

## Upgrading the Wiener index

EDUARDO A. CASTRO,<sup>a</sup> IVAN GUTMAN,<sup>b#</sup> DAMIAN MARINO<sup>a</sup> and  
PABLO PERUZZO<sup>a</sup>

<sup>a</sup>CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, C. C. 962, La Plata 1900, Argentina and <sup>b</sup>Faculty of Science, University of Kragujevac, P. O. Box 60, YU-34000 Kragujevac, Yugoslavia

(Received 18 June 2002)

**Abstract:** The Wiener index  $W$  is the oldest molecular-graph-based structure-descriptor. It is defined as the sum of the distances of all pairs of vertices of the molecular graph  $G$ , where the distance is the number of edges in the shortest path connecting the respective vertices, and where  $G$  is the hydrogen-depleted molecular graph. This seemingly very simple topological index could be "upgraded" (*a*) by using as the distance the sum of the bond lengths along the shortest path, or (*b*) by using the Euclidean distance between the respective pairs of atoms. Each of these "upgraded" Wiener indices could be computed either ( $\alpha$ ) for the hydrogen-depleted or ( $\beta$ ) for the hydrogen-filled molecular graph. We provide examples showing that none of the modifications ( $\alpha\alpha$ ), ( $\alpha\beta$ ), ( $\beta\alpha$ ), ( $\beta\beta$ ) yields better results than the ordinary Wiener index, and that there is a very good linear correlation between  $W$  and its "upgraded" variants.

**Keywords:** Wiener index, topological index, 3D-structure descriptors, chemical graph theory, QSPR, QSAR.

### INTRODUCTION

The Wiener index<sup>1</sup> ( $W$ ) is the oldest molecular-graph-based structure-descriptor<sup>2</sup> and its chemical applications<sup>3–5</sup> and mathematical properties<sup>6,7</sup> are well documented. The usual explanation as to why  $W$  is so successful in QSPR and QSAR studies is based on the fact that  $W$  represents a rough measure of the van der Waals molecular surface area.<sup>8</sup> Consequently, in the case of non-polar molecules (such as alkanes<sup>3</sup> and cycloalkanes<sup>5</sup>),  $W$  is proportional to the intermolecular forces,<sup>4,9</sup> and is thus related to a number of physico-chemical properties of the respective compounds (boiling point, heat of evaporation, heat of formation, chromatographic retention times, surface tension, vapor pressure, partition coefficients, *etc.*). The applicability of the Wiener index for the description of physico-chemical and pharmacologic properties of compounds very different from saturated hydrocarbons (*e.g.*, fungicidal derivatives of salicylhydroxamic acid,<sup>10</sup> antiviral 5-vinyl-

# Serbian Chemical Society active member

pyrimidine nucleoside analogs,<sup>11</sup> mixed amino acid complexes with copper (II)<sup>12,13</sup>) motivated us to examine variants of  $W$  in which the structural features of such non-hydrocarbon compounds could be incorporated.

#### "UPGRADING" THE WIENER INDEX

The original definition of the Wiener index is as follows.<sup>1,14</sup> Let  $G$  be a molecular graph<sup>14</sup> and  $u$  and  $v$  two of its vertices. Denote by  $d(u,v|G)$  the distance between  $u$  and  $v$  (= number of edges in the shortest path connecting  $u$  and  $v$ ). Then the Wiener index is

$$W = W(G) = \sum_{u,v} d(u,v|G) \quad (1)$$

with the summation going over all pairs of vertices of  $G$ . The molecular graph  $G$  used in formula (1) is hydrogen-depleted, *i.e.*, its vertices represent only carbon (and perhaps some other heavy) atoms, but not hydrogens.

The calculation of  $W$  is rather simple, either directly by means of formula (1) or by using some of the numerous shortcuts, applicable to particular classes of molecular graphs.<sup>6,7,14-16</sup> Needless to say, the distances  $d(u,v|G)$  in formula (1) are integers, and therefore  $W$  is also integer-valued. The obvious drawback of  $W$ , as defined *via* (1), is that it is insensitive to the existence and position of double and triple bonds as well as of heteroatoms.

Some possible modifications of formula (1), aimed at bringing the Wiener index closer to real-world molecules, come readily to mind:

(a) Structure-descriptors  $Wb$  and  $WbH$ : instead of the graph-theoretical distance  $d(u,v|G)$  – which is just the number of chemical bonds between the atoms  $u$  and  $v$  – one could use the sum of the bond lengths along the shortest path, provided the geometry of the molecule is known. This geometry needs to be computed by means of a suitable molecular-mechanics or molecular-orbital method. By means of  $Wb$  and  $WbH$ , single, double and triple bonds will be clearly distinguished, and their position in the molecule taken into account.

(b) Structure-descriptors  $We$  and  $WeH$ : Instead of  $d(u,v|G)$  one could use the Euclidean distance between the atoms  $u$  and  $v$ . Again, the molecular geometry must be known. The possible advantage of these indices would be that through-space, and not through-bond interactions would be accounted for.

( $\alpha$ ) Usually the molecular graph used for the computation of the Wiener index and its modifications is the hydrogen-depleted molecular graph, in which vertices represent only the heavy atoms (carbon and heteroatoms). This choice would result in the indices  $W$ ,  $Wb$  and  $We$ .

( $\beta$ ) If the Wiener index and its modifications would be calculated on the basis of the hydrogen-filled molecular graph, then instead of  $W$ ,  $Wb$  and  $We$ , one would obtain  $WH$ ,  $WbH$  and  $WeH$ , respectively.

The structure-descriptor  $We$  was studied in the past;<sup>17,18</sup> it was referred to as the three-dimensional (or 3D) Wiener index. In the case of acyclic molecules, the quantities  $W$  and  $WH$  were shown to be related by means of an exact mathematical relation:<sup>19</sup>

$$WH = 9W + (3n + 1)^2 \quad (2)$$

where  $n$  is the number of carbon atoms. To the authors' best knowledge, the other, above mentioned, variants of the Wiener index have so far not been considered in the chemical literature.

NUMERICAL WORK 1: THE SIX VARIANTS OF THE WIENER INDEX ARE LINEARLY CORRELATED

The fact that  $WH$  and  $W$  are linearly correlated is seen from Eq. (2). Therefore the relation between these two structure-descriptors was not examined any further.

In order to test the quality, predictive power and the differences between the remaining five variants of the Wiener index ( $W$ ,  $Wb$ ,  $WbH$ ,  $We$ ,  $WeH$ ), two data bases were used: a set of 39 acyclic monohydroxy alcohols ( $C_1$  to  $C_7$ ) from the paper<sup>20</sup> and a set of 25 acyclic alkenes from the paper.<sup>21</sup>

The optimized geometries of the respective molecules were determined by means of the program package Hyper Chem, employing the semiempirical AM1 molecular orbital method. The quantities  $Wb$ ,  $WbH$ ,  $We$ ,  $WeH$  were calculated from the appropriately modified formula (1), into which bond lengths and interatomic distances were substituted. All bond length and interatomic distances were in picometers. Consequently,  $Wb$ ,  $WbH$ ,  $We$ ,  $WeH$  are also expressed in pm-units (whereas  $W$  and  $WH$  are dimensionless quantities).

It was found that the original Wiener index  $W$  and its modifications are highly correlated and that these correlations are linear. For alcohols<sup>20</sup> the respective regression lines and their correlation coefficients  $R$  read:

$$\begin{array}{ll} Wb = 150.29 W + 35.38 & R = 0.9999 \\ We = 124.43 W - 83.02 & R = 0.9994 \\ WeH = 1233.3 W + 9168.9 & R = 0.9984 \end{array}$$

whereas for alkenes<sup>21</sup> the following values were obtained:

$$\begin{array}{ll} Wb = 156.25 W + 374.83 & R = 0.9917 \\ We = 117.3 W + 176.4 & R = 0.9966 \\ WeH = 1302.7 W + 7369.8 & R = 0.9976 \end{array}$$

In view of relation (2) and the very good linear correlation between  $W$  and  $Wb$ , there was no need to analyze the correlation between  $W$  and  $WbH$ , which must also be linear and very strong.

The  $R$ -values of the above regressions are all very close to unity, implying that a strong linear correlation exists between all the six variants of the Wiener index. In other words: all the "upgraded" versions of the Wiener index bear essentially the same information on molecular structure and none is significantly better than  $W$  itself.

## NUMERICAL WORK 2: STRUCTURE-PROPERTY RELATIONS BASED ON THE VARIANTS OF THE WIENER INDEX

In view of the analysis outlined in the preceding section one must expect that the QSPR relations based on the modified versions of  $W$  will be of similar quality as those based on  $W$  and that "upgrading" the Wiener index will yield little or no gain. This indeed was found by examining the relation between the various Wiener indices and the boiling points of alcohols, as well as the molar refractivities of alkenes.

For the linear correlation between the boiling points of alcohols and  $W$ ,  $Wb$ ,  $We$  and  $WeH$  the correlation coefficients are equal to 0.9367, 0.9375, 0.9349, and 0.9372, respectively. For the linear correlation between the molar refractivity of alkenes and  $W$ ,  $Wb$ ,  $We$ , and  $WeH$  the correlation coefficients are equal to 0.9779, 0.9812, 0.9652, and 0.9822, respectively. In both cases the increase of accuracy of the correlation when  $W$  is exchanged by  $Wb$ ,  $We$ , or  $WeH$  is statistically insignificant and negligible.

## CONCLUDING REMARKS

It has been shown that some, seemingly reasonable and theoretically justifiable, modifications of the Wiener index do not improve at all the applicability and value of this structure-descriptor in designing quantitative structure-property relations. This may be viewed as a somewhat surprising and somewhat disappointing finding. From another point of view, the same finding implies that the original Wiener index – now already half-a-century old – is a much more valuable topological index than one would expect from its extremely simple and seemingly naive definition, Eq. (1).

From a practitioner's point of view the dilemma:  $W$  or some of its "upgraded" versions  $Wb$ ,  $WbH$ ,  $We$ ,  $WeH$  has now been completely resolved. In QSPR and QSAR studies the modified Wiener indices will give no better results than the original Wiener index. The calculation of these modified Wiener indices is difficult, requiring the knowledge of molecular geometry. The calculation of the original Wiener index  $W$  is simple. Therefore, unless some convincing argument in favor of a modified Wiener index is offered, preference should be given to its original version, Eq. (1).

## ИЗВОД

## УСАВРШАВАЊЕ ВИНЕРОВОГ ИНДЕКСА

ЕДУАРДО А. КАСТРО, ИВАН ГУТМАН, ДАМИАН МАРИНО и ПАБЛО ПЕРУЦО

*Хемијски одсек Факултета ефикасних наука, Национални Универзитет Ла Плаја, Ла Плаја, Аргентина и  
Природно-математички факултет у Крајујевцу*

Винеров индекс  $W$  је најстарији, на молекулском графу засновани, структурни дескриптор. Дефинисан је као збир растојања између свих парова чворова молекулског графа  $G$ . При томе, растојање је једнако броју грана које леже на најкраћем путу који повезује одговарајуће чворове, а  $G$  је "безводонични" молекулски граф. Овај наизглед веома једноставан тополошки индекс могао би се "усавршити" тако што би се као растојање узимало (а) збир дужина веза дуж најкраћег пута, или (б) Еуклидско растојање између одговарајућих парова атома. Сваки од овако "усавршених" Винерових индекса могао би се рачунати на (α) "безводони-

чном" молекулском графу, или  $(\beta)$  на "водоничном" молекулском графу. Ми наводимо примере који показују да ни једна од модификација  $(a\alpha)$ ,  $(a\beta)$ ,  $(b\alpha)$ ,  $(b\beta)$  не даје боље резултате него обични Винеров индекс, као и да постоји добра линеарна корелација између  $W$  и његових побољшаних варијанти.

(Примљено 18. јуна 2002)

## REFERENCES

1. H. Wiener, *J. Am. Chem. Soc.* **69** (1947) 17
2. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim 2000
3. D. E. Needham, I. C. Wei, P. G. Seybold, *J. Am. Chem. Soc.* **110** (1988) 4186
4. I. Gutman, J. H. Potgieter, *J. Serb. Chem. Soc.* **62** (1997) 185
5. G. Rücker, C. Rücker, *J. Chem. Inf. Comput. Sci.* **39** (1999) 788
6. A. A. Dobrynin, R. Entringer, I. Gutman, *Acta Appl. Math.* **66** (2001) 211
7. A. A. Dobrynin, I. Gutman, S. Klavžar, P. Žigert, *Acta Appl. Math.*, **72** (2002) 247
8. I. Gutman, T. Körtvélyesi, *Z. Naturforsch.* **50a** (1995) 669
9. I. Gutman, J. H. Potgieter, *South Afr. J. Sci.* **92** (1996) 47
10. P. V. Khadikar, S. Karmarkar, S. Joshi, I. Gutman, *J. Serb. Chem. Soc.* **61** (1996) 89
11. S. Mendiratta, A. K. Madan, *J. Chem. Inf. Comput. Sci.* **34** (1994) 867
12. S. Nikolić, N. Raos, *Croat. Chem. Acta* **74** (2001) 621
13. N. Raos, *Croat. Chem. Acta* **75** (2002) 117
14. I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986
15. I. Gutman, *J. Mol. Struct. (Theochem)* **285** (1993) 137
16. I. Gutman, *J. Serb. Chem. Soc.* **58** (1993) 745
17. B. Bogdanov, S. Nikolić, N. Trinajstić, *J. Math. Chem.* **3** (1989) 299
18. M. Randić, B. Jerman-Blažić, N. Trinajstić, *Comput. Chem.* **14** (1990) 237
19. I. Gutman, D. Vidović, L. Popović, *J. Chem. Soc. Faraday Trans.* **94** (1998) 857
20. M. Randić, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **39** (1999) 261
21. M. Randić, M. Pompe, *SAR QSAR Environ. Res.* **10** (1999) 451.