

From Wiener Index to Molecules

Gonzalo Cerruela García, Irene Luque Ruiz,* and Miguel Ángel Gómez-Nieto

Department of Computing and Numerical Analysis, University of Córdoba, Campus Universitario de Rabanales, Albert Einstein Building, E-14071 Córdoba, Spain

Juan Antonio Cabrero Doncel and Antonio Guevara Plaza

Department of Languages and Computer Science, University of Málaga, Campus Universitario de Teatinos, E-29071 Málaga, Spain

Received July 6, 2004

In this paper we present an algorithm for the generation of molecular graphs with a given value of the Wiener index. The high number of graphs for a given value of the Wiener index is reduced thanks to the application of a set of heuristics taking into account the structural characteristics of the molecules. The selection of parameters as the interval of values for the Wiener index, the diversity and occurrence of atoms and bonds, the size and number of cycles, and the presence of structural patterns guide the processing of the heuristics generating molecular graphs with a considerable saving in computational cost. The modularity in the design of the algorithm allows it to be used as a pattern for the development of other algorithms based on different topological invariants, which allow for its use in areas of interest, say as involving combinatorial databases and screening in chemical databases.

1. INTRODUCTION

A nondirected and connected graph $G=(V, E, F)$ represents the topological structure of a chemical compound, when the set of vertices (or nodes) (V) represents the atoms of the molecule, the set of edges (E) represents the connections among the atoms, and F is a function that characterizes the types of edges and vertices. So, when the graph G represents the molecular structure of a chemical compound it is called a molecular graph.^{1–4}

Mathematical descriptors derived from molecular graphs have been widely used in QSPR and QSAR which are known as topological indices (TIs). Topological indices are based on a mathematical calculation over a matrix representing topological information of the molecular graphs.^{5–10}

Although, molecular graphs are usually represented by means of the adjacency matrix (A); a symmetrical matrix where each element $A(i, j)=1$, if the vertex i is adjacent to the vertex j , and 0 otherwise (and in which all the elements of the main diagonal are equal to 0). Matrices based on the distance between the graph vertices are used in the calculation of most of the topological descriptors.

The Wiener index¹¹ is one of the most widely used and referenced topological indices,^{12–17} based on the distance matrix (D), a symmetrical matrix where each element $D(i, j)$ has an integer value equal to the length of the shortest path between the graph vertices—the minimum number of edges that is necessary to travel in the G graph to reach the vertex j from the vertex i (or vice versa). This index is broadly utilized in computational chemistry to measure some topological properties (branching, cyclicity, centrality, etc.)

and in QSPR/QSAR studies.^{1,2,7,8,10,18,19} In recent years a large number of Wiener-type indices have been proposed (Hyper-Wiener, Detour, etc.) which are based on distance-related matrices (distance-path, detour, reciprocal, etc.).^{20–31}

The Wiener index can be easily calculated from the distance matrix (D) as follows

$$W = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N D(i, j) = \sum_i \sum_{j>i} D(i, j) \quad (1)$$

where N is the number of vertices of the G graph.

Some proposals have been carried out for the calculation of the Wiener index without the need to inspect the elements of the D matrix.^{32–35} When G is a complete graph (K_N graph), the Wiener index is given by the expression:

$$W = \frac{N^2 - N}{2} \quad (2)$$

And when G is a path graph, W can be calculated by the following expression:


$$W = \sum_{i=1}^{i=N-1} i(N-i) = \frac{N^3 - N}{6} \quad (3)$$

The expressions (2) and (3) mark the lower and upper bounds respectively of the Wiener index for a given N of a connected graph.

To calculate the Wiener index, a series of mathematical expressions considering the different types of molecular topologies (tree, stars, etc.) has been proposed, some of them being based on the decomposition of the distance matrix and


* Corresponding author phone: +34-957-21-2082; fax: +34-957-21-8630; e-mail: mailurui@uco.es.

Chart 1



0	1	2	3	2	1	=	9
1	0	1	2	3	2	=	9
2	1	0	1	2	3	=	9
3	2	1	0	1	2	=	9
2	3	2	1	0	1	=	9
1	2	3	2	1	0	=	9
+							54

$W=54/2=27$



0	1	1	2	3	3	=	10
1	0	1	2	3	3	=	10
1	1	0	1	2	2	=	7
2	2	1	0	1	1	=	7
3	3	2	1	0	1	=	10
3	3	2	1	1	0	=	10
+							54

$W=54/2=27$

calculating the Wiener index from the partial values of W calculated for each one of the submatrices.^{12,16,17,33–39}

If the structure of the molecular graph is unknown a priori, the Wiener index is usually calculated by means of the expression (1), with a low computational cost; however, the inverse process is somewhat more complex.

Given a value of the Wiener index $W=a$, more than one D matrix exists that satisfies this condition, as shown in Chart 1, where two different matrices (molecules) have the same value of $W=27$.

As the W value increases, the number of D matrices that satisfies this W value increases exponentially, since the size of the D matrix will vary over a wider range of values: from a complete graph K_N (the length of the shortest path between any couple of vertices is equal to 1), to a path graph, where all of its vertices and edges lie on a single straight line.

However, the adjacency matrix (derived from the molecular graph) does not uniquely represent the topological structure of a molecule. Organic compounds containing heteroatoms and multiple bonds must be represented as a vertex and edge-weighted molecular graph (VEW),^{14,15,27,40} which can be represented as weighted graph matrices.

A simple enough model for our problem is obtained without considering hydrogen atoms, weighting the vertices with the atomic number, and weighting the edges with the bond multiplicity. Thus, a so-called *connection matrix*⁴¹ (C) can be derived from this VEW representation, which completely represents the molecular graph G , since it keeps: a) the color (or type) of the vertices (atoms) via the elements $C(i,i)$ and b) the color (or type) of the edges (bonds) via the elements $C(i,j)$. The obtaining of the connection matrix (C) from the adjacency matrix (A) is again a combinatorial process in which the color of the vertex (atom type) and edge (bond type) should be kept in mind. Thus, although the molecular structure can be represented by means of a graph G , not all graphs G represent a chemically valid molecular structure. So, the combinatorial generation of graphs from the Wiener index yields nonvalid graphs, which should be discarded.

The computational cost of the process for obtaining molecular structures from the Wiener index will therefore be delimited by the following basic factors:

- The Wiener index value. As W increases, also the number and size of the possible molecular graphs increase.
- The set of colors or types of permitted atoms in the molecular structure and the occurrences of each of them, since each atom has a relationship capacity with the rest of the atoms of the molecule (coordination valency).
- The set of relationships allowed among the atoms, that is, the type of allowed connections and their occurrences.
- Other topological characteristics such as the number of cycles and size, etc.

In this paper we propose an algorithm that uses heuristics allowing the generation of chemically valid molecular structures with a cost lower than the theoretical computational cost. This algorithm generates SDF⁴² format representations of the chemical compounds that can be treated by other applications for use in solutions to many of the outlined problems in computational chemistry.

In this paper these basic determinant factors of the process of generation of molecular graphs from the Wiener index will be considered. Other factors related to the topology and activity of the molecule can be considered (for example, the number of H bond donors and H bond acceptors, etc.), and they can easily be included in the algorithm described in the following section.

2. DESCRIPTION OF THE ALGORITHM

Given a value (W) of the Wiener index (or an interval, or a set of values), the developed algorithm generates molecular structures following the strategy, as shown in the Context diagram of Figure 1:

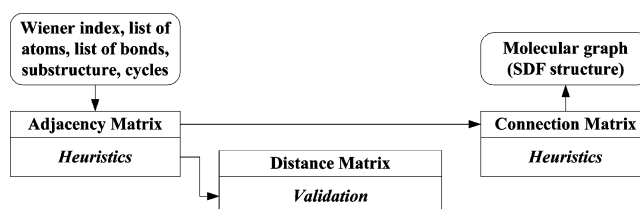


Figure 1. Context diagram of the developed algorithm.

- Construction of the adjacency matrix (A).
- Construction of the distance matrix (D).
- Construction of the connection matrix (C) and generation of the molecular structure in SDF format.

In each of the steps, the algorithm uses a series of heuristics in order to reduce the natural combinatorial explosion of the process.

2.1. Construction of the Adjacency Matrix. It is the first step of the algorithm, which determines most of the computational cost. So, in this step most of the proposed heuristics are applied:

2.1.1. Bounds for the Molecular Graph Size. Given a value (W) of the Wiener index, the first step is to obtain the bounds (lower and upper) between which a G graph can exist with a number of vertices N whose distance matrix D yields the Wiener index.

These bounds can be obtained in an approximate way considering: a) the path graph (for a lower bound) and b) a maximally complete graph (for an upper bound) that have the value of the given Wiener index.

Table 1. Heuristics Considering the Type of Connectivity Present in the Graphs^a

type of connectivity	N	occurrences (number of nodes)					
		minimum		maximum		values	
		even	odd	even	odd	even	odd
1-2	≥3	² c = 2	² c = 1		² c = N-2	mod(² c, 2) = 0	mod(² c, 2) ≠ 0
1-3	≥4	³ c = 1	none	³ c = N-1	none	any	none
1-4	≥5 ^b	⁴ c = 2	⁴ c = 1		⁴ c = N-2	mod(⁴ c, 2) = 0	mod(⁴ c, 2) ≠ 0
2-3	≥4		³ c = 2	³ c = N-2	³ c = N-1	mod(³ c, 2) = 0	
2-4	≥5		⁴ c = 1		⁴ c = N-2	any	
3-4	≥5	⁴ c = 2	⁴ c = 1		⁴ c = N-2	mod(⁴ c, 2) = 0	mod(⁴ c, 2) ≠ 0
1-2-3	≥4		³ c = 1	³ c = N-3	³ c = N-2	any	
	even		² c = 2	² c = N- ³ c-2 ^e	² c = N- ³ c-1 ^e	mod(² c, 2) = 0	
	odd		² c = 1	² c = N- ³ c-2 ^e	² c = N- ³ c-1 ^e	mod(² c, 2) ≠ 0	
1-2-4	≥5		⁴ c = 1		⁴ c = N-3	any	
	even	² c = 2 ^e	² c = 1 ^e	² c = N- ⁴ c-2		mod(² c, 2) = 0 ^e	mod(² c, 2) ≠ 0 ^e
	odd	² c = 1 ^e	² c = 2 ^e	² c = N- ⁴ c-2		mod(² c, 2) ≠ 0 ^e	mod(² c, 2) = 0 ^e
2-3-4	≥5		⁴ c = 1		⁴ c = N-3	any	
	even		³ c = 2	³ c = N- ⁴ c-2 ^d	³ c = N- ⁴ c-1 ^d	mod(³ c, 2) = 0	
	odd		³ c = 2	³ c = N- ⁴ c-1 ^d	³ c = N- ⁴ c-2 ^d	mod(³ c, 2) = 0	
1-2-3-4	≥5		⁴ c = 1		⁴ c = N-3	any	
	even		³ c = 1	³ c = N- ⁴ c-3 ^e	³ c = N- ⁴ c-2 ^e	mod(² c, 2) = 0 ^e	mod(² c, 2) ≠ 0 ^e
		² c = 2 ^e	² c = 1 ^e	² c = N- ⁴ c-3-2 ^f	² c = N- ⁴ c-3-1 ^f		
	odd		³ c = 1	³ c = N- ⁴ c-2 ^e	³ c = N- ⁴ c-3 ^e	any	
		² c = 1 ^e	² c = 2 ^e	² c = N- ⁴ c-3-2 ^f	² c = N- ⁴ c-3-1 ^f	mod(² c, 2) ≠ 0 ^e	mod(² c, 2) = 0 ^e

^a $\sum_i \sum_j A(i,j) = x$, x = value of connectivity. $\text{mod}(y,z) = y - \text{int}(y/z) \cdot z$. $\text{int}(x)$ is the integer part of x number. ^b $N \geq 8$ when even. ^c Considering even/odd the value of ³c. ^d Considering even/odd the value of ⁴c. ^e Considering ⁴c occurrences. ^f Considering ³c occurrences.

The lower bound (N_-)—path graphs—can be obtained solving the expression (3) as

$$N_- = \frac{1}{3}F(W)^{(1/3)} + F(W)^{-(1/3)} \quad (4)$$

where

$$F(W) = 81W + 3\sqrt{-3 + 729W^2} \quad (5)$$

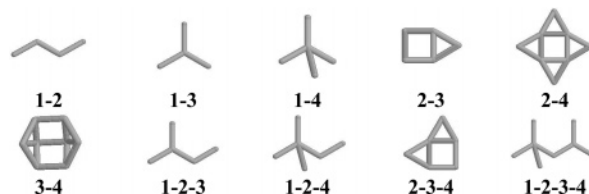
If we consider that the highest number of connections of the graph vertices is 4, then the highest number of independent cycles (SSSR) that can be formed in a graph of N vertices (for $N > 4$) is $N+1$. So, the smallest Wiener index for a graph of N vertices and $N+1$ cycles is equal to $N(N-3)$, that is there does not exist graphs (with all vertex 4-connected) with a Wiener index lower than $N(N-3)$. Therefore, the upper bound (N_+) is obtained by the expression:

$$N_+ = \frac{3 + \sqrt{9 + 4W}}{2} \quad (6)$$

2.1.2. Occurrences of Atom Connectivity. Since the adjacency matrix represents a connected graph and since each atom cannot maintain a number of connections greater than its coordination capacity (valency), the process of obtaining the adjacency matrix for a value V in the interval $[N_-, N_+]$ implies obtaining all the combinations with repetition of m elements (the maximum coordination number that an atom can have) over N (the number of atoms of the molecule or size of the adjacency matrix).

Initially we can set two trivial restrictions directly derivable from the molecular graph characteristics (Euler theorem^{2,3}):

1. The sum of all the connections of the atoms must be even:
2. The number of vertices with odd number of connections is always even.

Chart 2

Moreover, introducing the condition that the maximum coordination that an atom can have is 4, as is the case of the carbon atom, it is possible to restrict the number of possible combinations by means of a set of simple heuristics considering the different possible connections among the elements of the A matrix,

So, the different connectivities manifestable by the vertices of the adjacency matrix are considered. The possible cases are as follows: 1-2 (there only exist vertices with connectivity 1 and vertices with connectivity 2), 1-3, 1-4, 2-3, 2-4, etc.

Table 1 shows some of the restrictions established considering the total number of adjacencies of a given value between 1 and 4 as a function of the size of the adjacency matrix and Chart 2 shows some graph examples for different connectivity values.

The establishment of the bounds and the values of the adjacency matrix rows much restricts the combinatorial order of adjacency matrices which should be generated in the interval of values $[N_-, N_+]$ obtained for a given value of the Wiener index.

2.1.3. Occurrences of Atom Types. The occurrences of the types of atoms are directly related to the occurrences of the relationships among the atoms, since the atoms are limited in their relationships by their coordination valency.

The developed algorithm receives a list of the types and minimum and maximum occurrences of each atom type as input which will be considered in the molecular structures generated (see Figure 1). These values limit the possible set

of adjacency matrices built by means of checking the number of total connections of each element of the adjacency matrix and the minimum and maximum occurrences of each atom type (considering the maximum coordination valency of the atom).

2.1.4. Occurrences of Connection Types. The same as in the previous case the types of connections among the atoms are directly related to the number of relationships among them. As the allowed multiplicity of the connections increases (simple, double, triple), the number of relationships among the atoms of the molecule diminishes.

As above, the developed algorithm receives a list of the types and the minimum and maximum occurrences of each connection type as input which will be considered in the molecular structures generated (see Figure 1). These values will limit the possible set of adjacency matrices built.

2.1.5. Number and Size of the Cycles. The existence, number and size of the cycles is another restrictive factor in the following: a) the size of the adjacency matrix, since as the number of cycles increases, the distance among the graph vertices diminishes and, therefore, the size of the graph that satisfies a given value of the Wiener index increases and b) the number of the relationships among the graph vertices increases (degree of the vertices) and, therefore, the multiplicity of the relationship diminishes.

The presence of cycles in a graph can be directly obtained from the adjacency matrix by means of the following expression:

$$\eta = E - N + 1 \quad (7)$$

where η is the cyclomatic index (for a connected graph), E is the number of edges, and N is the number of vertices of the graph.

The value of η gives information about of the size of the SSSR (*Smallest Set of Smallest Rings*). For the calculation of this set we have found the algorithm proposed by Figueras⁴³ to be efficient.

As Figure 1 shows, the proposed algorithm receives the number and size of this SSSR set as input. This information is used by the algorithm to reduce (delete) those adjacency matrices that do not satisfy the input parameters.

2.1.6. Existence of Patterns. In the process of construction of molecular graphs from the Wiener index, it is convenient to consider that in all the generated graphs a pattern, that is, a common substructure may be present.

Such as common substructure, reduces the dimension of the problem, since this common substructure or pattern is represented by an adjacency matrix (A_p) which should be present in each of the adjacency matrices (A) generated by the algorithm ($A_p \subset A$).

In Figure 2 the activity diagram⁴⁴ corresponding to the generation of the adjacency matrix is shown. As we can appreciate

a) The information corresponding to the presence of a substructure is used initially to reduce the size of the adjacency matrix. So, the combinatorial problem decreases approximately in size $N - N_p$, where N is the value of the size of the adjacency matrix with the given Wiener index value, and N_p is the number of vertices of the substructure pattern.

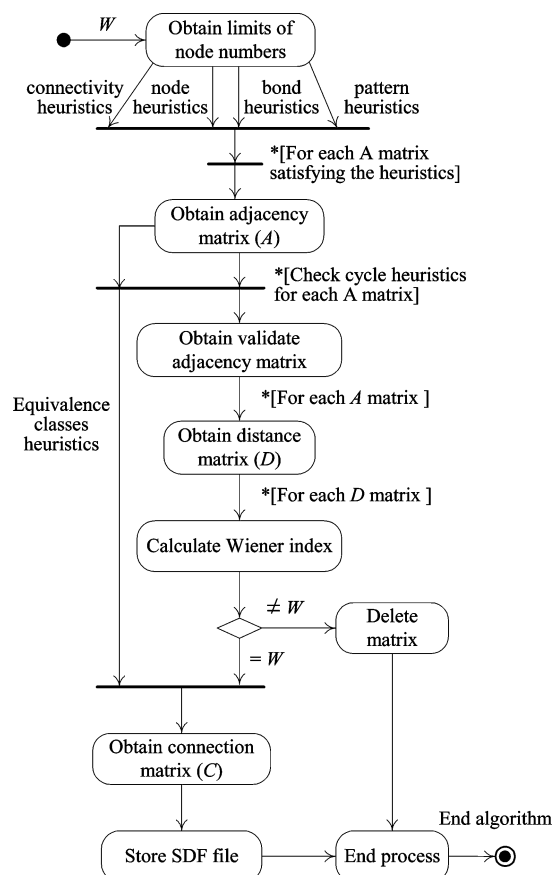


Figure 2. Activity diagram of the developed algorithm.

b) The information corresponding to the types of atoms (color of the vertices) and their number determines the relationships among the vertices of the adjacency matrix and, therefore, the values of the coefficients of Table 1.

c) The information corresponding to the types of connections and their number also determines the relationships among the vertices of the adjacency matrix. This information together with that corresponding to the types of atoms is used as an initial heuristic for the reduction of the size of the problem.

d) Last, the information corresponding to the number and size of the cycles is used, once the adjacency matrix has been generated.

2.2. Construction of the Distance Matrix. For each of the adjacency matrices (A) generated in the previous step the distance matrix (D) is built making use of the Floyd algorithm.⁴⁵

As the diagram in Figure 2 shows, for each D matrix obtained, the Wiener index is calculated, rejecting those adjacency matrices that do not yield the required W value.

2.3. Construction of the Connection Matrix. The connection matrix faithfully represents the topology or structure of a molecule, since besides keeping the color of the atoms and their relationships, it keeps the relationship type among the atoms (bonds).

It is a combinatorial process again, where with each connection matrix generated from the distance matrix, it combines the set of atoms and connections defined in the input to the algorithm, considering the minimum and maximum cardinalities defined (optionally) for each type.

The limitation of the combinatorial process will be given by this information, to satisfy each connection matrix:

1. Considering the type of atoms

$$\sum_j C(i,j) \leq v_i \quad (8)$$

where v_i is the coordination valency of the i atom.

2. Considering the established limits according to the type and occurrences of atoms and connections.

As Figure 2 shows this combinatorial explosion can be reduced again by means of the calculation of the classes of equivalence of vertices and connections.

The calculation of the equivalence classes of vertices⁴⁶ allows reduction of the combinatorial explosion so that the number of susceptible vertices to be colored in each step decreases to the number of existing equivalence classes of vertices in that step of the process of coloring. Although a computational cost exists for the calculation of the equivalence classes, the total cost of the process decreases, since typically the number of classes is much smaller than the number of vertices.

This same reduction can be carried out for the coloring of the connections. Thus, the graph represented by the connection matrix C is reduced to a line graph^{1,2,47} and the classes of equivalence of the vertices are obtained, which assumes that we know the equivalence classes of the connections.

Having obtained the connection matrices, with the use of Marvin's library,⁴² the corresponding SDF structures are generated. As appreciated in Figure 2, from the value of the Wiener index we obtain the minimum and maximum values corresponding to the size of the matrices that should be generated (number of atoms of the molecules). The application of the heuristics corresponding to the connectivity among the atoms (section 2.1.2), occurrences and color of the atoms (section 2.1.3) and occurrences and type of connections (section 2.1.4) are applied and a main set of adjacency matrices are obtained, which are analyzed to check the fulfillment of the heuristics corresponding to the presence of cycles (section 2.1.5) and, of course, their capacity to represent molecular graphs.

For each valid adjacency matrix the corresponding distance matrix is obtained, and the Wiener index is calculated eliminating those matrices that do not satisfy the required value.

Last, the heuristics corresponding to the equivalence classes are applied, which allows us to color (atoms and connections) the adjacency matrices and to build the connection matrices that represent the molecules. The result is stored in a SDF file that contains all the molecular structures generated by the algorithm.

3. CONSTRUCTION OF MOLECULAR GRAPHS

The proposed algorithm has been implemented in Java 1.4, for its portability and integration to other applications. The tests have been carried out on a personal computer, with processor Pentium III 400 MHz and 64 Mb.

Table 2 shows the results obtained for different orders of the value of the Wiener index, for the process of generation of molecules without coloring (only carbon atoms and simple connections are considered). Table 2 shows the following: the value of the Wiener number, the minimum and maximum

Table 2. Some Results Obtained Considering Only Simple Bonds and C Atoms in the Molecules^a

Wiener	$N_- - N_+$	$N_-^* - N_+^*$	A	C	T
10	4-5	4-5	18	2	0.02
20	5-6	5-6	21	7	0.02
30	6-7	6-7	73	21	0.02
40	7-8	7-8	279	23	0.03
50	7-8	7-8	1345	143	0.33
60	8-9	8-9	6619	546	0.98
70	8-10	8-10	9050	580	1.04
80	8-10	8-10	62691	3903	3.55
90	9-10	9-10	91413	3424	4.39
100	9-11	9-11	669562	22872	13.22
110	9-11	9-11	787083	29665	14.75
120	9-12	9-12	3587477	46829	59.17
130	10-12	10-12	4089351	207129	62.31

^a W: Wiener index, $N_- - N_+$: lowest and highest number of graph vertices in the generated molecules, $N_-^* - N_+^*$: lowest and highest number graph vertices estimated, A: number of the adjacency matrices generated with heuristics, C: number of valid connection matrices (molecules) generated, T: computational time in seconds.

size of the obtained adjacency matrices ($N_- - N_+$), the minimum and maximum size of the calculated adjacency matrices ($N_-^* - N_+^*$), the number of generated adjacency matrices, the number of generated connection matrices and the total time of computation in seconds, using the heuristics described in this paper. The approximate number of possible adjacency matrices, which can be generated for a graph of size N , is given by the following expression:

$$\prod_{i=1}^{i=N-1} \sum_{\substack{j=0 \\ i \geq j}}^{j=4} \binom{i}{j} \quad (9)$$

where the highest coordination of 4 is considered for the graph vertices. We can observe in Table 2 that the efficiency of the heuristic used in our algorithm is very high, increasing as the size of the adjacency matrix increases.

However, although the use of the heuristics considerably reduces the number of adjacency matrices generated with regard to when the heuristics are not used (equation 10), this number increases considerably as the W value increases. Evidently, as W increases, the size of the graph increases, and, therefore, the number of possible adjacency matrices also increases in a combinatorial expression, (although many of these adjacency matrices later on do not satisfy the value required for W) producing a decrease in efficiency when the W value increases.

This fact leads to the study of more heuristics to diminish the construction of adjacency matrices that later on are deemed deletable. However, given the computational cost that the construction of an adjacency matrix entails, the introduction of new heuristics in this step of the algorithm could give an increase in the total time of computation.

Table 3 shows the results obtained considering that from 1 to 2 atoms of oxygen and from 1 to 2 atoms of chlorine can be present in the generated molecules. The presence of heteroatoms initially influences the heuristics that consider the construction of the adjacency and connection matrices. Table 3 shows the data corresponding to the number of matrices for these two steps of the developed algorithm.

The heuristics which consider the presence of heteroatoms reduce the number of adjacency matrices generated with

Table 3. Some Results Obtained Considering Only Simple Bonds, One Cycle (of Any Size), Any C Atoms and One to Two O and Cl Atoms in the Molecules^a

Wiener	$N_- - N_+$	A	C	M	T
8	4-4	2	2	2	0.02
16	5-5	5	2	8	0.02
26	6-6	12	3	33	0.08
42	7-7	102	6	122	0.78
60	8-8	258	7	247	2.15
80	9-9	733	10	644	4.95
90	9-9	660	17	925	7.46
110	10-10	1891	18	1687	21.78
128	10-10	4938	13	963	7.51
135	10-11	4399	16	1441	26.71

^a W: Wiener index, $N_- - N_+$: lowest and highest number of graph vertices in the generated molecules, A: number of the adjacency matrices generated with heuristics, C: number of the uncolored connection matrices generated, M: number of generated molecules, T: computational time in seconds.

regard to when the heuristics are not applied (Table 3) and with regard to when there are no heteroatoms present (see the results for equal values in Table 2), which is translated as an increase in the efficiency of the algorithm.

We have observed that the values of the efficiency for the use of heuristics in the process of coloring oscillate with the increase in value of the Wiener index. This effect is because the heuristics based on the calculation of the equivalence classes for the coloring of the vertices are very dependent on the topology of the graphs, so that similar values of the Wiener index produce very different graphs in their topology compared to the present equivalence classes in each step of the coloring process.

4. DISCUSSION AND REMARKS

As an example of the application of the algorithm we show in Figures 3 and 4 the molecules generated by the developed

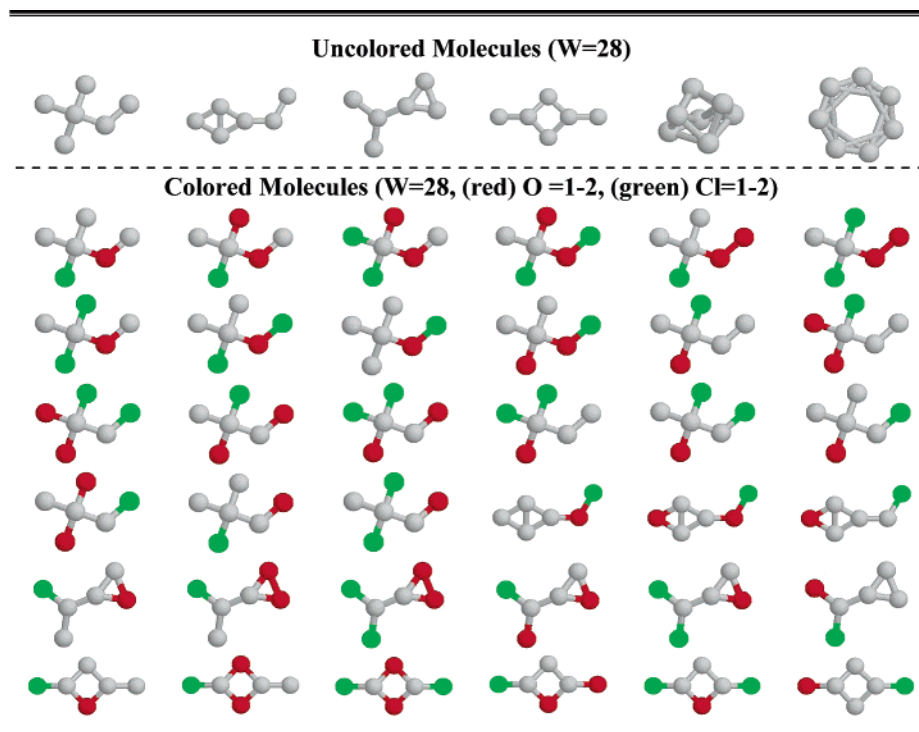
algorithm for some working conditions selected so that the set of generated molecules has a size that allows for their representation in the manuscript in a clear way.

Figure 3 shows the generated molecules for the value of the Wiener index equal to 28, in which from 1 to 2 atoms of oxygen and from 1 to 2 atoms of chlorine can be present, there being no restrictions for the type and occurrences of bonds and cycles. Figure 4 shows some of the generated molecules including pyridine as pattern and satisfying some restrictions in order to reduce the number of possible molecules: Wiener index from 75 to 101, one atom of oxygen, one cycle of size five and, of course, the characteristics corresponding to pyridine (three double bonds, one atom of nitrogen, and one cycle of size six).

As we can observe the proposed algorithm is suitable for the generation of molecular graphs from values of the Wiener index. The molecular graphs are generated in an efficient way, despite the combinatorial problem, thanks to the use of heuristics that consider the following: a) the properties of the related and non directed graphs, b) the indexes of coordination of the atoms, c) the different types of connections, d) the presence and occurrences of heteroatoms, e) the size and occurrences of the cycles and f) the presence of substructures.

However, as the size of the graphs increases (with high values of the Wiener index), the combinatorial problem leads to an excessive computational cost, due to the increase in the graph size and to the increase in the values of the interval $[N_-, N_+]$. In practice this inconvenience is minimized because molecules satisfying a known set of constraints (number and size of the cycles, occurrences and type of heteroatoms, etc.) are usually built.

The developed algorithm can be enlarged so that it includes other topological indices habitually utilized in computational chemistry. For that, it is necessary to propose models that

**Figure 3.** An example of molecules generated for Wiener index equal to 28.

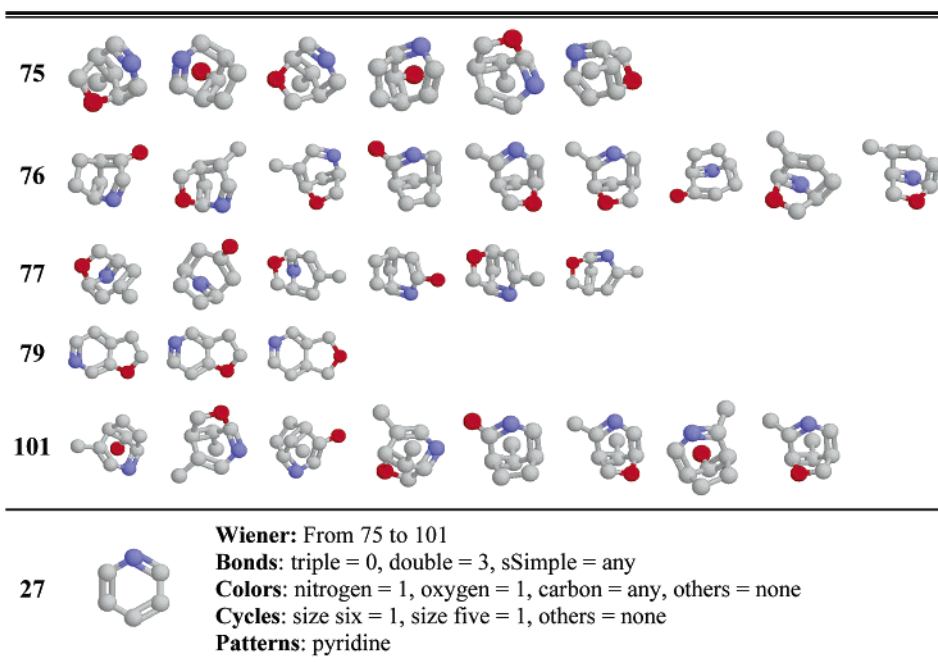


Figure 4. An example of molecules generated including pyridine as pattern.

allow us to predict, with a reasonable margin of error, the characteristics of the molecular graphs which satisfy a given value of the selected topological index and the development of new heuristics adapted to the characteristics and value of the topological index utilized.

Evidently the complexity and reliability of this calculation is dependent on the topological index in question. While for topological indices which are based on the adjacency matrix it is relatively simple, for those indices based on other types of matrices (minimum, maximum, reciprocal, inverse distance, etc.)^{1,4,5} and which consider chemical features (type of atoms, valency, electronegativity, distances of connection, etc.), this calculation may be complex and subject to considerable deviations.

Some topological indices based on the adjacency matrix are obtained by means of the following expression:

$$Index = b \sum_i \sum_j A(i,j) \times (v_i \times v_j)^a \quad (10)$$

where $v_k = \sum_i A(k,i)$ is the coordination valency of atom k .

From the previous expressions lower and upper bounds for the graph size can be extracted for a given value of the considered index. In the case of the Randić⁸ connectivity index (${}^1\chi$), where $b=1/2$ and $a=-1/2$, for $n \geq 5$, and maximum vertex degree of 4, we note the following bounds:

$$\text{Path graphs: } N_- = 2({}^1\chi - \sqrt{2}) + 3 \quad (11)$$

$$\text{4-regular graphs: } N_+ = {}^1\chi \times 2 \quad (12)$$

However although for other topological indices the error obtained in the values of N_- and N_+ can be appreciable the use of the heuristics considerably reduces the combinatorial problem, for which the proposed algorithm can be adapted to the construction of combinatorial databases based on these topological indices. At the moment we are working on the development of models that consider different topological invariants for their application to different fields such as the

construction of combinatorial databases, screening in chemical databases,^{48,49} prediction of molecular structures in model QSPR, etc.

ACKNOWLEDGMENT

The authors wish to express their most sincere gratitude to the reviewers of this paper. The valuable effort, comments and suggestions of the reviewers and the added corrections to this paper have been of great aid to the authors.

REFERENCES AND NOTES

- (1) Rouvray, D. H.; Balaban, A. T. *Chemical Applications of Graph Theory. Applications of Graph Theory*. Wilson, R. J., Beineke, L. W., Eds.; Academic Press: 1979; pp 177–221.
- (2) Gross, J.; Yellen, J. *Graph Theory and Its Applications*. CRC Press: 1999.
- (3) Diestel, R. *Graph Theory*. Electronic Edition. Springer-Verlag: 2000.
- (4) Ivanciuc, O.; Balaban, A. T. *The Graph Description of Chemical Structures*. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 59–167.
- (5) Lucic, B.; Lukovits, I.; Nikolic, S.; Trinajstić, N. Distance-Related Indexes in the Quantitative Structure–Property Relationship Modeling. *J. Chem. Inf. Comput. Sci.* **2001**, 41(3), 527–535.
- (6) Mandloi, M.; Sikarwar, A. Sapre, N. S. A Comparative QSAR Study Using Wiener, Szeged, and Molecular Connectivity Indices. *J. Chem. Inf. Comput. Sci.* **2000**, 40 (1), 57–62.
- (7) Diudea, M. V. *QSPR/QSAR Studies by Molecular Descriptors*. Nova Science Publishers: Huntington, NY, 2001.
- (8) Randić, M. *Topological Indices*. In *Encyclopedia of Computational Chemistry*; Scheleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018–3032.
- (9) Randić, M.; Zupan, J. On Interpretation of Well-Known Topological Indices. *J. Chem. Inf. Comput. Sci.* **2001**, 41 (3), 550–560.
- (10) *Topological Indices and Related Descriptors in QSAR and QSPR*. Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Singapore, 1999; 811 pp, 90–5699–239–2.
- (11) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- (12) Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. Some Recent Results in the Theory of the Wiener Number. *Indian J. Chem.* **1993**, 32A, 651–661.
- (13) Ivanciuc, O. QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **2000**, 40 (6), 1412–1422.

- (14) Ivanciuc, O.; Klein, D. J. Computing Wiener-Type Indices for Virtual Combinatorial Libraries Generated from Heteroatom-Containing Building Blocks. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (1), 8–22.
- (15) Ivanciuc, O.; Klein, D. J. Building-Block Computation of Wiener-Type Indices for Virtual Screening of Combinatorial Libraries. *Croat. Chem. Acta* **2002**, *75* (2), 577–601.
- (16) MATCH (Communications in Mathematical Chemistry and in Computer Chemistry), **1997**, 35. Fifty Years of the Wiener Number.
- (17) Nikolic, S.; Trinajstić, N.; Mihalic, Z. The Wiener Index: Development and Applications. *Croat. Chem. Acta* **1995**, *68*, 105–129.
- (18) Bonchev, D.; Balaban, A. T.; Mekenyan, O. Generalization of the Graph Center Concept. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 106–113.
- (19) Mekenyan, O.; Bonchev, D.; Trinajstić, N. A. Topological Characterization of Cyclic Structures with Acyclic Branches. *MATCH* **1981**, *11*, 145–168.
- (20) Balaban, A. T.; Mills, D.; Ivanciuc, O. Basak S. C. Reverse Wiener Indices. *Croat. Chem. Acta* **2000**, *73* (4) 923–941.
- (21) Bonchev, D. The Overall Wiener Index. A New Tool for Characterization of Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 582–592.
- (22) Bytautas, L.; Klein, D. J. Mean Wiener Numbers and Other Mean Extensions for Alkane Trees. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (2), 471–481.
- (23) Estrada, E.; Gutman, I. A Topological Index Based on Distance of Edges of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (4), 850–853.
- (24) Gutman, I. A New Hyper-Wiener Index. *Croat. Chem. Acta* **2004**, *77* (1–2) 61–64.
- (25) Gutman, I.; Linert, W.; Lukovits, I.; Tomovic, Z. The Multiplicative Version of the Wiener Index. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (1), 113–116.
- (26) Gutman, I.; Vukievi, D.; Zerovnik, J. A Class of Modified Wiener Indices. *Croat. Chem. Acta* **2004**, *77* (1–2) 103–109.
- (27) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 169–220.
- (28) Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. Wiener Index Extension by Counting Even/Odd Graph Distances. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 536–549.
- (29) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *17*, 50–52.
- (30) Lukovits, I. The Detour Index. *Croat. Chem. Acta* **1996**, *69*, 873–882.
- (31) Nikolić, S.; Trinajstić, N.; Mihalić, Z. The Detour Matrix and the Detour Index. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 279–306.
- (32) Schultz, H. P. Topological Organic Chemistry. 13. Transformation of Graph Adjacency Matrixes to Distance Matrixes. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (5), 1158–1159.
- (33) Gutman, I. A New Method for the Calculation of the Wiener Number of Acyclic Molecules. *J. Mol. Struct. (THEOCHEM)* **1993**, *285*, 137–142.
- (34) Gutman, I. Formula for the Wiener Number of Trees and Its Extension to Graphs Containing Cycles. *Graph. Theory Notes. NYAS* **1994**, *27*, 9–15.
- (35) Garfield, E. R.; Robinson, R. W.; Rouvray, D. H. Determination of the Wiener Molecular Branching Index for the General Trees. *J. Comput. Chem.* **1985**, *6*, 598–609.
- (36) Bonchev, D.; Klein, D. J. On the Wiener Number of Thorn Trees, Stars, Rings and Rods. *Croat. Chem. Acta* **2002**, *75* (2), 613–620.
- (37) Dobrynin, A. A.; Entringer, R.; Gutman, I. Wiener Index of Trees: Theory and Applications. *Acta Appl. Math.* **2001**, *66*, 211–249.
- (38) Dobrynin, A. A.; Gutman, I.; Klavzar, S.; Zigert, P. Wiener Index of Hexagonal Systems. *Acta Appl. Math.* **2002**, *72*, 247–294.
- (39) Smolenskii, E. A. The Wiener Distance Matrix for Acyclic Compounds and Polymers. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (2), 522–528.
- (40) Ivanciuc, O.; Ivanciuc, T. Matrices and Structural Descriptors Computed from Molecular Graph Distances. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 221–277.
- (41) (a) CCL.net. Computational Chemistry List. <http://www.ccl.net/ccal/documents/molecular-modeling/node3.shtml>. (b) Molconn-Z <http://www.eslc.vabiotech.com/molconn/>. (c) The Chemistry Development Kit (CDK), <http://cdk.sourceforge.net/>. (d) Dalke Scientific. Connection matrix vs node with edges for molecules, http://www.dalke-scientific.com/writings/diary/archive/2003/09/21/connection_matrix.html (e) Bock, J. R.; Gough, D. A. A New Method to Estimate Ligand–Receptor Energetics. *Molecular Cellular Proteomics* **2002**, *1*, 904–910.
- (42) (a) MDL Information Systems. www.mdll.com/downloads/public/ctfile/ctfile.pdf. (b) www.chemaxon.com/marvin/doc/user/mol-csmol-doc.html.
- (43) Figueras, J. Ring Perception Using Breadth-First Search. *J. Chem Inf. Comput. Sci.* **1996**, *36*, 986–991.
- (44) Rumbaugh, J.; Jacobson, I.; Booch, G. *The Unified Modeling Language. Reference Manual*. Addison-Wesley Longman Inc., USA, 1999.
- (45) Preiss B. R. *Data Structures and Algorithms with Object-Oriented Design Patterns in Java*. John Wiley & Sons: 2000.
- (46) Rucker, G.; Rucker, C. Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relations among Atoms into Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 187–191.
- (47) Estrada, E.; Guevara, N.; Gutman, I. Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 428–431.
- (48) Luque Ruiz, I.; Cerruela García, G.; Gómez-Nieto, M. A. Representation of the Molecular Topology of Cyclical Structures by means of Cycle Graphs: 3. A Hierarchical Model of Screening of Chemical Databases. *J. Chem. Inf. Comput. Sci.* (Accepted for publication in ASAP).
- (49) Luque Ruiz, I.; Cerruela García, G.; Gómez-Nieto, M. A. Representation of the Molecular Topology of Cyclical Structures by means of Cycle Graphs: 4. A Screening Tool of Chemical Databases. *J. Chem. Inf. Comput. Sci.* (To be submitted.)

CI049788L