MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **53** (2005) 215-224

ISSN 0340 - 6253

THE GRAPH OF ATOMIC ORBITALS AND ITS BASIC PROPERTIES. 1. WIENER INDEX

Ivan Gutman,[a] Andrey A. Toropov[b] and Alla P. Toropova

[a] *Faculty of Science, University of Kragujevac, Serbia & Montenegro*
e-mail: gutman@knez.uis.kg.ac.yu

[b] *Scientific Research Institute "Algorithm-Engineering", 700125, F. Khodjaev street 25, Tashkent, Uzbekistan*
e-mail: aa_toropov@yahoo.com

(Received June 5, 2004)

Abstract

The graph of atomic orbitals (GAO) is a novel type of molecular graphs, recently put forward by two of the present authors. The definition of GAO is re-stated in precise graph theoretic terms. A connection between the Wiener index of GAO and the Wiener index of the ordinary molecular graphs, both hydrogen-depleted and hydrogen-filled, is established.

INTRODUCTION

Most researches in chemical graph theory are done on, or by means of, molecular graphs pertaining to covalent organic compounds [1–4]. Many different types of molecular graphs were defined and examined. Of these, the graph in which heavy (usually carbon) atoms are represented by vertices, whereas hydrogen atoms are ignored, is employed in the vast majority of cases. Such graphs are simply called "*molecular graphs*". If, however, we want to distinguish them from other types of molecular graphs, then the former are referred to as "*hydrogen-depleted molecular graphs*". The "*hydrogen-filled molecular graphs*" contain vertices representing hydrogen atoms.[1]

In Fig. 1 are depicted the hydrogen-depleted (G_C) and the hydrogen-filled (G_H) molecular graphs of methyl-isopropylether.

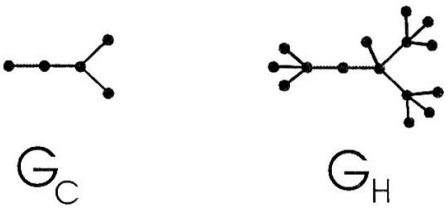

Fig. 1. Molecular graphs of methyl-isopropylether, $CH_3OCH(CH_3)_2$. Diagram G_C stands for the hydrogen-depleted, whereas G_H for the hydrogen-filled molecular graph.

In a series of recently published articles [6–15], two of the present authors introduced a novel type of molecular graphs, the "*graph of atomic orbitals*", GAO. All the papers [6–15] were application-oriented and reported various QSPR and QSAR

[1] Arthur Cayley was the first to define and consider these two types of molecular graphs [5]. He named them "*kenograms*" and "*plerograms*". Recall that in 1874, when the paper [5] was published, the name "*graph*" was not yet coined.

studies, in which pertinently chosen invariants of the GAO were employed. In all these papers the vertices of the GAO were weighted by adjustable parameters, reflecting the nature of the group of atomic orbitals that this vertex represented. In view of this, and in view of the terminology used in [6–15], the peculiar and interesting graph-theoretic features of GAOs could have been overlooked by the rest of chemical-graph-theoretical community. The aim of this article is to emphasize the genuine graph-theoretic nature of GAOs, and to try to establish some of their general graph-theoretic properties.

Throughout this paper we consider the GAOs as simple (schlicht) graphs, that is as graphs without weighted and directed edges, without weighted vertices and without self-loops. Their definition is described in the subsequent section.

DEFINITION OF GAO

In both the hydrogen-depleted and hydrogen-filled molecular graphs, vertices represent individual atoms. The idea behind the GAO is to represent by a vertex a group of atomic orbitals of the respective atom. These groups of atomic orbitals are the following:

atom	groups of atomic orbitals	n_i
H	$1s^1$	1
C	$1s^2\ 2s^2\ 2p^2$	3
N	$1s^2\ 2s^2\ 2p^3$	3
O	$1s^2\ 2s^2\ 2p^4$	3
F	$1s^2\ 2s^2\ 2p^5$	3
S	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^4$	5
Cl	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$	5
Br	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^5$	8
I	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 5s^2\ 5p^5$	11

Definition 1. Let M be a molecule and G_H its hydrogen-filled molecular graph. Let $V(G_H) = \{v_1, v_2, \ldots, v_N\}$ be the vertex set of G_H. The *graph of atomic orbitals* (GAO), pertaining to M, is obtained from G_H by replacing each of its vertex v_i, $i = 1, 2, \ldots, N$, by a set $\mathcal{V}_i$, consisting of n_i distinct vertices, where the value of n_i depends on the type of atom represented by vertex v_i, as specified in the above table.

A vertex of GAO belonging to $\mathcal{V}_i$ is adjacent to a vertex belonging to $\mathcal{V}_j$ if and only if the vertices v_i and v_j of G_H are adjacent.

An immediate consequence of Definition 1 is that two vertices of a GAO, belonging to the same set $\mathcal{V}_i$, are never adjacent.

In Fig. 2 is depicted the GAO of methyl-isopropylether.

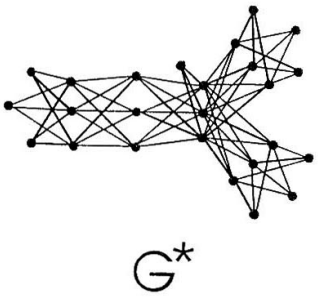

Fig. 2. The graph of atomic orbitals of methyl-isopropylether, cf. Fig. 1.

Comparing Figs. 1 and 2 we see that the structure of a GAO appears to be much more complicated than the structure of an ordinary molecular graph. Yet, as demonstrated elsewhere [6–15], the GAO renders a suitable basis for QSPR and QSAR studies (of course, provided its vertices are appropriately weighted).

In what follows we determine certain mathematical properties of GAOs that are independent of any weighting of their vertices.

DISTANCES IN GAO

If G is a connected graph and x and y are its vertices, then the distance between x and y, denoted by $d(x, y|G)$ is the length of (= number of edges in) the shortest path connecting x and y in G. The sum of the distances between all pairs of vertices of G is the Wiener index of G, denoted by $W(G)$. More details on distances in graphs

and their Wiener indices are found elsewhere [2,16–19].

Throughout this paper we use the following notation and conventions. By G_C, G_H, and G^* are denoted, respectively, the hydrogen-depleted molecular graph, the hydrogen-filled molecular graph, and the GAO, pertaining to the same molecule M. As before, the vertex set of G_H is $\mathcal{V}(G_H) = \{v_1, v_2, \ldots, v_N\}$. The vertices of G_H are labelled so that $v_1, v_2, \ldots, v_n$ represent the heavy atoms of M, whereas $v_{n+1}, v_{n+2}, \ldots, v_N$ the hydrogen atoms. If so, then the hydrogen-depleted molecular graph G_C has n vertices, and $\mathcal{V}(G_C) = \{v_1, v_2, \ldots, v_n\}$. Using the notation explained in Definition 1 we have that

$$\mathcal{V}(G^*) = \bigcup_{i=1}^{N} \mathcal{V}_i$$

implying that G^* possesses $n_1 + n_2 + \cdots + n_N$ vertices.

The basic result concerning distances in a GAO is the following Lemma, which is a straightforward consequence of the way in which the molecular graphs G_C, G_H, and G^* are constructed.

Lemma 2. Using the notation specified in Definition 1, if $x, y \in \mathcal{V}_i$ and $x \neq y$, then $d(x, y|G^*) = 2$. If $x \in \mathcal{V}_i$ and $y \in \mathcal{V}_j$, such that $i \neq j$, then $d(x, y|G^*) = d(v_i, v_j|G_H)$. Furthermore, if $i, j \leq n$ (i. e., if neither v_i nor v_j represent hydrogen atoms of M), then $d(x, y|G^*) = d(v_i, v_j|G_C)$.

Bearing Lemma 2 in mind we can now express the Wiener index of a GAO as:

$$W(G^*) = \sum_{i=1}^{N} 2\binom{n_i}{2} + \sum_{1 \leq i < j \leq N} n_i\, n_j\, d(v_i, v_j|G_H)\ . \tag{1}$$

The first term on the right-hand side of (1) is obtained by observing that in the set $\mathcal{V}_i$ there are $\binom{n_i}{2}$ vertex pairs, each at distance 2. The second term corresponds to vertex pairs of G^*, belonging to different sets $\mathcal{V}_i$ and $\mathcal{V}_j$. There are $n_i n_j$ such pairs, each at distance $d(v_i, v_j|G_H)$.

A vertex v_i of G_H pertains either to some heavy atom of M, in which case $n_i \geq 3$, or to hydrogen, in which case $n_i = 1$. This means that $n_i \geq 3$ holds for $1 \leq i \leq n$, whereas $n_i = 1$ for $n + 1 \leq i \leq N$. In view of this,

$$\begin{aligned} W(G^*) \ &= \ \sum_{i=1}^{n} 2\binom{n_i}{2} + \sum_{1 \leq i < j \leq n} n_i\, n_j\, d(v_i, v_j|G_H) + \sum_{1 \leq i \leq n < j \leq N} n_i\, d(v_i, v_j|G_H) \\ &+ \ \sum_{n < i < j \leq N} d(v_i, v_j|G_H)\ . \end{aligned} \tag{2}$$

In the most frequently occurring case, when all heavy atoms belong to the second row of the periodic system, i. e., when $n_i = 3$ for all $i = 1, 2, \ldots, n$, formula (2) is simplified as

$$W(G^*) = 6n + 9 \sum_{1 \le i < j \le n} d(v_i, v_j | G_H) + 3 \sum_{1 \le i \le n < j \le N} d(v_i, v_j | G_H) + \sum_{n < i < j \le N} d(v_i, v_j | G_H) \quad (3)$$

Combining Eq. (3) with the expressions for the Wiener indices of the hydrogen-depleted and hydrogen-filled molecular graphs:

$$W(G_C) = \sum_{1 \le i < j \le n} d(v_i, v_j | G_H)$$

$$W(G_H) = \sum_{1 \le i < j \le n} d(v_i, v_j | G_H) + \sum_{1 \le i \le n < j \le N} d(v_i, v_j | G_H) + \sum_{n < i < j \le N} d(v_i, v_j | G_H)$$

we arrive at

Theorem 3. The Wiener indices of the graph of atomic orbitals, the hydrogen-depleted, and the hydrogen-filled molecular graphs (pertaining to the same molecule in which all heavy atoms are from the second row of the periodic system) are related as

$$W(G^*) = 6n + 6W(G_C) + 3W(G_H) - 2W_H(G_H) \quad (4)$$

where n is the number of vertices of the hydrogen-depleted molecular graph, and $W_H(G_H)$ is the sum of distances between pairs of vertices corresponding to hydrogen atoms.

Formulas (3) and (4) hold for the vast majority of organic compounds. In particular, they are applicable to hydrocarbons.

WIENER INDEX OF GAO OF ALKANES

In the case of alkanes (compounds consisting of n carbon and $2n + 2$ hydrogen atoms) the relations between the Wiener indices of the various molecular graphs are significantly simpler:

Theorem 4. If G^*, G_C, and G_H are the graph of atomic orbitals, hydrogen-depleted molecular graph, and hydrogen-filled molecular graph, respectively, of an alkane with n carbon atoms, then the respective Wiener indices are related as

$$W(G^*) = 25\,W(G_C) + 15\,n^2 + 14\,n + 1 \tag{5}$$

$$W(G^*) = \left(\frac{5}{3}\right)^2 W(G_H) - 9\,n^2 - \left(n + \frac{4}{3}\right)^2 \tag{6}$$

$$W(G_H) = 9\,W(G_C) + (3\,n + 1)^2 \,. \tag{7}$$

Proof. Eq. (7) was proven in an earlier work [20]. Therefore, bearing in mind identity (4), what remains to be proven is an expression for $W_H(G_H)$ in terms of $W(G_C)$ and n. This expressions reads

$$W_H(G_H) = 4\,W(G_C) + (2\,n + 1)(3\,n + 1) \,. \tag{8}$$

Its proof is given in the Appendix.

Substituting (7) and (8) back into (4), we obtain (5). Combining (5) and (7), we obtain (6). □

APPENDIX: PROOF OF FORMULA (8)

Because formula (8) is independent of the GAO concept, its proof is given separately. Let T be an n-vertex chemical tree, i. e., a hydrogen-depleted molecular graph of an alkane C_nH_{2n+2}. Its vertex set is $\mathcal{V}(T) = \{v_1, v_2, \ldots, v_n\}$. Let the degree (= number of first neighbors) of the vertex v_i of T be δ_i.

By T_H we denote the hydrogen-filled molecular graph, corresponding to T. It is an $(3n + 2)$-vertex tree, possessing $2n + 2$ pendent vertices (= vertices of degree 1). T_H is obtained from T, by attaching $4 - \delta_i$ pendent vertices to the vertex $v_i \in \mathcal{V}(T)$, and doing this for all $i = 1, 2, \ldots, n$.

The quantity $W_H(T_H)$ is just the sum of distances between all pairs of pendent vertices of T_H. In other words, $W_H(T_H)$ is the sum of (topological) distances between all pairs of hydrogen atoms of the alkane molecule considered.

Any two pendent vertices of T_H, attached to the same vertex of T, are at distance 2. The sum of all such distances is equal to

$$\sum_{i=1}^{n} 2\binom{4-\delta_i}{2} = \sum_{i=1}^{n}(4-\delta_i)(3-\delta_i)$$

$$= \sum_{i=1}^{n}(4-\delta_i)(4-\delta_i) - \sum_{i=1}^{n}(4-\delta_i) = \sum_{i=1}^{n}(4-\delta_i)^2 - (2n+2) \ . \qquad (9)$$

For $i \neq j$, there exist $(4-\delta_i)(4-\delta_j)$ pairs of pendent vertices, one attached to v_i, the other attached to v_j. Each such pair is at distance $d(v_i, v_j|T) + 2$. The sum of all such distances is equal to

$$\sum_{1\le i<j\le n} (4-\delta_i)(4-\delta_j)[d(v_i, v_j|T) + 2]$$

$$= 2 \sum_{1\le i<j\le n} (4-\delta_i)(4-\delta_j) + \sum_{1\le i<j\le n} (4-\delta_i)(4-\delta_j)\, d(v_i, v_j|T) \ . \qquad (10)$$

Now,

$$2 \sum_{1\le i<j\le n} (4-\delta_i)(4-\delta_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}(4-\delta_i)(4-\delta_j) - \sum_{i=1}^{n}(4-\delta_i)(4-\delta_i)$$

$$= \left[\sum_{i=1}^{n}(4-\delta_i)\right]^2 - \sum_{i=1}^{n}(4-\delta_i)^2 = (2n+2)^2 - \sum_{i=1}^{n}(4-\delta_i)^2 \ . \qquad (11)$$

The sum of (9) and (10) is equal to $W_H(T_H)$. By taking into account (11) we obtain

$$W_H(T_H) = \left[\sum_{i=1}^{n}(4-\delta_i)^2 - (2n+2)\right] + \left[(2n+2)^2 - \sum_{i=1}^{n}(4-\delta_i)^2\right]$$

$$+ \sum_{1\le i<j\le n} (4-\delta_i)(4-\delta_j)\, d(v_i, v_j|T) \ . \qquad (12)$$

The last term on the right-hand side of (12) can be rewritten as

$$16 \sum_{1\le i<j\le n} d(v_i, v_j|T) - 4 \sum_{1\le i<j\le n} (\delta_i + \delta_j)\, d(v_i, v_j|T) + \sum_{1\le i<j\le n} \delta_i \delta_j\, d(v_i, v_j|T) \ . \qquad (13)$$

Evidently, $\sum_{1\le i<j\le n} d(v_i, v_j|T)$ is just the Wiener index $W(T)$ if the hydrogen-depleted molecular graph. In [21] it has been demonstrated that

$$\sum_{1\le i<j\le n} (\delta_i + \delta_j)\, d(v_i, v_j|T) = 4W(T) - n(n-1) \qquad (14)$$

$$\sum_{1\le i<j\le n} \delta_i \delta_j\, d(v_i, v_j|T) = 4W(T) - (2n-1)(n-1) \ . \qquad (15)$$

When identities (14) and (15) are substituted back into (13), and this combined with (12), we obtain after some calculation:

$$W_H(T_H) = 4\,W(T) + 6\,n^2 + 5\,n + 1$$

which is just another way of writing Eq. (8). □

References

[1] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1983; 2nd revised ed. 1992.

[2] I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer–Verlag, Berlin, 1986.

[3] D. Bonchev, D. H. Rouvray (Eds.), *Chemical Graph Theory – Introduction and Fundamentals*, Gordon & Breach, New York, 1991.

[4] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.

[5] A. Cayley, On the mathematical theory of isomers, *Phil. Mag.* **67** (1874) 444–446.

[6] A. A. Toropov, A. P. Toropova, QSPR modeling of the formation constants for complexes using atomic orbital graphs, *Russ. J. Coord. Chem.* **26** (2000) 398–405.

[7] A. A. Toropov, A. P. Toropova, QSPR modeling of the stability constants of biometal complexes with phosphate derivatives of adenosine, *Russ. J. Coord. Chem.* **26** (2000) 792–797.

[8] A. Mercader, E. A. Castro, A. A. Toropov, QSPR modeling of the enthalpy of formation from elements by means of correlation weighting of local invariants of atomic orbital molecular graphs, *Chem. Phys. Lett.* **330** (2000) 612–623.

[9] A. A. Toropov, A. P. Toropova, Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graphs of local invariants, *J. Mol. Struct. (Theochem)* **538** (2001) 287–293.

[10] A. A. Toropov, A. P. Toropova, QSPR modeling of stability of complexes of adenosine phosphate derivatives with metals absent from the complexes of the teaching access, *Russ. J. Coord. Chem.* **27** (2001) 574–578.

[11] A. A. Toropov, A. P. Toropova, QSAR modeling of mutagenicity based on graph of atomic orbitals, *Internet Electron. J. Mol. Des.* **1** (2002) 109–113.

[12] A. A. Toropov, A. P. Toropova, QSPR modeling of complex stability by optimization of correlation weights of the hydrogen bond index and the local graph invariants, *Russ. J. Coord. Chem.* **28** (2002) 877–880.

[13] A. A. Toropov, A. P. Toropova, QSPR modeling of alkane properties based on graph of atomic orbitals, *J. Mol. Struct. (Theochem)* **637** (2003) 1-10.

[14] A. A. Toropov, A. P. Toropova, I. V. Nesterov, O. M. Nabiev, Comparison of QSAR models of anti-HIV-1 potencies based on labeled hydrogen filled graph and graph of atomic orbitals, *J. Mol. Struct. (Theochem)* **640** (2003) 175–181.

[15] A. A. Toropov, A. P. Toropova, Nearest neighboring code and hydrogen bond index in labeled hydrogen–filled graph and graph of atomic orbitals: Application to model of normal boiling points of haloalkanes, to be published.

[16] F. Buckley, F. Harary, *Distance in Graphs*, Addison–Wesley, Redwood, 1990.

[17] I. Gutman, Y. N. Yeh, S. L. Lee, Y. L. Luo, Some recent results in the theory of the Wiener number, *Indian J. Chem.* **32A** (1993) 651–661.

[18] A. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* **66** (2001) 211–249.

[19] A. A. Dobrynin, I. Gutman, S. Klavžar, P. Žigert, Wiener index of hexagonal systems, *Acta Appl. Math.* **72** (2002) 247–294.

[20] I. Gutman, D. Vidović, L. Popović, On graph representation of organic molecules – Cayley's plerograms vs. his kenograms, *J. Chem. Soc. Faraday Trans.* **94** (1998) 857–860.

[21] I. Gutman, Selected properties of the Schultz molecular topological index, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1087–1089.