Design of topological indices.
Part 4*. Reciprocal distance matrix, related local vertex invariants and topological indices*

Ovidiu Ivanciuc[1], Teodor-Silviu Balaban[2] and Alexandru T. Balaban[1]

[1]*Department of Organic Chemistry, Polytechnic Institute, Splaiul Independentei 313, 77206 Bucharest, Romania*

[2]*Institute of Organic Chemistry of the Romanian Academy, Splaiul Independentei 202, Bucharest, Romania*

A new molecular graph matrix, the reciprocal distance (RD) matrix, is defined. Its nondiagonal elements are equal to the reciprocals of the topological distances between the corresponding vertices, while the diagonal elements are all equal to zero. Based on the RD matrix, a real-number local vertex invariant, RDS was proposed, and three topological indices, namely RDSUM, RDSQ, and RDCHI, were defined. Their degeneracy was investigated and proved to be lower than that of the topological index W based on the distance matrix. The correlational ability of the new molecular descriptors was tested against van der Waals molecular surfaces and boiling points of alkanes, showing a satisfactory monoparametric dependence.

1. Introduction

Molecular topology determines a large number of molecular properties ranging from physico-chemical and thermodynamic properties to chemical reactivity and biological activity. In this respect, chemical applications of graph theory have undergone dramatic expansions in recent years [2–5].

From a practical point of view, an efficient way of coding the topology of a chemical structure is represented by topological indices [6,7]. A topological index (TI) is a numerical quantity which characterizes the bonding topology of a molecule. Such indices reflect in different ways the size and shape of the molecules they characterize, and also provide some measure of the degree of molecular branching.

By removing all hydrogen atoms from the chemical formula of a chemical compound containing covalent bonds, we obtain the hydrogen-depleted graph (or

---

*For part 3 of this series, see ref. [1].
*This paper is dedicated to Frank Harary on the occasion of his 70th anniversary. We agree with N. Trinajstić's proposal to call RDSUM the "Harary number".

molecular graph) of that compound, whose vertices correspond to non-hydrogen atoms. In the particular case of hydrocarbons, the vertices of the molecular graph denote carbon atoms.

A few useful graph-theoretical definitions will be introduced. Let $G = (V, E)$ be a graph G with N vertices, q edges, the cyclomatic number $\mu = q - N + 1$, the vertex set $V(G)$, and the edge set $E(G)$. The cyclomatic number μ represents the number of cycles in the graph, DEG_i denotes the degree of the vertex i in G, i.e. the number of edges incident with the vertex i. Acyclic graphs have $\mu = 0$ and are called trees.

The topology of a chemical structure can be coded in matrix form by the use of the adjacency matrix. The adjacency matrix of a graph G with N vertices, $A(G) = A$, is the square $N \times N$ symmetric matrix which contains information about the connectivity of vertices in G. Its entries are defined as

$$(A)_{ij} = \begin{cases} 1 & \text{for vertices } i, j \text{ adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

The sum of entries over row i or column i in $A(G)$ is DEG_i.

As an example, the molecular graph and the adjacency matrix of 3-methylhexane (T_1) are given below. This graph is the smallest identity tree because its only symmetry operation is the identity (no vertices are equivalent).
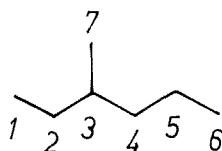
T_1

$$A(T_1) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The distance matrix of a graph G with N vertices, $D(G) = D$, is a square $N \times N$ symmetric matrix, whose entries $(D)_{ij}$, are equal to the number of edges between vertices i and j on the shortest path between them. Thus, entries 1 are the same as in $A(G)$, and its main diagonal entries are equal to zero as in $A(G)$, but all other entries are integers larger than 1. In the recent literature, there are many

efficient algorithms available for computing the distance matrix for any molecular graph [8,9]. The numerous applications of the distance matrix to various branches of chemistry were recently surveyed [10]. The distance matrix of the graph T_1 is shown below:

$$\mathbf{D}(T_1) = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 3 \\ 1 & 0 & 1 & 2 & 3 & 4 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 & 1 \\ 3 & 2 & 1 & 0 & 1 & 2 & 2 \\ 4 & 3 & 2 & 1 & 0 & 1 & 3 \\ 5 & 4 & 3 & 2 & 1 & 0 & 4 \\ 3 & 2 & 1 & 2 & 3 & 4 & 0 \end{bmatrix}.$$

The distance sum of the vertex i, DS_i, is defined as the sum of the topological distances between vertex i and every vertex in the molecular graph, i.e. the sum over row i or column i in $\mathbf{D}(G)$,

$$DS_i = \sum_{j=1}^{N} (\mathbf{D})_{ij}.$$

2. Topological indices

The high correlational ability of TIs with a wide range of physico-chemical and biological properties was theoretically interpreted by Klein [11,12]. He suggested that TIs are low-order cluster expansions of the chemical structure, with a rapid convergence. A brief presentation of some of the most important TIs will follow.

A graph-theoretic invariant (GTI) is termed additive [11,12] if

$$GTI(G_1 \cup G_2) = GTI(G_1) + GTI(G_2),$$

where $G_1 \cup G_2$ denotes the graph composed of two disconnected components G_1 and G_2. Such an additive TI is the topological index W, introduced in 1947 by Wiener [13,14] for predicting the alkane boiling points. Wiener extended the application of the W index to other physical properties of alkanes such as heats of formation, heats of vaporization, molar volumes, and molar refractions.

In the original work by Wiener, the W index was not formulated explicitly in graph-theoretical terms. Hosoya [15] pointed out that the Wiener index can be obtained from the distance matrix of the molecular graph, and is equal to the half-sum of the elements of the distance matrix $\mathbf{D}$. The Wiener index may be expressed also as the half-sum of the distance sums of the molecular graph:

$$W = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{D})_{ij} = \frac{1}{2} \sum_{i=1}^{N} DS_i.$$

The relatively high degeneracy of W may be reduced by using information theory, as shown by Bonchev and Trinajstić [16], who devised ways for increasing the correlation ability while decreasing the degeneracy of topological indices.

Another additive graph-theoretic invariant is the Randić connectivity index, defined as [17]

$$\chi = \sum_{E(G)} (\text{DEG}_i \, \text{DEG}_j)^{-1/2},$$

where DEG_i and DEG_j denote the degrees of the two endpoints of an edge e_{ij} in the molecular graph. The summation is extended over all edges in the molecular graph. The TI χ is the most used one as a molecular descriptor in structure–property and structure–activity relationship studies [18,19].

The highly discriminating TI J (average distance sum connectivity) was defined by the formula [20,21]

$$J(G) = \frac{q}{\mu + 1} \sum_{E(G)} (\text{DS}_i \, \text{DS}_j)^{-1/2},$$

where DS_i and DS_j denote the distance sums of the endpoints of an edge in the molecular graph and the summation goes over all edges in the molecular graph.

Attempts were made to reveal the physical meaning of a selected set of the most used TIs [22,23]. The interesting finding was that TIs based on the distance matrix, namely W and J, as well as indices derived from information theory, I_D^E, $\bar{I}_D^E$, I_D^W and $\bar{I}_D^W$, reflect accurately van der Waals areas, while indices such as $^3\chi$ (connectivity calculated for paths of length 3 instead of edges, which are paths of length 1, as in Randić's connectivity χ defined above) represent a measure of the molecular van der Waals volumes, as indicated by correlations with molar refractions or densities of alkanes.

3.   Reciprocal distance matrix

From the definition of the Wiener index, it is apparent that the larger the distance sum, the greater the associated weight for a given vertex.

Usually, any interaction must decrease with the increase of the distance between the interacting particles. We define a new molecular graph matrix, in which the weight of the distance between two atoms decreases with the increase of the distance.

The reciprocal distance matrix of a graph G with N vertices, $\text{RD}(G) = \text{RD}$, is a square $N \times N$ symmetric matrix whose entries $(\text{RD})_{ij}$ are equal to the reciprocal of the distance between vertices i and j for non-diagonal elements, and are equal to zero for the diagonal elements.

$$(\text{RD})_{ij} = \begin{cases} 0 & \text{for } i = j, \\ d_{ij}^{-1} & \text{for } i \neq j. \end{cases}$$

The RD matrix of the tree T_1 corresponding to the molecular graph of 3-methylhexane is presented below (the real-number entries are truncated here to two decimal digits):

$$\mathrm{RD}(T_1) = \begin{bmatrix} 0 & 1 & 0.50 & 0.33 & 0.25 & 0.20 & 0.33 \\ 1 & 0 & 1 & 0.50 & 0.33 & 0.25 & 0.50 \\ 0.50 & 1 & 0 & 1 & 0.50 & 0.33 & 1 \\ 0.33 & 0.50 & 1 & 0 & 1 & 0.50 & 0.50 \\ 0.25 & 0.33 & 0.50 & 1 & 0 & 1 & 0.33 \\ 0.20 & 0.25 & 0.33 & 0.50 & 1 & 0 & 0.25 \\ 0.33 & 0.50 & 1 & 0.50 & 0.33 & 0.25 & 0 \end{bmatrix}.$$

The RD matrix was successfully used in a study concerning computer generation of acyclic graphs based on local vertex invariants and topological indices [24].

4. New graph invariants (local and global) based on the reciprocal distance matrix

The reciprocal distance sum of the vertex i, RDS_i, is defined as the sum of the elements of the RD matrix over row i or column i:

$$\mathrm{RDS}_i = \sum_{j=1}^{N} (\mathrm{RD})_{ij}.$$

The RDS vector for the tree T_1 is given below for vertices 1–7, respectively.

$$\mathrm{RDS}(T_1) = \{2.61667, 3.58333, 4.33333, 3.83333, 3.41667, 2.53333, 2.91667\}.$$

This local vertex (atomic) invariant (LOVI) is a real number, like several new LOVIs defined recently [25].

We note that the RDS vector is identical to the recently defined vertex topological index VTI_{10} [26].

In the following section, three new global (molecular) invariants (topological indices) which are real numbers will be introduced.

On the basis of the RD matrix, we define the RDSUM index, equal to the half-sum of the elements of the reciprocal distance matrix RD. The RDSUM index may also be expressed as the half-sum of the RDS vector of the molecular graph:

$$\mathrm{RDSUM} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathrm{RD})_{ij} = \frac{1}{2} \sum_{i=1}^{N} \mathrm{RDS}_i.$$

Another index defined on the basis of the RDS vector is the RDSQ index:

$$RDSQ = \sum_{E(G)} (RDS_i \, RDS_j)^{1/2}.$$

In analogy with the Randić χ index, we define the RDCHI index:

$$RDCHI = \sum_{E(G)} (RDS_i \, RDS_j)^{-1/2}.$$

In table 1, we present the values of the three topological indices RDSUM, RDSQ, and RDCHI for the set of 19 alkanes with 4–7 carbon atoms.

Table 1

Topological indices RDSUM, RDSQ, and RDCHI for the set
of 19 alkanes with 4–7 carbon atoms.

Alkane	RDSUM	RDSQ	RDCHI	A^a	BP^b
butane	4.33333	6.78174	1.33420	–	– 0.5
2-Me-propane	4.50000	7.34847	1.22474	–	– 10.5
pentane	6.41667	10.69008	1.50919	–	36.5
2-Me-butane	6.66667	11.50536	1.40069	–	27.9
2,2-Me$_2$-propane	7.00000	12.64911	1.26491	–	9.5
hexane	8.70000	15.05183	1.67761	–	68.7
2-Me-pentane	9.00000	16.00403	1.57682	–	60.2
3-Me-pentane	9.08333	16.24542	1.55857	–	63.2
2,3-Me$_2$-butane	9.33333	17.06395	1.47474	–	58.1
2,2-Me$_2$-butane	9.50000	17.63881	1.43021	–	49.7
heptane	11.15000	19.78208	1.84000	334.36	98.4
2-Me-hexane	11.48333	20.82106	1.74635	322.91	90.0
3-Me-hexane	11.61667	21.19445	1.72321	316.67	91.8
3-Et-pentane	11.75000	21.56689	1.69795	303.15	93.5
2,4-Me$_2$-pentane	11.83333	21.90865	1.65407	309.97	80.5
2,3-Me$_2$-pentane	12.00000	22.38932	1.62717	303.11	89.8
2,2-Me$_2$-pentane	12.08333	22.72680	1.60437	306.53	79.2
3,3-Me$_2$-pentane	12.25000	23.20776	1.57545	297.20	86.0
2,2,3-Me$_3$-butane	12.50000	24.02922	1.50909	292.89	80.9

[a]Molecular van der Waals area (Å^2).
[b]Boiling point at normal pressure (°C).

While RDSUM and RDSQ increase both with molecular size and with branching degree, RDCHI (like χ) increases with molecular size but decreases with molecular branching, as seen in table 1. The intermolecular ordering of isomeric alkanes (table 1) for all three new TIs parallels exactly that induced by I_D^W, and differs from that induced by J or by Bertz's graph derivatives [27] only in the permuted order of 3-ethylpentane and 2,4-dimethylpentane.

In a study concerning structural selectivities of TIs [28], the Wiener index was found to have a good correlational ability, but a low selectivity (high degeneracy) when compared with the set of studied TIs. The first two degenerate pairs of isomeric alkanes with identical Wiener index are encountered for heptanes, namely:

$$W(2,2\text{-}Me_2\text{-}C_5) = W(2,3\text{-}Me_2\text{-}C_5) = 46,$$

$$W(3\text{-}Et\text{-}C_5) = W(2,4\text{-}Me_2\text{-}C_5) = 48.$$

The degeneracy of the three new indices was tested up to and including the pentadecanes. While RDSUM was found to be degenerate for two nonanes,

$$RDSUM(2,2\text{-}Me_2\text{-}C_7) = RDSUM(2,3\text{-}Me_2\text{-}C_7) = 17.55000$$

and for higher alkanes, the other indices were nondegenerate for all alkanes tested. However, they will be degenerate for some octadecanes which were shown to present degeneracy for indices based on distances [29]. A pair of octadecanes with identical distance degree sequence and degenerate distance-based TIs is represented by the trees T_2 and T_3,

$$T_2 \qquad\qquad T_3$$
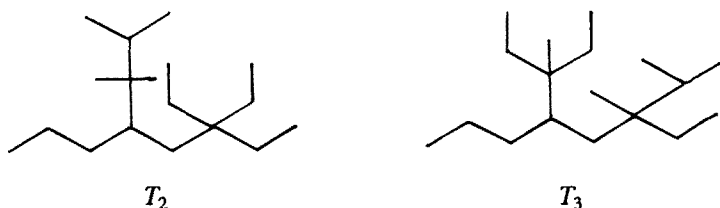
$$W(T_2) = W(T_3) = 583,$$
$$RDSUM(T_2) = RDSUM(T_3) = 55.23571,$$
$$RDSQ(T_2) = RDSQ(T_3) = 113.74407,$$
$$RDCHI(T_2) = RDCHI(T_3) = 2.59963.$$

One must note that high selectivity is not sufficient to obtain good results in structure–property correlations if the values of the TIs do not reflect properly the connection between the chemical structure and a certain property. On the other hand, a low degeneracy of a TI is highly desirable, ensuring a selective characterization among isomers.

Monoparametric correlations with van der Waals areas of the nine heptane isomers were tested for the TIs with the following results, where r is the correlation coefficient and s is the standard deviation:

	r	s
$A = 164.830\ (\pm 36.821) + \quad 3.017(\pm 0.764)W$	0.979	2.823
$A = 667.641(\pm 146.279) - \quad 30.206(\pm 12.336)RDSUM$	-0.949	4.408
$A = 518.118\ (\pm 93.106) - \quad 9.494(\pm 4.234)RDSQ$	-0.940	4.767
$A = 104.763\ (\pm 93.440) + 123.111(\pm 56.059)RDCHI$	0.938	4.856

As is apparent from the equations, the Wiener index gives the best correlation with the van der Waals area; another finding is the negative correlation coefficient between molecular area and indices RDSUM and RDSQ.

Monoparametric correlations with boiling points (BP) (at normal pressure) for all 72 alkanes with $n = 4-9$ carbon atoms were tested for the TIs with the following results:

	r	s
$BP = \quad 20.599(\pm 11.137) + \quad 1.229(\pm 0.142)W$	0.940	12.538
$BP = -25.904(\pm 14.053) + \quad 9.159(\pm 0.911)RDSUM$	0.954	10.988
$BP = -11.936(\pm 15.352) + \quad 4.331(\pm 0.521)RDSQ$	0.935	13.024
$BP = \quad 191.930(\pm 52.525) + 171.346(\pm 29.543)RDCHI$	0.878	17.528

It is known that the boiling points of normal alkanes show a nonlinear increase with an increasing number of carbon atoms. It was demonstrated that the variation can be interpreted in terms of the fractal dimension of alkanes [30]. Following this finding, monoparametric correlations between boiling points and the logarithm of the TIs under consideration were tested for the same set of 72 alkanes:

	r	s
$BP = -158.158(\pm 15.559) + 148.139(\pm 8.488)\lg(W)$	0.984	6.526
$BP = -172.507(\pm 23.981) + 245.412(\pm 20.599)\lg(RDSUM)$	0.967	9.391
$BP = -197.201(\pm 30.998) + 215.629(\pm 21.550)\lg(RDSQ)$	0.954	11.032
$BP = \quad -53.816(\pm 26.300) + 672.991(\pm 105.345)\lg(RDCHI)$	0.896	16.243

For the same reasons, monoparametric correlations between boiling points and the square root of the TIs under investigation were tested for the same set of 72 alkanes:

	r	s
$BP = \quad -47.043(\pm 12.226) + \quad 18.883(\pm 1.426)W^{1/2}$	0.973	8.499
$BP = -134.060(\pm 21.249) + \quad 64.026(\pm 5.495)RDSUM^{1/2}$	0.965	9.589
$BP = -107.762(\pm 22.928) + \quad 41.565(\pm 4.303)RDSQ^{1/2}$	0.951	11.389
$BP = -485.245(\pm 97.893) + 449.076(\pm 73.618)RDCHI^{1/2}$	0.888	16.851

One can conclude that in linear monoparametric correlations with normal boiling points of alkanes, RDSUM behaves somewhat better that W (comparably to Hosoya's index Z). It should be noted, however, that the above r and s values are only fair, and that for correlations with boiling points, much better topological indices were developed (e.g. ref. [31]).

In the above correlations with boiling points, the molecular size rather than shape is the factor of prime importance. The ability of the TIs to reflect the molecular shape was tested in correlation with the boiling points of the set of 35 nonanes. The shortcomings of the indices are clearly revealed by the very low correlation coefficient, equal to -0.315 for RDSUM, -0.333 for RDSQ, and 0.403 for RDCHI, denoting that the main contribution to the three TIs under study is represented by the molecular size, and with a small contribution of the molecular shape. A similar observation concerning the Wiener index was recently published [32].

Table 2 presents the intercorrelation matrix of the TIs χ, J, W, RDSUM, RDSQ, and RDCHI for the set of 72 alkanes with 4–9 carbon atoms. The high intercorrelation coefficient between RDSUM and RDSQ indicates that these TIs

Table 2

Intercorrelation matrix of TIs χ, J, W, RDSUM, RDSQ, and RDCHI.

	RDCHI	RDSQ	RDSUM	W	J	χ
χ	0.930	0.900	0.927	0.968	0.465	1.000
J	0.120	0.773	0.731	0.408	1.000	
W	0.920	0.887	0.915	1.000		
RDSUM	0.745	0.998	1.000			
RDSQ	0.697	1.000				
RDCHI	1.000					

express approximately the same type of structural information. On the other hand, as expected from the definition, RDCHI correlates poorly with RDSUM and RDSQ. The correlation with three widely used TIs brings new information: the Randić χ index correlates best with the three RD matrix indices, followed by the Wiener index, whereas the TI J does not correlate with RDCHI and expresses only a low correlation with RDSUM and RDSQ.

Independently, graph-theoretic invariants similar to RD and RDSUM were recently defined by Trinajstić et al. [33].

A different approach to LOVIs and TIs where the weight of distances decreases with their increase appeared recently [34].

Acknowledgement

References

[1]   O. Ivanciuc and A.T. Balaban, MATCH, submitted.
[2]   A.T. Balaban, J. Chem. Inf. Comput. Sci. 25(1985)334.
[3]   A.T. Balaban, J. Mol. Struct. 120(1985)117.
[4]   K. Balasubramanian, Chem. Rev. 85(1985)599.
[5]   N. Trinajstić, D.J. Klein and M. Randić, Int. J. Quant. Chem. Quant. Chem. Symp. 20(1986)699.
[6]   A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Topics Curr. Chem. 114(1983)21.
[7]   P.J. Hansen and P.C. Jurs, J. Chem. Educ. 65(1988)574.
[8]   W.R. Müller, K. Szymanski, J.V. Knop and N. Trinajstić, J. Comput. Chem. 8(1987)170.
[9]   P. Senn, Comput. Chem. 12(1988)219.
[10] D.H. Rouvray, in: *Mathematical and Computational Concepts in Chemistry*, ed. N. Trinajstić (Ellis Horwood, Chichester, 1986) ch. 25, p. 295.
[11] D.J. Klein, in: *Mathematical and Computational Concepts in Chemistry*, ed. N. Trinajstić (Ellis Horwood, Chichester, 1986) ch. 16, p. 171.
[12] D.J. Klein, Int. J. Quant. Chem. Quant. Chem. Symp. 20(1986)153.
[13] H. Wiener, J. Am. Chem. Soc. 69(1947)17.
[14] H. Wiener, J. Am. Chem. Soc. 69(1947)2636.
[15] H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.
[16] D. Bonchev and N. Trinajstić, J. Chem. Phys. 67(1977)4517.
[17] M. Randić, J. Am. Chem. Soc. 97(1975)6609.
[18] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[19] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis* (Wiley, New York, 1986).
[20] A.T. Balaban, Chem. Phys. Lett. 89(1982)399.
[21] A.T. Balaban, Pure Appl. Chem. 55(1983)199.
[22] I. Motoc and A.T. Balaban, Rev. Roum. Chim. 26(1981)593.
[23] J.K. Labanowski, I. Motoc and R.A. Dammkoehler, Comput. Chem. 15(1991)47.
[24] T.-S. Balaban, P.A. Filip and O. Ivanciuc, J. Math. Chem. 11(1992)79.
[25] A.T. Balaban, J. Chem. Inf. Comput. Sci. 32(1992)23; J. Math. Chem., in press;
      A.T. Balaban and T.S. Balaban, J. Math. Chem. 8(1991)383; J. Chim. Phys. 89(1992)1735.
[26] O. Ivanciuc, Rev. Roum. Chim. 34(1989)1361.
[27] S.H. Bertz, Discr. Appl. Math. 19(1988)65, reprinted in: *Applications of Graphs in Chemistry and Physics*, ed. J.W. Kennedy and L.V. Quintas (North-Holland, Amsterdam, 1988).
[28] M. Razinger, J.R. Chrétien and J.E. Dubois, J. Chem. Inf. Comput. Sci. 25(1985)23.
[29] O. Ivanciuc, T.-S. Balaban and A.T. Balaban, J. Math. Chem., in press.
[30] D.H. Rouvray and R.B. Pandey, J. Chem. Phys. 85(1986)2286.
[31] P.A. Filip, T.S. Balaban and A.T. Balaban, J. Math. Chem. 1(1987)61.
[32] D.H. Rouvray, J. Comput. Chem. 8(1987)470; J. Mol. Struct. THEOCHEM 54(1989)187.
[33] N. Trinajstić et al., J. Math. Chem., in press.
[34] M. Diudea, O. Minailiuc and A.T. Balaban, J. Comput. Chem. 12(1991)527.