CROATICA CHEMICA ACTA CCACAA **69** (1) 9–16 (1996)

ISSN 0011-1643
CCA–2283

Original Scientific Paper

Design of Topological Indices. Part 3.* New Identification Numbers for Chemical Structures: MINID and MINSID

Ovidiu Ivanciuc and Alexandru T. Balaban

Department of Organic Chemistry, Faculty of Chemical Technology, University »Politehnica« of Bucharest, Splaiul Independentei 313, 77206 Bucharest, Roumania

Received October 20, 1994; revised February 13, 1995; accepted February 14, 1995

Identification (ID) numbers are structural invariants based on weighted paths and walks in the molecular graph. They are useful molecular descriptors in the structure-property and structure-activity relationship studies, with very low degeneracy and high structure discriminating ability. For cyclic structures, the number of paths and walks increases very fast and the computation of ID numbers for large polycyclic molecular graphs is very time consuming. New identification numbers, MINID and MINSID, are defined using weighted distances of the molecular graph. Their computation is easier and faster than that of the previously defined ID numbers.

INTRODUCTION

In recent years, many different graph invariants have been proposed for the characterization of structural features of chemical species. These are usually referred to in the chemical literature as topological indices.[2–5] Such indices reflect in different ways the size and shape of the molecules they characterize and also provide some measure for the degree of molecular branching.

A number of useful graph definitions will be introduced. Let G = (V,E) be a graph G with N vertices, q edges, the vertex set V = V(G) and the edge set E = E(G). Let V = $\{v_1, v_2,...v_N\}$ be the labeling of G. The edge conneting

* Part 2, Ref. 1.

vertices v_i and v_j is denoted by e_{ij}. A walk in a graph is a sequence of edges, starting from any vertex and ending on any vertex, with the property that the ending vertex of the jth edge in the sequence is the beginning of the $(j+1)$th edge. The walk is called a path if all the vertices (and thus necessarily all the edges) are distinct. DEG_i denotes the degree of vertex v_i in G, *i.e.*, the number of edges incident with vertex v_i. The distance between the vertices v_i and v_j is denoted by d_{ij}, and is equal to the number of bonds on the shortest path between the vertices v_i and v_j.

The distance sum of vertex v_i is the sum of the topological distances between vertex v_i and every vertex in the molecular graph:[6]

$$DS_i = \sum_{j=1}^{N} d_{ij} \quad . \tag{1}$$

IDENTIFICATION NUMBERS

Randić has recently proposed a novel structural descriptor, called the molecular identification number ID,[7] defined by equations (2–4).

The mapping from the edge set E(G) of the chemical graph G to the set of real numbers R, f: E(G) $\rightarrow$ R, is defined by:

$$f(e_{ij}) = (DEG_i\ DEG_j)^{-1/2} \quad . \tag{2}$$

Let p be a path of m edges in G. Then, the mapping from the path set of graph G to the set of real numbers R is defined by:

$$f^*(p) = \prod_{}^{m} f(e_{ij}) \quad . \tag{3}$$

The connectivity ID number of the molecular graph is defined as:

$$ID(G) = N + \sum_{p} f^*(p) \tag{4}$$

where N is the number of elements in the vertex set V(G) of G. The summation in Eq. (4) goes over all distinct paths in G.

The selectivity of the connectivity ID number was tested for all alkane trees with up to 20 vertices.[8] In this large set of alkanes, 124 pairs and one triple of non-isomorphic alkane trees were found, having the same connectivity ID number. Randić proposed a prime ID number, denoted ID',[9] which has shown much greater selectivity than the connectivity ID numbers.[10]

Balaban[11] found a drawback of ID and ID' numbers: their values cluster according to the vertex partition. He suggested a different approach, namely to replace vertex degrees (DEG_i) in the original ID formula by distance sums (DS_i), obtaining a new highly selective identification number, SID.

Another identification number was defined by using weighted walks,[10] and a 3-dimensional identification number was recently used[12] to characterize conformations of small alkanes and cycloalkanes.

All the above identification numbers, *i.e.* ID, ID', WID and SID, are easy to compute only for acyclic molecules and for those with a small number of cycles. For large polycyclic molecular graphs, there is a very large number of paths and walks, respectively, and this slows dawn the computation of the above identification numbers.

DISTANCE IDENTIFICATION NUMBERS: MINID AND MINSID

We introduce two new identification numbers, MINID and MINSID, based on weighted distances. Thus, the enumeration of walks/paths in a graph is replaced by an easier task, that of determining weighted distances.

Let d_{ij} be the distance between vertices v_i and v_j, consisting of m edges. Then, the mapping from the distance set of graph G to the set of real numbers R is defined by:

$$f'(d_{ij}) = \prod^{m} f(e_{ij}) \tag{5}$$

where $f(_{ij})$ is defined in expression (2).

The identification number MINID of G is defined as:

$$\text{MINID(G)} = N + \sum_{d} f'(d_{ij}) \tag{6}$$

where N is the number of elements in the vertex set V(G) of G. The summation in Eq. (6) goes over all distances in G.

Obviously, the computation of MINID numbers for large polycyclic graphs is much faster than that of ID numbers. The similarity between the Randić ID number and the MINID number gives two common deficiencies: the clustering according to the vertex partition, and the degeneracy due to the fact that only $2^{-1/2}$, $3^{-1/2}$ and $6^{-1/2}$ can appear as real number coefficients in their expressions. A more convenient distance identification number is defined on the basis of distance sums.

Let d_{ij} be the distance between vertices v_i and v_j, consisting of m edges. Then, the mapping from the distance set of graph G to the set of real numbers R is defined by:

$$g'(d_{ij}) = \prod^{m} g(e_{ij}) \tag{7}$$

where $g(e_{ij})$ is:

$$g(e_{ij}) = (DS_i\ DS_j)^{-1/2}\ . \tag{8}$$

The identification number MINSID of the molecular graph G is defined as:

$$MINSI\Delta(G) = N + \sum_{d} g'(d_{ij}) \tag{9}$$

where N is the number of elements in the vertex set V(G) of G. The summation in Eq. (9) goes over all distances in G.

For acyclic molecular graphs, the path set and the distance set are identical. It follows that if the molecular graph is a tree T, the path ID numbers and the distance ID numbers are equal:

$$ID(T) = MINID(T)$$
$$SID(T) = MINSID(T)$$

TABLE I

Topological indices ID, SID, MINID and MINSID for graphs of alkyl-substituted 6-membered cycloalkanes

No.	Name	ID	SID	MINID	MINSID
1	cC_6	11.81250	6.74999	10.87500	6.74486
2	Me–cC_6	13.67683	7.67404	12.75352	7.67134
3	1,1-Me_2–cC_6	15.45879	8.64800	14.54810	8.64638
4	1,2-Me_2–cC_6	15.53341	8.62773	14.61540	8.62614
5	1,3-Me_2–cC_6	15.53296	8.61136	14.67044	8.60992
6	1,4-Me_2–cC_6	15.53283	8.59945	14.72574	8.59814
7	Et–cC_6	15.69267	8.57727	14.73885	8.57601
8	1,1,2-Me_3–cC_6	17.31156	9.61017	16.40132	9.60914
9	1,1,3-Me_3–cC_6	17.31071	9.58656	16.46285	9.58562
10	1,1,4-Me_3–cC_6	17.31046	9.56947	16.52486	9.56861
11	1,2,3-Me_3–cC_6	17.38253	9.58470	16.52877	9.58376
12	1,2,4-Me_3–cC_6	17.38202	9.56594	16.59289	9.56508
13	1,3,5-Me_3–cC_6	17.38172	9.56389	16.66175	9.56310
14	1-Et-1-Me–cC_6	17.46310	9.57430	16.52952	9.57345
15	1-Et-2-Me–cC_6	17.54893	9.55193	16.60258	9.55109
16	1-Et-3-Me–cC_6	17.54945	9.53367	16.65873	9.53291
17	1-Et-4-Me–cC_6	17.54959	9.52016	16.71458	9.51947
18	Pr–cC_6	17.70059	9.49237	16.73152	9.49179
19	*i*Pr–cC_6	17.52861	9.53336	16.56936	9.53267

For cyclic molecular graphs, the set of distances represents a subset of the set of paths. Therefore, if G is a cyclic graph, the path ID number are greater than distance ID numbers:

$$ID(G) > MINID(G)$$
$$SID(G) > MINSID(G)$$

In Table I, we present values of ID numbers ID, SID, MINID and MINSID for 19 isomeric alkylcyclohexanes. The abbreviations used in the nomenclature of alkanes are: Me = methyl, Et = ethyl, Pr = propyl, *i*Pr = isopropyl and the ring of six carbon atoms is symbolized by cC_6. In Table II, values of distance ID numbers MINID and MINSID are presented for the 23 polycyclic benzoids (polyhexes) with 3, 4 and 5 benzenoid rings (Figure 1).

It can be seen from Table I that MINID and MINSID order compounds similarly, but not identically; MINID orders compounds quite similarly to *J*: higher MINID values correspond to a higher degree of branching. As seen

TABLE II

Topological indices MINID and MINSID for the polyhexes from Figure 1

Polyhexe	MINID	MINSID
1	21.30911	14.69419
2	21.32995	14.66920
3	27.61351	18.61321
4	27.58895	18.59163
5	27.58838	18.57246
6	27.60911	18.56214
7	27.62984	18.54360
8	24.61397	16.69715
9	33.89422	22.53347
10	33.86857	22.51808
11	33.86835	22.50246
12	33.86770	22.48490
13	33.88836	22.47947
14	33.88845	22.48005
15	33.91478	22.51762
16	33.88920	22.49757
17	33.88841	22.48719
18	33.90925	22.47543
19	33.90908	22.47030
20	33.92980	22.45795
21	30.91815	20.60286
22	30.91872	20.60606
23	30.90597	20.57931

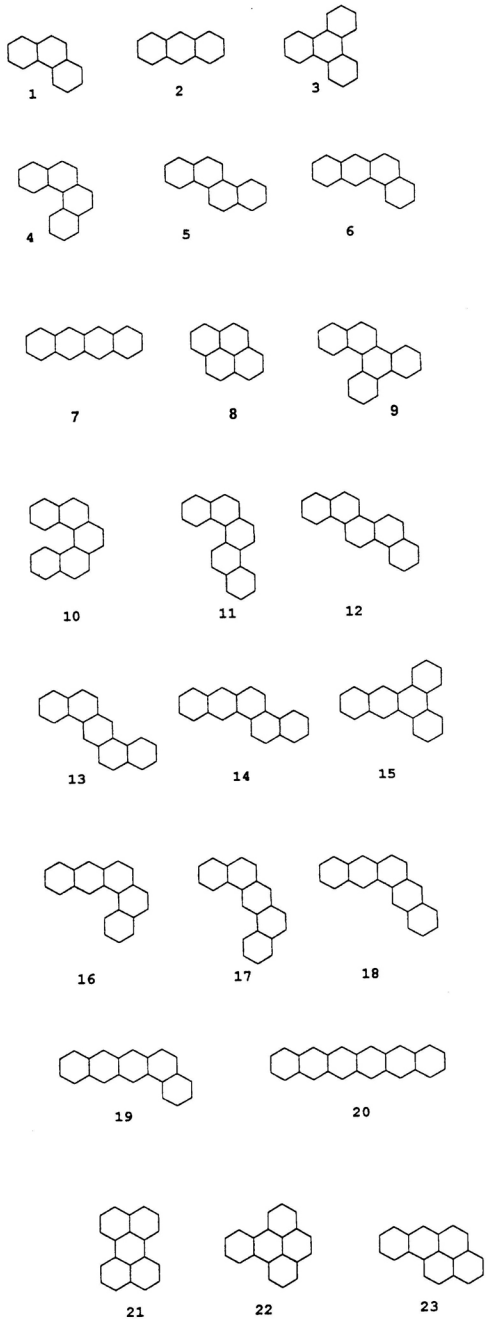

Figure 1. Polyhexes with 3, 4 and 5 benzenoid rings whose distance ID numbers MINID and MINSID are presented in Table II.

in Table II *cata-* and *peri*-condensed benzenoids with the same number of rings differ markedly in their MINID and MINSID values because they have different numbers, N, of carbon atoms.

CONCLUDING REMARKS

The path and walk ID numbers, namely ID, ID', SID and WID, are useful graph theoretic descriptors. Their main drawback is the long computer time needed for their computation for large polycyclic graphs.

The weighted distance ID numbers, namely MINID and MINSID, are computationally simpler than any ID number defined before. Therefore, the weighted distance ID numbers are proposed for further work in the structure–property and structure–activity relationship studies. We are now trying to establish the range of applicability and the structural selectivity of the ID numbers MINID and MINSID.

Acknowledgement. – We thank the Ministry of Research and Technology for financial support of this research under Grants 572B TA4 and TB11.

REFERENCES

1. Part 2, A. T. Balaban, D. Ciubotariu, and O. Ivanciuc, *MATCH (Commun. Math. Chem.)* **25** (1990) 41.
2. A. T. Balaban, I. Motoc, D. Bonchev, and O. Mekenyan, *Top. Curr. Chem.* **114** (1983) 21.
3. A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **25** (1985) 334.
   A. T. Balaban, *J. Mol. Struct. (Theochem)* **120** (1985) 117.
   A. T. Balaban, *J. Mol. Struct. (Theochem)* **165** (1988) 243.
   A. T. Balaban, *Croat. Chem. Acta* **66** (1993) 447.
   A. T. Balaban, *Rev. Roum. Chim.* **39** (1994) 245.
   A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **34** (1994) 398.
4. P. J. Hansen and P. C. Jurs, *J. Chem. Educ.* **65** (1988) 574.
5. N. Trinajstić, *Chemical Graph Theory,* CRC Press, Boca Raton, Florida, 1983, Vol. 2, p. 105; A. Sabljić and N. Trinajstić, *Acta Pharm. Jugosl.* **31** (1982) 189.
6. D. Bonchev, A. T. Balaban, and O. Mekenyan, *J. Chem. Inf. Comput. Sci.* **20** (1980) 106.
7. M. Randić, *J. Chem. Inf. Comput. Sci.* **24** (1984) 164.
8. K. Szymansky, W. R. Müller, J. V. Knop, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **25** (1985) 413.
9. M. Randić, *J.Chem. Inf. Comput. Sci.* **26** (1986) 134.
10. K. Szymansky, W. R. Müller, J. V. Knop, and N. Trinajstić, *Int. J. Quantum Chem.: Quantum Chem. Symp.* **20** (1986) 173.
11. A. T. Balaban, in: *Graph Theory and Topology in Chemistry*, R. B. King and D. H. Rouvray (Eds.), *Stud. Phys. Theor. Chem.* Elsevier, Amsterdam, **51** (1987) 159.
12. M. Randić, B. Jerman-Blažić, and N. Trinajstić, *Comput. Chem.* **14** (1990) 237.

SAŽETAK

Dizajn topoloških indeksa. 3. dio. Novi brojevi za identifikaciju kemijskih strukture: MINID i MINSID

Ovidiu Ivanciuc i Alexandru T. Balaban

Identifikacijski (ID) brojevi strukturne su invarijante zasnovane na uteženim stazama i šetnjama u molekularnom grafu. Oni su korisni molekularni deskriptori u proučavanju odnosa između strukture i svojstava, te strukture i aktivnosti, i pokazuju vrlo nisku degeneraciju, s visokom sposobnošću razlikovanja struktura. Broj staza i šetnji za cikličke strukture vrlo brzo raste pa je i računanje ID brojeva za velike policikličke molekularne grafove vremenski zahtjevno. Novi identifikacijski brojevi, MINID i MINSID, definirani su s pomoću uteženih udaljenosti u molekularnom grafu. Njihovo je izračunavanje lakše i brže nego za dosad definirane ID brojeve.