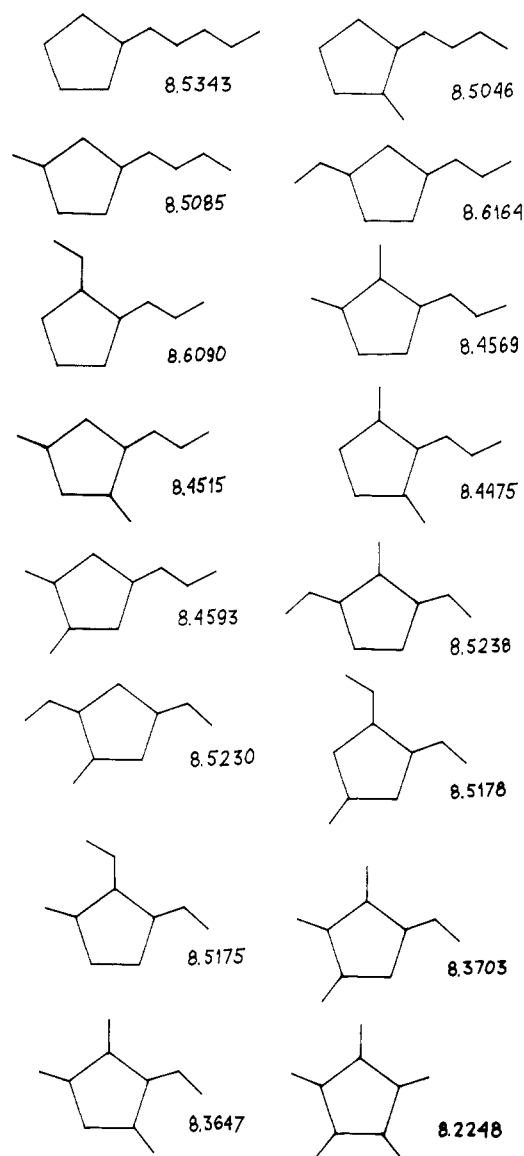
8.5343   8.5046

8.5085   8.6164

8.6090   8.4569

8.4515   8.4475

8.4593   8.5238

8.5230   8.5178

8.5175   8.3703

8.3647   8.2248

Figure 7. Variations in ring ID values with the type of ring substitution.

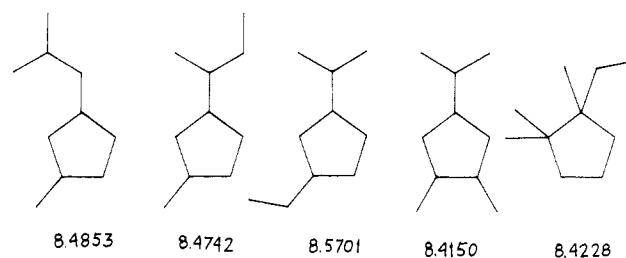8.4853   8.4742   8.5701   8.4150   8.4228

Figure 8. Five-membered ring in selected terpenes and the corresponding ring ID values.

(substitutions) in rings and formulated their observations in a number of rules. For example, one of their rules specifies that the number of newly formed branches (accompanied by a reduction of the length of the side chain) decreases the value of W. We illustrate in Figure 7 the above situation on 10 atomic skeletons having a five-membered ring. We have indicated the value of the ring ID for the structures considered and see that the same regularity found for W also holds for the ring ID. Similarly, we find a regular decrease in ring ID when branching in the side chain is varied. Finally, as the number of branches increases, we see that the ring ID decreases. If we restrict attention only to rings with unbranched substituents, we find the following: For rings with a single substituent, the ring ID value 8.5343 is found; for rings with

RING ID NUMBERS

J. Chem. Inf. Comput. Sci., Vol. 28, No. 3, 1988 **147**

two substituents, the interval of ring ID values is 8.5046–8.6164; for rings with three substituents, the interval becomes 8.4475–8.5238; and for rings with four substituents, ring ID decreases to 8.3647–8.3703. Finally, when all five substituted places have methyls, we have ring ID of 8.2248. Thus, we see a steady decrease in ring ID as the number of substituents increases. Ring index is sensitive to the positions and the number of substituents as is illustrated in Figure 8 on selected terpenes having a five-membered ring. The increase and then decrease of ring ID numbers on increasing substitution in a small polycyclic molecule having a constant number of atoms reflects subtle topological factors. To quote from Lewis Carrol, "...the further off to England, the nearer is to France"; i.e., increasing substitution for one ring is accompanied by a decreasing substitution for another ring. It is interesting to observe that the initial ring is also in this way affected in a subtle manner in its ring ID number.

CONCLUDING REMARKS

While it is certain that molecular descriptors based on selected (single or collection of) molecular invariants will not and cannot replace molecular codes that are unique and that fully incorporate information on a structure, molecular descriptors nevertheless serve a useful purpose in chemical documentation, even though their prime intent is to assist in discerning regularities in structure–property relationships. When one focuses attention on molecular fragments, in particular rings as illustrated here, it appears that suitably designed molecular invariants are not only useful but may become essential. We have illustrated suitability of weighted paths as descriptors and have seen not only that they discriminate among similar structural environments but that the descriptors for apparently similar rings generate similar numerical values. This becomes of interest when large files are screened, as it allows one to narrow down the range of targeted fragments. Therefore, it may be recommended as a useful practice to include ring ID values and other similar well-defined substructure components as additional molecular descriptors for structure files. This will not only facilitate fast substructure searches for well-defined substructure components but may also point to some local molecular properties. Let us indicate also that information on arbitrary fragments, including the individual bonds and even individual atoms, may in this way be extracted and emphasized. For example, one may use chemical intuition and experience to extract similar information for selected bonds only, such as the "strategic" bonds of Corey,[28] which are targets of particular chemical reactions. Finally, the suggested methodology based on the use of weighted paths has considerable flexibility by admitting variations in weighting procedures that may better suit special needs.

REFERENCES

(1) Read, R. C. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 1335.
(2) Knop, J. V.; Müller, W. R.; Jeričević, Z.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91.
(3) Randić, M. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 136.
(4) International Union of Pure and Applied Chemistry (Cowley Centre, Oxford, OX4 3YF, England).
(5) Herndon, W. C. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; p 231. Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171. Bonchev, D.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 233. Bonchev, D.; Mekenyan, O.; Balaban, A. T. In *Mathematics and Computational Concepts in Chemistry*; Trinajstić, N., Ed.; Ellis Horwood: Chichester, U.K., 1986; p 34.
(6) Randić, M. *J. Chem. Phys.* **1960**, *60*, 3920. Randić, M. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 171.
(7) Hosoya, H. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.
(8) Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 2636. Wiener, H. *J. Phys. Chem.* **1948**, *52*, 425, 1082.
(9) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(10) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
(11) Heilbronner, E.; Schmelzer, A. *Nouv. J. Chim.* **1980**, *4*, 23.
(12) Fisher, E. *Chem. Ber.* **1894**, *27*, 2985.
(13) Randić, M. In *Molecular Basis of Cancer (Part A: Macromolecular Structure, Carcinogens, and Oncogens)*; Alan R. Liss: New York, 1985; p 309.
(14) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
(15) Randić, M. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164.
(16) Randić, M.; Jerman-Blažič, B.; Grossman, S. C.; Rouvray, D. H. *Math. Modell.* **1986**, *8*, 571. Grossman, S. C.; Jerman-Blazic Dzonova, B.; Randić, M. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1986**, *12*, 123. Randić, M.; Jerman-Blažič, B.; Rouvray, D. H.; Seybold, P. G.; Grossman, S. C. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 245. Randić, M.; Grossman, S. C.; Jerman-Blažič, B.; Rouvray, D. H.; El-Basil, S. *Math. Comput. Modell.* **1988**, *11*, 837.
(17) Randić, M. *Croat. Chem. Acta* (submitted for publication).
(18) Randić, M. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1984**, *11*, 137.
(19) Okuyama, T.; Miyashita, Y.; Kanaya, S.; Katsumi, H.; Sasaki, S.; Randić, M. *J. Comput. Chem.* (in press).
(20) Randić, M. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 134.
(21) Szymanski, K.; Müller, W. R.; Knop, J. V.; Trinajstić, N. *Croat. Chem. Acta* **1986**, *59*, 719.
(22) Szymanski, K.; Müller, W. R.; Knop, J. V.; Trinajstić, N. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1986**, *20*, 173.
(23) Balaban, A. T. *J. Mol. Struct. (THEOCHEM)* (in press).
(24) Randić, M.; Brissey, G. M.; Spencer, R. B.; Wilkins, C. L. *Comput. Chem.* **1979**, *3*, 5.
(25) Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1226.
(26) Bonchev, D.; Mekenyan, Ov.; Knop, J. V.; Trinajstić, N. *Croat. Chem. Acta* **1969**, *52*, 361. Bonchev, D.; Mekenyan, Ov.; Trinajstić, N. *Int. J. Quantum Chem.* **1980**, *17*, 845. Mekenyan, Ov.; Bonchev, D.; Trinajstić, N. *MATCH* **1979**, *6*, 93. Mekenyan, Ov.; Bonchev, D.; Trinajstić, N. *Int. J. Quantum Chem.* **1981**, *19*, 929.
(27) Mekenyan, Ov.; Bonchev, D.; Trinajstić, N. *MATCH* **1981**, *11*, 145.
(28) Corey, E. J. *Q. Rev., Chem. Soc.* **1971**, *25*, 455.