

## CTI: A NOVEL CHARGE-RELATED TOPOLOGICAL INDEX WITH LOW DEGENERACY

Plamen A. DEMIREV\* and Alexander S. DYULGEROV

*Institute of Organic Chemistry, Bulgarian Academy of Sciences, Bl. 9, Sofia 1113, Bulgaria*

Ivan P. BANGOV

*Institute of Mathematics, Bulgarian Academy of Sciences, Bl. 8, Sofia 1113, Bulgaria*

Received 18 January 1991; revised 14 April 1991

### Abstract

A new type of topological index is proposed. Its definition is based on the concept of atomic charge distributions in organic molecules. Introduction of electronic in addition to purely topologic factors allows consideration of heteroatom-containing structures as well. It is demonstrated that the index has a low degree of degeneracy, thus suggesting it can be used for coding chemical structures, discrimination of redundancies in structure  $\theta$  generation programs, and studies of quantitative structure – activity relationships for heteroatom-containing structures.

### 1. Introduction

The trend for mathematization of chemistry has brought into the limelight the close relationship between the mathematical notion of *graph* and the chemical notion of *structure*. Formally, it can be defined as follows [1]:

$$g = (\mathcal{V}, \mathcal{E}, f) \Leftrightarrow s = (\mathcal{A}, \mathcal{B}, f), \quad (1)$$

where  $g$  is a graph,  $s$  a chemical structure,  $\mathcal{V}$  is the set of vertices  $v_i$ , corresponding to the set  $\mathcal{A}$  of atoms  $a_i$  in the structure,  $\mathcal{E}$  is the set of edges  $e_i$ , corresponding to the set  $\mathcal{B}$  of chemical bonds  $b_i$ , and  $f$  is an incidence function (mapping) that assigns vertices (atoms) to the edges (bonds). For the case of simple non-oriented graphs, there are additional requirements on the incidence function  $f$  [2]. The set  $\mathcal{E}$  should then be a subset of  $\mathcal{V} \times \mathcal{V}$ , where the Cartesian product  $\mathcal{V} \times \mathcal{V}$  denotes all pairs of vertices  $(v_i, v_j)$ .  $\mathcal{E}$  should also possess the properties of *symmetry* (i.e. from  $(v_i, v_j)$  belonging to  $\mathcal{E}$  follows that  $(v_j, v_i)$  also belongs to  $\mathcal{E}$ ) and *antireflexivity* (if  $(v_i, v_j) \in \mathcal{E}$ , then  $v_i \neq v_j$ ) (see e.g. [2] for different definitions of graphs).

\*Present address: Department of Radiation Sciences, Uppsala University, Box 535, S-75121 Uppsala, Sweden.

A main deficiency of any particular realization of the graph–structure relationship (1), in spite of its clear mathematical formulation, is its non-invariance under permutation of the numbering of vertices. Thus, the adjacency matrix **A** (a symmetric square matrix obeying the following relations:  $A_{ii} = 0$ ,  $A_{ij} = 1$  when two vertices (atoms)  $i$  and  $j$  form an edge (bond), and  $A_{ij} = 0$  otherwise) is dependent on the numbering of the vertices. The same is true for the more elaborate distance matrix **D**, which is of the same order as **A** but whose entries are the minimal distances between vertices  $i$  and  $j$  (vertex relabelling leads in both cases to permutations of rows and/or columns in **A** and **D**). The invariant sets of a graph are defined as sets of numbers that do not depend on the orientation of the graph and its numbering [3]. Different procedures for the construction of graph invariants are outlined in [4]. Some obvious examples of graph invariants include the number of vertices and edges of a graph, the vertex degree, the distance sum of a vertex, etc. [4].

On the other hand, the graph–structure relationship (1) is used for a successful mapping of the set of chemical structures onto the field of real numbers. This is accomplished by constructing the so-called *topological indices* (**TIs**) from different graph-theoretical invariants. The **TIs** can formally be represented by the following relation:

$$\mathbf{TI} = \mathcal{G}(\mathbf{g}), \quad (2)$$

where  $\mathcal{G}$  is an operator, mapping the multidimensional matrix (either **A** or **D**) representation  $\mathbf{g}$  of a graph onto the (one-dimensional) space of the real numbers. A basic requirement is that mapping of graphs  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ , which are isomorphic, should result in  $\mathcal{G}(\mathbf{g}_1) = \mathcal{G}(\mathbf{g}_2)$  [2]. Of course, the opposite assertion need not be valid: equal **TI** may correspond to different graphs (chemical structures) – degeneracy of the **TI**. The so-called "isomorphism disease", which according to [2] is not an uncommon syndrome among beginning practitioners in the field, stems from that requirement for lower degeneracy of the **TIs**. Numerous other requirements on the **TIs** are listed in [5]. The use of **TIs** in different branches of chemistry, including their ability to correlate with various physicochemical and biological properties of different molecules, is extremely well documented. There are several books and numerous reviews published on the subject [4, 6 – 10]. However, despite their ubiquity, **TIs** possess some apparent deficiencies:

*First*, there is no mathematically rigorous proof that each newly defined index is nondegenerate in all cases. The nondegeneracy requirement is also important when **TIs** are used for the description and numerical coding of chemical structures among their other uses. For many practical applications, such as avoiding duplications in the case of computer structure generation, chemical compound nomenclature, documentation and retrieval, structure – property (structure – activity) correlations, **TIs** with the highest possible degree of discrimination are needed.

*Second*, in most cases graphs constructed on the basis of (1), the so-called *chemical* (or *molecular*) *graphs*, are hydrogen depleted, representing only the carbon

atom skeleton structures. They do not contain information on the nature of the atoms and chemical bonds in the corresponding molecule. Treatment of "real" structures without overlooking the hydrogen and/or heteroatoms and multiple bonds is needed in many practical applications, e.g. quantitative structure – activity correlations, computer drug design, etc. That is why new approaches for modification of **TIs** and construction of more elaborate ones have to be advanced.

*Third*, a full characterization of branching in molecular structure, based on **TIs**, has not yet been achieved [3,5,11,12]. The problem of hierarchical ordering of structures and finding criteria for describing molecular complexity is a challenging one, and has still to be resolved.

Starting from the first **TI** introduced almost 150 years ago [13], a great variety of indices has been constructed. There are more than 130 indices proposed up to now. The introduction of new **TIs** is aided in many cases by intuition rather than some reasonable physicochemical argument or mathematical rigour. In 1947, Wiener devised his famous index, which is still in use [14]:

$$W = 1/2 \sum_i \sum_j D_{ij} . \quad (3)$$

However, time has proven that **W** has a high degree of degeneracy; hence, it is not appropriate for dealing with isomorphism (e.g. discrimination of isomers).

One of the more widely exploited **TIs** introduced so far is the molecular connectivity index, proposed by Randić [15]:

$$\chi = 1/2 \sum_i \sum_j A_{ij} \left( \sum_k \sum_l A_{ik} A_{jl} \right)^{-1/2} . \quad (4)$$

It was specifically designed to characterize molecular branching. Despite its lower degeneracy and its potency in correlational studies, this index is also not ultimately universal and applicable for all instances, e.g. for characterizing heteroatom-containing structures.

In a recent paper, Balaban et al. [16] introduced the concept **Local Invariant Sets (LOIS)**, which are comprised of the so-called local indices. The local indices characterize separate vertices of the chemical graph. **LOIS** are further exploited for the construction of global indices, using already known formulations. These global indices, as most of the other indices, are based on hydrogen depleted graphs. However, there is a possibility to characterize different atoms by setting the nondiagonal adjacency matrix entries to different values (e.g. atomic numbers, etc.)

In the present paper, we report an effort to construct a novel topological index which takes into consideration the electronic structure of the molecule as well. To this end, we exploit the **LOIS** concept by introducing local invariants based on the notion of partial atomic charges. This approach allows us to deal with "real"

chemical structures, while also accounting for the presence of hydrogen and/or heteroatoms. Some other applications of the proposed **TI** are discussed as well.

## 2. Construction of local vertex invariants

The definition (2) suggests that in order to construct a **TI**, an appropriate function (with the matrix elements of **A** or **D** as arguments), featuring certain structural properties, should be devised. The great diversity of **TIs** stems from the variety of functions used. A direct procedure for the construction of **TIs** has already been outlined [16]. The first step, the "assignment", consists of defining the **Local Vertex Invariants (LOVIs)**, which comprise the **LOIS**. Balaban et al. [16] obtain the different sets of **LOVIs** as a solution of a system of linear equations that include a matrix derived from the adjacency matrix. The second "operational" step is the computation of a global **TI** as a function of the **LOVIs**. (The terms "*local vertex invariant, local invariant set*" and the acronyms **LOVI**, **LOIS** are borrowed from [16].) This two-step procedure is rather straightforward and general no matter which approach is used for the definition of **LOVIs**.

Bangov has recently introduced a novel **LOVI**, which was termed Atom in Structure Invariant Index (**ASII**) [17 – 19]. It is defined by the following formula:

$$\text{ASII} = \text{ASII}_0 - N_{\text{H}} + Q, \quad (5)$$

where  $\text{ASII}_0$  is an initial value characterizing the type and hybridization state of a given atom  $i$  (table 1).  $N_{\text{H}}$  is the number of attached hydrogen atoms to the atom  $i$ .  $Q$  is the charge density (net atomic charge) of the atom, calculated through some of the fast and comparably reliable empirical methods. We employ the **Iterative Partial Equalization of Orbital Electronegativities (IPEOE)** procedure of Gasteiger et al. [20]. It defines the orbital electronegativity in terms of purely atomic properties – ionization potentials and electron affinities. This procedure provides net atomic charge values, which are different for nonsymmetric atoms and equal for equivalent (symmetric) atoms. The nonoverlapping values of the charge densities enable differentiation of atoms of different types and hybridization. The first term in expression (5) accounts for the type and hybridization state of the atom, while the second term accounts for the attached hydrogen atoms. The connectivity within the structure is estimated by the third term – the net atomic charge. It depends on the entire connectivity of the molecule, although only "through the bond" interactions are taken into account. As long as the **IPEOE** computational scheme involves successively  $\alpha, \beta, \gamma, \dots$  etc. environment, this term might act as an **LOVI** in itself. While the **ASII**s are not obtained via the approach prescribed by Balaban et al. [16] for construction of **LOVIs** (i.e. as a solution of a system of equations), they nevertheless fulfill the conditions for a local graph invariant listed in [16]. The **ASII**s reflect not only topological but more subtle chemical properties of the molecular structure.

Table 1  
Initial Atom-in-Structure Invariant Index ( $ASII_0$ )  
values for atoms in different hybridization states

Atom (hybridization state)	$ASII_0$
C	
$sp^3$	4
sp	7
$sp^2$ (olefinic)	11
$sp^2$ (aromatic)	13
N	
$sp^3$	15
$sp^2$	18
sp	20
O	
$sp^3$	23
$sp^2$	25
S	28
F	32
Cl	33
Br	34
I	35

The  $ASII$ s were employed in the automatic assignment of  $^{13}C$  NMR signals to carbon atoms [17]. Partitioning of carbon atoms into different classes according to their  $^{13}C$ -H multiplicities (in the following order: singlet, doublet, triplet and quartet) and ranking both  $ASII$ s and chemical shifts in each class provides their automatic assignment. In most cases, juxtaposing of the two ranks coincides with the correct assignment.

The  $ASII$ s were also employed for perception of the topological symmetry of the atoms in a given structure and comparing it to the chemical equivalence of the  $^{13}C$  NMR signals [17]. Thus, structures with symmetry that does not correlate with the signal equivalence class are removed. The  $ASII$ s have also been used for partitioning of the vertices into different automorphism classes in a new structure generation approach [19]. In all cases examined, we have confirmed the correlation between  $ASII$ s and the corresponding vertex features.

### 3. Global index construction

To define a global index on the basis of the  $ASII$ s with low or hopefully no degeneracy, an appropriate function has to be introduced in the "operational" step.

Let us form the set  $\mathcal{X}$  of ASII's which we shall denote further as  $x_i, x_j \in \mathcal{X}$ . It is clear from relation (1) that a global graph invariant may be composed of local invariants that account for both vertices and edges. The information for the graph vertices is in our case included in the  $x_i$ . We define each element of  $\mathcal{E}$  by the product:  $x_i x_j$  (i.e. specify the form of the incidence function in (1)). Subsequently, each product is weighted and summed:

$$\text{CrI} = \sum_{i \neq j} w_{ij} x_i x_j. \quad (6)$$

We define the novel global invariant that is obtained as a Charge-related Index (CrI). It is obvious that the CrI is invariant since the sum (6) does not change under permutation of the numbering of the graph.

An explicit realization of CrI was introduced recently [19]. It is the global ASIIg index, defined by choosing the weights  $w_{ij} = w = (1/\sum_l x_l)$ . This index is computed by the following formula:

$$\text{ASIIg} = 10 \left( \sum_{i \neq j} \sum (w x_i x_j) \right)^{1/2}. \quad (7)$$

The global ASIIg index is extremely useful in dealing with isomers. Our experience [19] in applying the ASIIg for elimination of duplications in the case of structure generation has confirmed our belief that it has a very low degree of degeneracy. Based on our experience with the global ASIIg, we suggest a new set of weights for the global CrI. The  $w_{ij}$  are defined as the reciprocal of the corresponding distance matrix  $\mathbf{D}$  elements:  $w_{ij} = 1/\mathbf{D}_{ij}$  ( $i \neq j$ ). Hence, the new index that we propose – Charge-related Topological Index (CTI) – is computed according to:

$$\text{CTI} = \sum_{i \neq j} \sum (x_i x_j / \mathbf{D}_{ij}). \quad (8)$$

#### 4. The charge topological index (CTI) – a highly discriminating topological index

The weight  $w_{ij}$  for pairs  $(x_i, x_j)$  of vertices which form edges is equal to 1 ( $\mathbf{D}_{ij} = 1$  for  $i, j$  labeling adjacent vertices  $v_i, v_j$ ). It is always greater than the weights for pairs of vertices not forming an edge. Hence, for a set of isomers the different connectivities are expressed both by different terms  $\mathbf{Q}$  in (5) and by different weights  $1/\mathbf{D}_{ij}$  in (8). We suggest that the CTI has a high discriminating power and a very low degree of degeneracy. Evidently this conjecture cannot be rigorously proven (the same applies to the global ASIIg), but we will be indebted if any counterexamples are found. An illustration of the discrimination power of the CTI is given in fig. 1,

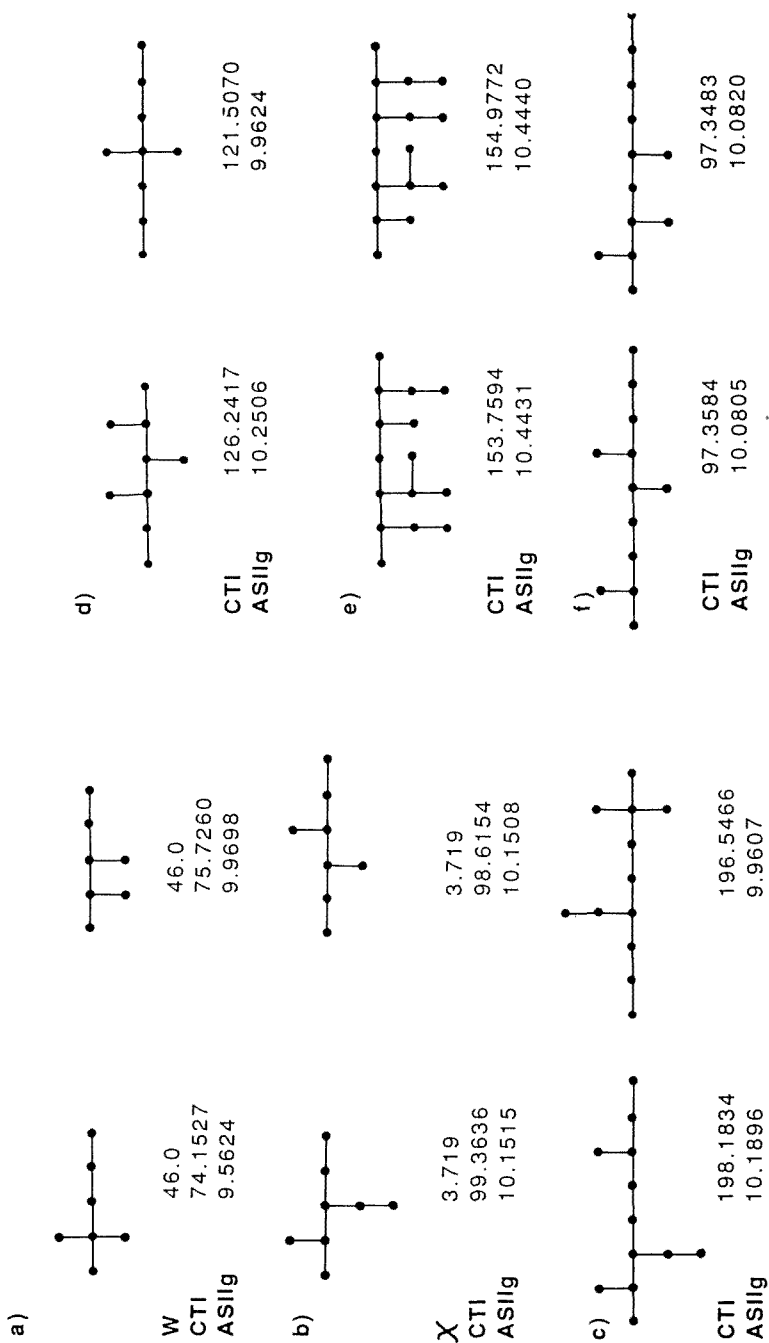


Fig. 1. Pairs of nonisomorphic graphs of alkanes, displaying some degeneracies (identical topological indexes), but different values for the CTI and the ASIlg: (a) same Wiener index (W), (b) same Randić index ( $\chi$ ), (c) same Balaban index [21], (d) same isospectral index [3], (e) same ID number [22], (f) operationally degenerate isomers for some of the TI, introduced by Balaban [16].

where the **CTI** values for different pairs of alkane isomers are presented. All these isomer pairs show some degeneracy for different other **TIs**, but the **CTI** is different for all of them. The values of the **CTI** and **ASIIg** for the series of lower alkanes are listed in table 2. Eighteen different hydrocarbon structures, containing four carbon

Table 2  
Charge-related topological index (**CTI**) and **ASIIg**, calculated for the series of lower alkanes ( $n = 2-7$ )

<i>n</i>	Alkane	<b>CTI</b>	<b>ASIIg</b>
2	ethane	0.8682	7.0711
3	propane	4.0641	8.1109
4	n-butane	9.5193	8.9255
4	2-methylpropane	9.6103	8.5526
5	n-pentane	16.3392	9.2492
5	2-methylbutane	17.3298	9.3031
5	2, 2-dimethylpropane	17.5292	8.8020
6	n-hexane	24.1502	9.4230
6	2-methylpentane	25.5174	9.4480
6	3-methylpentane	26.4147	9.7261
6	2, 3-dimethylbutane	27.4097	9.7179
6	2, 2-dimethylbutane	27.5181	9.4863
7	n-heptane	32.7421	9.5314
7	2-methylhexane	34.3215	9.5424
7	3-methylhexane	35.5936	9.7733
7	2, 4-dimethylpentane	36.0662	9.5524
7	3-ethylpentane	36.8652	9.9987
7	2, 2-dimethylpentane	37.0763	9.5624
7	2, 3-dimethylpentane	37.8630	9.9698
7	3, 3-dimethylpentane	38.8731	9.9482
7	2, 2, 3-trimethylbutane	39.8726	9.9269

atoms (saturated and unsaturated; acyclic as well as cyclic) are shown in fig. 2, together with their corresponding **CTI**. The **CTI** and the **ASIIg** for cyclic structures, displaying assignment degeneracy for various **LOIS** [16], are presented in fig. 3. The enthalpies of formation of  $C_2-C_8$  alkanes [23] versus their respective **CTI** are plotted in fig. 4 in order to illustrate the potential applicability of the **CTI** for structure – property correlations. Another illustration (fig. 5) includes the correlation between the motor octane numbers of the  $C_4-C_9$  normal alkanes [16] and their **CTI**.













CTI	ASIIg	CTI	ASIIg
9.5193	8.9255		126.9828
9.6103	8.5526		160.6609
18.9505	8.9605		161.7319
19.9265	9.2523		165.9374
67.2720	8.0408		169.9367
68.4782	7.9465		183.7697
93.4664	8.7732		390.1898
99.9763	9.3839		493.8174
125.5936	7.7530		9.9235
126.8092	7.6859		

Fig. 2. Values of the CTI and the ASIIg for different  $C_4$ -atom structures.

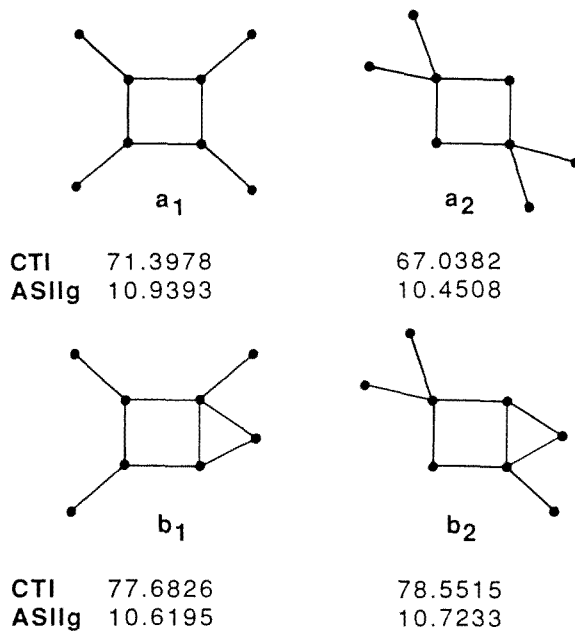


Fig. 3. CTI and ASIIg for cyclic graphs, displaying assignment degeneracy for various LOIS [16] (structures  $b_1$  and  $b_2$  have also the same distance sum connectivity index).

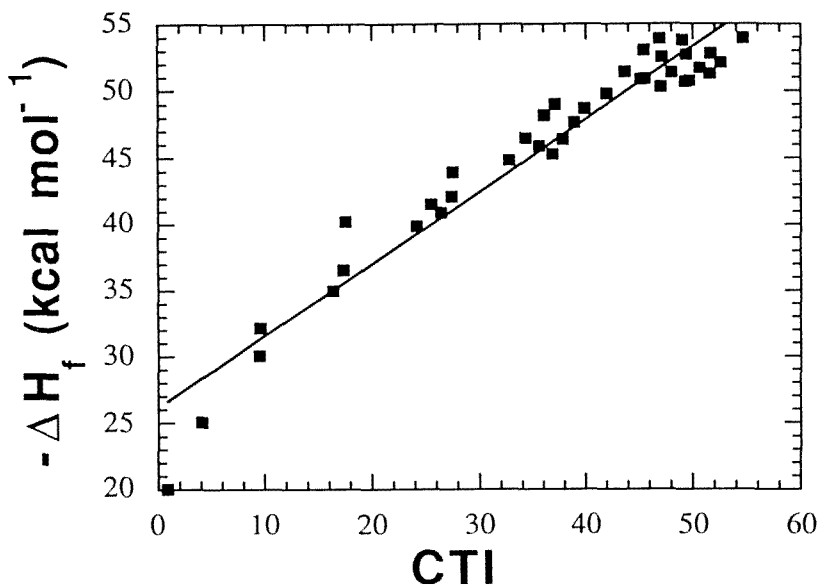


Fig. 4. Correlation ( $R = 0.968$ ) between the enthalpies of formation ( $-\Delta H_f$ , data taken from [23]) of the  $C_2$ - $C_8$  alkanes and their CTI (values listed in tables 2 and 4).

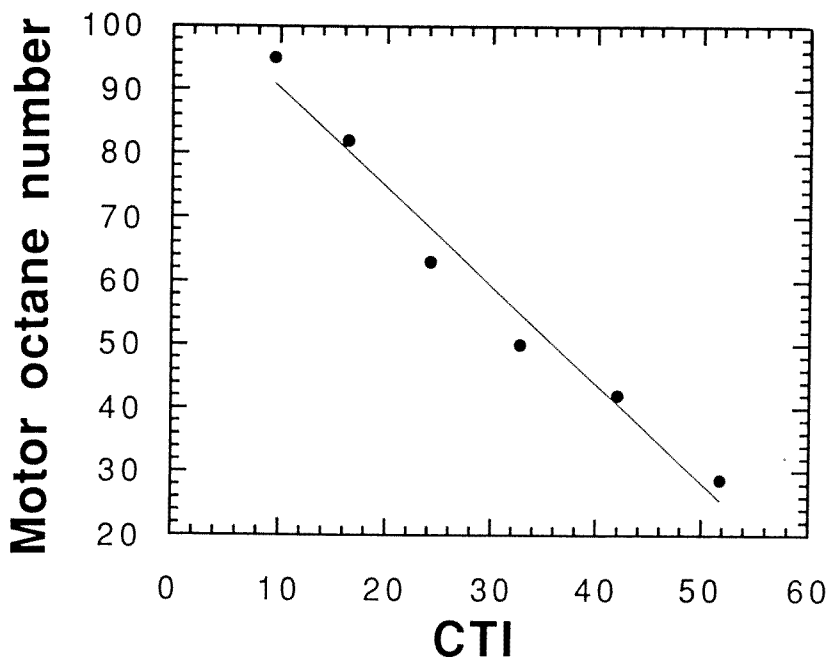


Fig. 5. Correlation ( $R = 0.986$ ) between the motor octane number [16] of the normal  $C_4$ – $C_9$  alkanes and their CTI (CTI of *n*-nonane 51.7677, other values listed in tables 2 and 4).

## 5. CTI for heteroatom-containing structures

The CTI can be assigned to structures containing heteroatoms as well. The discrimination of heteroatoms in the chemical graphs has always been a drawback in the construction of TIs, although several approaches to counteract this difficulty have been forwarded [2, 24, 25] (see also references therein for other approaches). Their essence consists of devising recipes for modification of the distance matrix **D** for the sake of including multiple bonds as well as heteroatoms. Balaban has modified the averaged distance sum connectivity index **J** [21] in a way that takes into account the relative electronegativities or the relative covalent radii of different heteroatoms as compared to the C atom [25]. In the present approach, introduction of heteroatoms is naturally accomplished through the local ASII. An illustration of the application of CTI for heteroatomic systems (different isomeric amines –  $C_4H_{11}N$ ) is presented in fig. 6. The dependence of the CTI values on different parameters is complex and thus does not allow the extraction of any clear-cut correlations. One may note that CTI increases from primary to tertiary amines. It is, however, evident that the type of C atom(s) connected to the N also plays a significant role. The CTI has also been calculated for the series of lower alcohols and the data are presented in table 3. As in the case with amines, the CTI values depend on several different


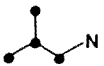

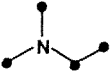
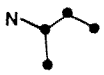
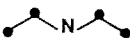
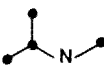

	CTI	ASIIg
	61.8384	4.4838
	65.7839	4.5823
	66.9244	4.8600
	70.2266	5.1993
	73.6155	5.2666
	74.7375	5.4408
	75.7826	5.5016
	81.7102	5.8898

Fig. 6. Charge-related topological index (CTI) and ASIIg, calculated for different isomeric amines –  $C_4H_{11}N$ .

Table 3

Charge-related topological index (CTI) and ASIIg, calculated for the series of lower alcohols ( $n = 2-5$ )

No.	$n$	Alcohol	CTI	ASIIg
1	2	ethanol	56.3998	3.1243
2	3	propane-3-ol	78.9678	3.2428
3	3	propane-2-ol	92.9637	3.8743
4	4	butane-4-ol	97.7223	3.3510
5	4	2-methylpropane-3-ol	103.9319	3.4069
6	4	butane-3-ol	117.8204	3.9906
7	4	2-methylpropane-2-ol	132.0219	4.5216
8	5	pentane-5-ol	114.7124	3.4534
9	5	2-methylbutane-4-ol	118.8611	3.5034
10	5	2-methylbutane-1-ol	124.9558	3.5375
11	5	2, 2-dimethylpropane-3-ol	131.3143	3.6072
12	5	pentane-4-ol	137.9558	4.0729
13	5	pentane-3-ol	144.0450	4.1020
14	5	2-methylbutane-3-ol	145.0783	4.1408
15	5	2-methylbutane-2-ol	159.1723	4.6379

factors. The CTI increases both with branching and in the direction from primary through secondary to tertiary alcohols. One may argue that the dependence on the second factor is more pronounced, since the n-pentane-5-ol precedes s-butanol and t-butanol follows after all primary pentanols.

In the present study, we do not aim to correlate the CTI with any physicochemical parameter of the corresponding structure (similar correlations are currently being investigated and will be the subject of a future account). We would like to speculate, though, that the physical meaning behind the CTI might be connected to the sum of "through the bond" Coulomb interaction energy terms between the pairs of different atoms in the structure as following from eq. (8).

## 6. CTI and the branching of chemical structures

The concept of branching in chemical structures is not a new concept [5, 15]. It has always been linked to the more profound idea of molecular complexity in general. It is intuitively clear that a rough measure of branching in a molecular structure may be provided by the number of its branching points. Thus, in the case of, for example, isomeric alkanes, the two extremes should be the normal (linear) alkanes – the non-branched species – and the alkanes with maximum number of quaternary carbon atoms – the most branched species. Despite this, there has been a long-standing argument as to how molecular branching should be defined, even when relatively simple structures are considered. The problems and challenges in characterizing molecular branching and molecular complexity have been extensively reviewed [11, 12]. A way to circumvent the difficulties arising in finding an appropriate definition of branching is to try to characterize it in quantitative rather than in qualitative terms. Therefore, different mathematical criteria for comparing and hierarchical ordering of (mainly acyclic alkane) structures have been introduced [26]. The use of Young diagram techniques, random walk on trees and characteristic graph polynomials should be noted among other mathematical approaches [12]. Different TIs have also been employed for quantitative description of branching in molecular structures, since TIs reflect both the size and the shape of a molecule [12]. Bertz introduced a new mathematical model of chemical complexity through the notion of a graph derivative [11]. While he criticizes the TI approach for the ordering of graphs and molecular structures as a "force-fit" one, he argues that the values of a "good" TI should increase with the increase in molecular branching. He also notes that J (the averaged distance sum connectivity index) parallels the ordering of the isomeric pentanes, hexanes, heptanes and partially the octanes, obtained using the graph derivative approach [11].

It appears that branching is a purely topological property, although the particular arrangement of atoms in a molecule influences the intramolecular interactions and its electronic configuration on the whole. Since a combination of topologic and electronic factors is taken into account when computing the CTI, we tried to use

Table 4

Charge-related topological index (CTI) and  $ASIIg$ , calculated for the series of octanes (the order of different isomers, according to the branching index of Bertz [11], is given in brackets)

Compound	CTI	$ASIIg$
octane (1)	41.9796	9.6055
2-methylheptane (2)	43.6962	9.6092
3-methylheptane (4)*	45.1788	9.8061
2, 5-dimethylhexane (3)	45.4882	9.6124
4-methylheptane (5)	45.5535	9.8070
2, 2-dimethylhexane (12)*	46.8757	9.6178
3-ethylhexane (7)*	47.0352	9.9999
2, 4-dimethylhexane (6)	47.1357	9.7960
2, 3-dimethylhexane (8)	48.0352	9.9747
2, 2, 4-trimethylpentane (13)*	48.9985	9.6214
3, 4-dimethylhexane (9)	49.3077	10.1508
3, 3-dimethylhexane (14)*	49.4228	9.9561
2-methyl-3-ethylpentane (10)	49.6818	10.1515
2, 3, 4-trimethylpentane (11)	50.6825	10.1183
3-ethyl-3-methylpentane (15)	51.5943	10.2822
2, 2, 3-trimethylpentane (16)	51.6971	10.0916
2, 3, 3-trimethylpentane (17)	52.5966	10.2434
2, 2, 3, 3-tetramethylbutane (18)	54.6152	10.1806

it for ordering of chemical structures. The results for the case of the  $C_2$ – $C_7$  alkanes are presented in table 2, and those for the octanes in table 4. We note, in comparing the CTI ordering to the "logically derived order of branching" advanced by Bertz [11], that the lower alkanes up to  $C_6$  are following the same ordering. Only one of the eleven heptanes, namely 2, 4-dimethylpentane, does not fit into Bertz's classification scheme. Six out of the eighteen octanes do not parallel the ordering suggested by Bertz, which is the same number when compared to the Balaban index  $J$  [11].

## 7. Program

The recipe for the calculation of the CTI has been implemented in a computer code for an IBM compatible PC. To this end, a previously created program for structure generation and automatic assignment of  $^{13}C$  NMR chemical shifts [17 – 19] has been modified in order to incorporate the above-described algorithm. The procedures for coding and chemical structure input are detailed elsewhere [17]. The adjacency ( $A$ ) and the distance ( $D$ ) matrices are then constructed, which allows calculation of the Wiener ( $W$ ) and Randić ( $R$ ) topological indices and the  $ASIIg$  as well. The program is written in Turbo PASCAL (Borland International Inc., v.5.0). It is user-oriented and in menu-driven interactive format. Possibilities for using other empirical methods

for calculation of the charge densities, such as the method of Abraham et al. [27], are also built into the program.

## 8. Conclusion

A new type of topological index, termed the Charge Topological Index, has been proposed. Its definition, based on the concept of atomic charge distributions in organic molecules, follows a recipe suggested by Balaban [16] for the introduction of global TIs. The computational procedure is more complicated, including calculations of quantities (e.g. charges) which are not directly connected to purely molecular topological notions. The requirement for computer calculation of the CTI is not a serious impediment to its use, since in almost all applications of TIs nowadays the personal computer is a standard tool employed by chemists. The introduction of purely atomic factors in addition to the topological ones allows consideration of heteroatom-containing structures. It has been demonstrated that the CTI has a low degree of degeneracy as well. Thus, we suggest that the CTI is not a mere addition to the plethora of different TIs, but that it can find some real practical applications. These may include, among others, coding of chemical structures, discrimination of redundancies in structure  $\theta$  generation programs, and studies of quantitative structure – activity relationships for heteroatom-containing structures in, for example, computer drug design.

## Acknowledgements

One of us (P.A.D.) acknowledges the Swedish Natural Sciences Research Council (NFR) for partial financial support.

## References

- [1] W. Gellert, H. Kuestner, M. Hellwich and H. Kaestner (eds.), *Mathematics at a Glance. A Compendium* (VEB Bibliogr. Inst. Leipzig, 1975).
- [2] I. Gutman and O. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer, Berlin, 1986).
- [3] D.H. Rouvray, *J. Mol. Struct.* 185(1989)187.
- [4] M.I. Stankevich, I.V. Stankevich and N.S. Zefirov, *Usp. Khimii* 57(1988)337
- [5] A.T. Balaban, *J. Mol. Struct. (THEOCHEM)* 165(1988)243.
- [6] A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, *Top. Curr. Chem.* 114(1983)21.
- [7] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Chichester, 1983).
- [8] N. Trinajstić, *Chemical Graph Theory*, Vol. 2 (CRC Press, Boca Raton, FL, 1983), ch. 4, p. 105.
- [9] R.B. King (ed.), *Chemical Applications of Topology and Graph Theory* (Elsevier, Amsterdam, 1983).
- [10] R.B. King and D.H. Rouvray (eds.), *Graph Theory and Topology in Chemistry* (Elsevier, Amsterdam, 1987).
- [11] St. Bertz, *Discr. Appl. Math.* 19(1988)65.
- [12] D.H. Rouvray, *Discr. Appl. Math.* 19(1988)317.
- [13] H. Kopp, *Ann. Chem. Pharm.* 50(1844)71.

- [14] H. Wiener, J. Amer. Chem. Soc. 69(1947)2636.
- [15] M. Randić, J. Amer. Chem. Soc. 97(1975)171.
- [16] P.A. Filip, T.S. Balaban and A.T. Balaban, J. Math. Chem. 1(1987)61.
- [17] I.P. Bangov, Anal. Chim. Acta 209(1988)29.
- [18] I.P. Bangov, Commun. Dept. Chem. Bulg. Acad. Sci. 21(1988)194.
- [19] I.P. Bangov, J. Chem. Inf. Comput. Sci. 30(1990)277.
- [20] J. Gasteiger and M. Marsili, Tetrahedron 36(1980)3219;  
M. Marsili and J. Gasteiger, in: *Data Processing in Chemistry*, Studies in Physical and Theoretical Chemistry Series 16 (Elsevier – PWN, Amsterdam – Warsaw, 1981), p. 56.
- [21] A.T. Balaban, Chem. Phys. Lett. 89(1982)399.
- [22] M. Randić, J. Chem. Inf. Comput. Sci. 28(1988)142.
- [23] D. Scott, J. Chem. Phys. 60(1974)3144.
- [24] M. Barysz, G. Jashavi, R. Lall, V. Srivastava and N. Trinajstić, in: ref. [9], p. 222.
- [25] A.T. Balaban, Math. Chem. 21(1986)115.
- [26] I. Gutman and M. Randić, Chem. Phys. Lett. 47(1977)15.
- [27] R.J. Abraham, L. Griffiths and P. Loftus, J. Comput. Chem. 3(1982)407.