

Novel 2D TOMOCOMD-CARDD Descriptors: Atom-based Stochastic and non-Stochastic Bilinear Indices and their QSPR Applications

Yovani Marrero Ponce^{1,2*} & Francisco Torrens²

¹Department of Pharmacy, Faculty of Chemistry-Pharmacy and Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

²Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (València), Spain.

 Fax: 53-42-281130, 281455  Phone: 53-42-281192, 281473  e-mails:
ymarrero77@yahoo.es; yovani.marrero@uv.es; ymponce@gmail.com or
yovanimp@qf.uclv.edu.cu

ABSTRACT

Novel atom-based molecular descriptors based on a bilinear map similar to use defined in linear algebra are presented. These molecular descriptors, called "local (atom, group and atom-type) and total (global) bilinear indices", are proposed here as a new molecular parametrization easily calculated from the 2D molecular information. The proposed non-stochastic and stochastic molecular fingerprints try to match molecular structure provided by the molecular topology by using the k^{th} non-stochastic (Marrero-Ponce, Y. J. Chem. Inf. Comput. Sci. **2004**, 44, 2010 and Marrero-Ponce, Y. *Molecules* **2003**, 8, 687) and stochastic (Marrero-Ponce, Y., et al. *J. Mol. Struc. (Theochem)* **2005**, 717, 67 and Marrero-Ponce, Y.; Castillo-Garit, J. A. *J. Comput.-Aided Mol. Des.* DOI: DO00017575) graph-theoretic electronic-density matrices, M^k and S^k , respectively. That is to say, the k^{th} non-stochastic and stochastic bilinear indices are calculated using M^k and S^k as matrix operators of bilinear transformations. Moreover, chemical information is codified by using different pair combinations of atomic weightings (atomic mass, polarizability, van der Waals volume, and electronegativity). The prediction ability in Quantitative Structure-Property Relationships (QSPR) of the new molecular descriptors was tested by analysing regressions of these descriptors for six selected properties of octane isomers. It was clearly demonstrated that prediction ability was higher than those showed by other 2D/3D well-known sets of molecular descriptors. The obtained results suggest that with the present method it is possible to obtain a good estimation of these physicochemical properties for octanes. The approach described in the present report appears to be a prominent method to find quantitative models for description of physicochemical and biological properties.

Keywords: TOMOCOMD-CARDD Software, Non-Stochastic and Stochastic Atom-Based Bilinear Indices, QSPR Model, Physicochemical Properties, Octane Isomers.

1. INTRODUCTION

Molecular descriptors (MDs) have deserved more and more attention from chemists along the later years.¹ In this connection, a large number of QSAR/QSPR studies have been reported in recent literature that use MDs in prediction of the physicochemical and

biological properties of molecules. MDs are numbers that characterize a specific aspect of molecular structure. The important common feature to all those MDs is the independence of their numerical values on renumbering atoms in a chemical structure.^{3,4}

At present, there are a great number of MDs that can be used in QSAR/QSPR and drug-design studies.³⁻⁵ Among them, the so-called topological indices (TIs) are among most useful MDs known nowadays.⁶⁻⁸ TIs are MDs derived from graph-theoretical invariants and codify structural information contained in 'molecular connectivity'.⁶⁻⁹

The introduction of new TIs has focused the efforts of the scientists working on molecular topology.¹⁰⁻¹² In fact, our research group has proposed several atom- and bond-based topological and topographic MDs. In this sense, Estrada's group has defined several bond-based MDs using the Randić-type graph-theoretical invariant, spectral moment (trace sum)-theoretical invariant and so on.¹²⁻¹⁹ On the other hand, González and co-workers have developed a new method based on the Markov chain theory, which has been successfully employed in QSPR and QSAR studies.²⁰⁻²³ More recently, one of the present authors (M-P. Y) has introduced new sets of total and atom-level MDs relevant to QSAR/QSPR studies and 'rational' drug design, atom-based quadratic $q_k(x)$ and linear indices $f_k(x)$.²⁴⁻²⁷ These MDs are based on the calculation of quadratic and linear maps similar to those defined in linear algebra.²⁸ Our research team has focused its efforts mainly on their application in drug design, so that many new drug(lead)-like compounds within different pharmacological skills have been found out.²⁹⁻³⁵ This approach has been successfully applied to the prediction of several physicochemical and pharmacokinetic properties of organic compounds.^{25,36-40} In addition, these MDs have been extended to consider three-dimensional features of small/medium-sized molecules based on the

trigonometric-3D-chirality-correction factor approach.^{27,41} Finally, promising results have been found in the modeling of the interaction between drugs and HIV Ψ -RNA packaging-region in the field of bioinformatics using macromolecular indices.^{42,43} An alternative formulation of our approach for structural characterization of proteins was also carried out recently.^{44,45}

The main purpose of the present paper is to present new sets of MDs, namely *non-stochastic and stochastic bilinear indices* and establish their abilities (both total and local) for the description of the molecular structure by correlating them with six selected physicochemical properties of octane isomers.

2. METHODOLOGY

In previous reports, we outline outstanding features concerned with the theory of 2D atom-based TOMOCOMD-CARDD descriptors. This method codifies the molecular structure by means of mathematical quadratic, linear and bilinear transformations. In order to calculate these algebraic maps for a molecule, the atom-based molecular vector, \bar{x} (vector representation) and k^{th} “non-stochastic and stochastic graph-theoretic electronic-density matrices”, \mathbf{M}^k and \mathbf{S}^k correspondingly (matrix representations), are constructed.²⁴⁻⁴¹ Such atom-adjacency relationships and chemical-information codification will be applied in the present report to generate a series of atom-based MDs, *atom, group and atom-type as well as total bilinear indices*, to be used in drug design and chemoinformatic studies.

Therefore the structure of this section will be as follows: 1) a background in atom-based molecular vector and non-stochastic and stochastic graph-theoretic electronic-density matrices will be described in the next subsections (2.1 and 2.2, respectively), and 2) an outline of the mathematical definition of bilinear maps and a definition of our procedures will be develop in subsections 2.3 and 2.4, correspondingly.

2.1. Chemical Information and Atom-based Molecular Vector

The atom-based molecular vector (\bar{x}) used to represent small-to-medium size organic chemicals have been explained elsewhere in some detail.²⁴⁻⁴¹ The components (x) of \bar{x} are numeric values, which represent a certain standard atomic property (atom label). That is to say, these weights correspond to different atom properties for organic

molecules. Thus, a molecule having 5, 10, 15,..., n atomic nuclei can be represented by means of vectors, with 5, 10, 15,..., n components, belonging to the spaces \mathfrak{R}^5 , \mathfrak{R}^{10} , \mathfrak{R}^{15} , ..., \mathfrak{R}^n , respectively; where n is the dimension of the real set (\mathfrak{R}^n). That is to say, \bar{x} is the n -dimensional property vector of the atoms (atomic nuclei) in a molecule.

This approach allows us to encode organic molecules such as 3-mercapto-pyridine-4-carbaldehyde through the molecular vector $\bar{x} = [x_{N1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}, x_{C6}, x_{C7}, x_{O8}, x_{S9}]$ (see also Table 1 for molecular structure). This vector belongs to the product space \mathfrak{R}^9 . However, diverse kinds of atomic weights (x) can be used for codifying information related to each atomic nucleus in the molecule. These atomic labels are chemically meaningful numbers such as atomic Log P,⁴⁶ surface contributions of polar atoms,⁴⁷ atomic molar refractivity,⁴⁸ atomic hybrid polarizabilities,⁴⁹ Gasteiger-Marsilli atomic charge,⁵⁰ atomic masses (M),⁵¹ the van der Waals volumes (V),⁵¹ the atomic polarizabilities (P),⁵¹ atomic electronegativity (E) in Pauling scale⁵² and so on.

Table 1 (click here)

Now, if we are interested to codify the chemical information by means of two different molecular vectors, for instance, $\bar{x} = [x_1, \dots, x_n]$ and $\bar{y} = [y_1, \dots, y_n]$; then different combinations of molecular vectors ($\bar{x} \neq \bar{y}$) are possible when a weighting scheme is used. In the present report, we characterized each atomic nucleus with the following parameters: atomic masses (M),⁵¹ the van der Waals volumes (V),⁵¹ the atomic polarizabilities (P),⁵¹ and atomic electronegativity (E) in Pauling scale.⁵² The values of these atomic labels are shown in Table 2. From this weighting scheme, six (or twelve if $\bar{x}_M - \bar{y}_V \neq \bar{x}_V - \bar{y}_M$) combinations (pairs) of molecular vectors ($\bar{x}, \bar{y}; \bar{x} \neq \bar{y}$) can be computed, $\bar{x}_M - \bar{y}_V$, $\bar{x}_M - \bar{y}_P$, $\bar{x}_M - \bar{y}_E$, $\bar{x}_V - \bar{y}_P$, $\bar{x}_V - \bar{y}_E$, and $\bar{x}_P - \bar{y}_E$. Here, we used the symbols $\bar{x}_W - \bar{y}_Z$, where the subscripts W and Z mean two atomic properties from our weighting scheme and a hyphen (-) expresses the combination (pair) of two selected atom-label chemical properties. In order to illustrate this, let us consider the same organic molecule as in the example above (3-mercapto-pyridine-4-carbaldehyde) and the following weighting scheme: M and V ($\bar{x}_M - \bar{y}_V = \bar{x}_V - \bar{y}_M$). The following molecular vectors, $\bar{x} = [14.01, 12.01, 12.01, 12.01, 12.01, 12.01, 12.01, 16.0, 32.07]$ and $\bar{y} = [15.599, 22.449, 22.449, 22.449, 22.449, 22.449, 22.449, 11.494, 24.429]$ are obtained when we use M and V as chemical weights for codifying each atom in the example molecule in \bar{x} and \bar{y} vectors, respectively.

Table 2 (click here)

2.2. Background in Non-Stochastic and Stochastic Graph–Theoretic Electronic-Density Matrices

In molecular topology, molecular structure is expressed, generally, by the hydrogen-suppressed graph. That is, a molecule is represented by a graph. Informally a graph G is a collection of vertices (points) and edges (lines or bonds) connecting these vertices.⁵³⁻⁵⁵ In more formal terms, a simple graph G is defined as an ordered pair $[V(G), E(G)]$ which consists of a nonempty set of vertices $V(G)$ and a set $E(G)$ of unordered pairs of elements of $V(G)$, called edges.⁵³⁻⁵⁵ In this particular case we are not dealing with a simple graph but with a so-called pseudograph (G). Informally, a pseudograph is a graph with multiple edges or loops between the same vertices or the same vertex. Formally: a pseudograph is a set V of vertices along a set E of edges, and a function f from E to $\{\{u,v\} | u,v \text{ in } V\}$ (The function f shows which pair of vertices are connected by which edge). An edge is a loop if $f(e) = \{u\}$ for some vertex u in V .^{24,25,56}

In the earlier reports we have introduced new molecular matrices that describe changes along the time in the electronic distribution throughout the molecular backbone. The $n \times n$ k^{th} non-stochastic graph–theoretic electronic-density matrix of the molecular pseudograph (G), \mathbf{M}^k , is a square and symmetric matrix, where n is the number of atoms (atomic nuclei) in the molecule.²⁴⁻⁴¹ The coefficients ${}^k m_{ij}$ are the elements of the k^{th} power of $\mathbf{M}(G)$ and are defined as follows:²⁴⁻⁴¹

$$m_{ij} = P_{ij} \text{ if } i \neq j \text{ and } \exists e_k \in E(G)$$

(1)

$$= L_{ii} \text{ if } i = j$$

$$= 0 \text{ otherwise}$$

where $E(G)$ represents the set of edges of G . P_{ij} is the number of edges (bonds) between vertices (atomic nuclei) v_i and v_j , and L_{ii} is the number of loops in v_i .

The elements $m_{ij} = P_{ij}$ of such a matrix represent the number of chemical bonds between an atomic nucleus i and other j . The matrix \mathbf{M}^k provides the numbers of walks of length k that link every pair of vertices v_i and v_j . For this reason, each edge in \mathbf{M}^1

represents 2 electrons belonging to the covalent bond between atomic nuclei i and j ; e.g. the inputs of \mathbf{M}^1 are equal to 1, 2 or 3 when single, double or triple bonds, correspondingly, appears between vertices v_i and v_j . On the other hand, molecules containing aromatic rings with more than one canonical structure are represented by a pseudograph. It happens for substituted aromatic compounds such as pyridine, naphthalene, quinoline, and so on, where the presence of pi (π) electrons is accounted by means of loops in each atomic nucleus of the aromatic ring. Conversely, aromatic rings having only one canonical structure, such as furan, thiophene and pyrrol are represented by a multigraph. In order to illustrate the calculation of these matrices, let us consider the same molecule selected in the previous section. Table 1 depicts the molecular structure of this compound and its labeled molecular pseudograph. The zero ($k = 0$), first ($k = 1$), second ($k = 2$) and third ($k = 3$) powers of the non-stochastic graph-theoretic electronic-density matrices are also given in this Table.

As can be seen, \mathbf{M}^k are graph-theoretic electronic-structure models, like an “extended Hückel theory (EHT) model”. The \mathbf{M}^1 matrix considers all valence-bond electrons (σ - and π -networks) in one step and its power ($k = 0, 1, 2, 3, \dots$) can be considered as interacting-electron chemical-network models in k step. The complete model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.⁵⁷

The present approach is based on a simple model for the intramolecular movement of all outer-shell electrons. Let us consider a hypothetical situation in which a set of atoms is free in space at an arbitrary initial time (t_0). At this time, the electrons are distributed around the atomic nuclei. Alternatively, these electrons can be distributed around cores in discrete intervals of time t_k . In this sense, the electron in an arbitrary atom i can move (step-by-step) to other atoms at different discrete time periods t_k ($k = 0, 1, 2, 3, \dots$) throughout the chemical-bonding network.

On the other hand, the k^{th} stochastic graph-theoretic electronic-density matrix of G , \mathbf{S}^k , can be directly obtained from \mathbf{M}^k . Here, $\mathbf{S}^k = [{}^k s_{ij}]$, is a square matrix of order n ($n =$ number of atomic nuclei) and the elements ${}^k s_{ij}$ are defined as follows:^{30,31,34,35}

$${}^k S_{ij} = \frac{{}^k m_{ij}}{{}^k \text{SUM}_i} = \frac{{}^k m_{ij}}{{}^k \delta_i}$$

(2)

where, ${}^k m_{ij}$ are the elements of the k^{th} power of \mathbf{M} and the SUM of the i th row of \mathbf{M}^k are named the k -order vertex degree of atom i , ${}^k \delta_i$. It should be remarked that the matrix \mathbf{S}^k in Eq. 2 has the property that *the sum of the elements in each row* is 1. An $n \times n$ matrix with nonnegative entries having this property is called a “stochastic matrix”.²⁸ The k^{th} s_{ij} elements are the transition probabilities with the electrons moving from atom i to j in the discrete time periods t_k . It should be also pointed out that k^{th} element s_{ij} takes into consideration the molecular topology in k step throughout the chemical-bonding (σ - and π -) network. In this sense, the ${}^2 s_{ij}$ values can distinguish between hybrid states of atoms in bonds. For instance, the self-return probability of second order (${}^2 s_{ii}$) [i.e., the probability with which electron returns to the original atom at t_2], varies regularly according to the different hybrid states of atom i in the molecule, e.g. an electron will have a higher probability of returning to the sp C atom than to the sp^2 (or sp^3) C atom in t_2 [$p(\text{C}_{sp}) > p(\text{C}_{sp^2}) > p(\text{C}_{sp^2 \text{ arom}}) > p(\text{C}_{sp^3})$] (see Table 1 for more details). This is a logical result if the electronegativity scale of these hybrid states is taken into account.

2.3. Mathematical Bilinear Forms: A Theoretical Framework

In mathematics, a **bilinear form** in a real vector space is a mapping $b : V \times V \rightarrow \mathfrak{R}$, which is linear in both arguments.⁵⁸⁻⁶³ That is, this function satisfies the following axioms for any scalar α and any choice of vectors $\bar{v}, \bar{w}, \bar{v}_1, \bar{v}_2, \bar{w}_1$ and \bar{w}_2 .

- i. $b(\alpha \bar{v}, \bar{w}) = b(\bar{v}, \alpha \bar{w}) = \alpha b(\bar{v}, \bar{w})$
- ii. $b(\bar{v}_1 + \bar{v}_2, \bar{w}) = b(\bar{v}_1, \bar{w}) + b(\bar{v}_2, \bar{w})$
- iii. $b(\bar{v}, \bar{w}_1 + \bar{w}_2) = b(\bar{v}, \bar{w}_1) + b(\bar{v}, \bar{w}_2)$

That is, b is *bilinear* if it is linear in each parameter, taken separately.

Let V be a real vector space in \mathfrak{R}^n ($V \in \mathfrak{R}^n$) and consider that the following vector set, $\{\bar{e}_1, \bar{e}_2, \dots, \bar{e}_n\}$ is a basis set of \mathfrak{R}^n . This basis set permits us to write in unambiguous form any vectors \bar{x} and \bar{y} of V , where $(x^1, x^2, \dots, x^n) \in \mathfrak{R}^n$ and $(y^1, y^2, \dots, y^n) \in \mathfrak{R}^n$ are the coordinates of the vectors \bar{x} and \bar{y} , respectively. That is to say,

$$\bar{x} = \sum_{i=1}^n x^i \bar{e}_i \tag{3}$$

and,

$$\bar{y} = \sum_{i=1}^n y^i \bar{e}_i$$

(4)

Subsequently,

$$b(\bar{x}, \bar{y}) = b(x^i \bar{e}_i, y^j \bar{e}_j) = x^i y^j b(\bar{e}_i, \bar{e}_j) \quad (5)$$

if we take the a_{ij} as the $n \times n$ scalars $b(\bar{e}_i, \bar{e}_j)$, That is,

$$a_{ij} = b(\bar{e}_i, \bar{e}_j), \text{ to } i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, n \quad (6)$$

Then,

$$b(\bar{x}, \bar{y}) = \sum_{i,j} a_{ij} x^i y^j = [X]^T A [Y] = \begin{bmatrix} x^1 & \dots & x^n \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix} \quad (7)$$

As it can be seen, the defined equation for b may be written as the single matrix equation (see Eq. 7), where $[Y]$ is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{y} in a basis set of \mathfrak{R}^n , and $[X]^T$ (a $1 \times n$ matrix) is the transpose of $[X]$, where $[X]$ is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{x} in the same basis of \mathfrak{R}^n .

Finally, we introduce the formal definition of **symmetric bilinear form**. Let V be a real vector space and b be a bilinear function in $V \times V$. The bilinear function b is called symmetric if $b(\bar{x}, \bar{y}) = b(\bar{y}, \bar{x}), \forall \bar{x}, \bar{y} \in V$.⁵⁸⁻⁶³ Then,

$$b(\bar{x}, \bar{y}) = \sum_{i,j} a_{ij} x^i y^j = \sum_{i,j} a_{ji} x^j y^i = b(\bar{y}, \bar{x})$$

(8)

2.4. Non-Stochastic and Stochastic Atom-Based Bilinear Indices: Total (Global)

Definition

The k^{th} non-stochastic and stochastic bilinear indices for a molecule, $b_k(\bar{x}, \bar{y})$ and ${}^s b_k(\bar{x}, \bar{y})$, are computed from these k^{th} non-stochastic and stochastic graph-theoretic electronic-density matrices, \mathbf{M}^k and \mathbf{S}^k as shown in Eqs. 9 and 10, respectively:

$$b_k(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ij} x^i y^j$$

(9)

$${}^s b_k(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ij} x^i y^j$$

(10)

where n is the number of atoms in the molecule, and x^1, \dots, x^n and y^1, \dots, y^n are the coordinates or components of the molecular vectors \bar{x} and \bar{y} in a canonical basis set of \mathfrak{R}^n .

The defined equations (9) and (10) for $b_k(\bar{x}, \bar{y})$ and ${}^s b_k(\bar{x}, \bar{y})$ may be also written as the single matrix equations:

$$b(\bar{x}, \bar{y}) = [X]^T \mathbf{M}^k [Y]$$

(11)

$${}^s b(\bar{x}, \bar{y}) = [X]^T \mathbf{S}^k [Y] \quad (12)$$

Where $[Y]$ is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{y} in the canonical basis set of \mathfrak{R}^n , and $[X]^T$ is the transpose of $[X]$, where $[X]$ is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{x} in the canonical basis of \mathfrak{R}^n . Therefore, if we use the canonical basis set, the coordinates $[(x^1, \dots, x^n)$ and $(y^1, \dots, y^n)]$ of any molecular vectors (\bar{x} and \bar{y}) coincide with the components of those vectors $[(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)]$. For that reason, those coordinates can be considered as weights (atomic labels) of the vertices of the molecular pseudograph, due to the fact that components of the molecular vectors are values of some atomic property that characterizes each kind of atomic nuclei in molecule.

It should be remarked that non-stochastic and stochastic bilinear indices are symmetric and non-symmetric bilinear forms, respectively. Therefore, if in the following weighting scheme, M and V are used as atomic weights to compute these MDs, two different sets of stochastic bilinear indices, ${}^{M-V} s \mathbf{b}_k^{\mathbf{H}}(x, y)$ and ${}^{V-M} s \mathbf{b}_k^{\mathbf{H}}(x, y)$ [because $\bar{x}_M \bar{y}_V \neq \bar{x}_V \bar{y}_M$] can be obtained and only one group of non-stochastic bilinear indices (${}^{M-V} s \mathbf{b}_k^{\mathbf{H}}(x, y) = {}^{V-M} s \mathbf{b}_k^{\mathbf{H}}(x, y)$) because in this case $\bar{x}_M \bar{y}_V = \bar{x}_V \bar{y}_M$ can be calculated.

2.5. Non-Stochastic and Stochastic Atom-Based Bilinear Indices: Local (Atomic, Group, and Atom-type) Definition.

In the last decade, Randić⁶⁴ proposed a list of desirable attributes for a MD. Therefore, this list can be considered as a methodological guide for the development of new TIs. One of the most important

criteria is the possibility of defining the descriptors locally. This attribute refers to the fact that the index could be calculated for the molecule as a whole but also over certain fragments of the structure itself.

Sometimes, the properties of a group of molecules are related more to a certain zone or fragment than to the molecule as a whole. Thereinafter, the global definition never satisfies the structural requirements needed to obtain a good correlation in QSAR and QSPR studies. The local indices can be used in certain problems such as:

- Research on drugs, toxics or generally any organic molecules with a common skeleton, which is responsible for the activity or property under study.
- Study on the reactivity of specific sites of a series of molecules, which can undergo a chemical reaction or enzymatic metabolism.
- In the study of molecular properties such as spectroscopic measurements, which are obtained experimentally in a local way.
- In any general case where it is necessary to study not the molecule as a whole, but rather some local properties of certain fragments, then the definition of local descriptors could be necessary.

Therefore, in addition to *total bilinear indices* computed for the whole molecule, a local-fragment (atomic, group or atom-type) formalism can be developed. These descriptors are termed *local non-stochastic and stochastic bilinear indices*,

$b_{kL}(\bar{x}, \bar{y})$ and ${}^s b_{kL}(\bar{x}, \bar{y})$, respectively. The definition of these descriptors is as follows:

$$b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ijL} x^i y^j$$

(13)

$${}^s b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ijL} x^i y^j$$

(14)

Where ${}^k m_{ijL}$ [${}^k s_{ijL}$] is the k^{th} element of the row “ i ” and column “ j ” of the local matrix \mathbf{M}^k_L [\mathbf{S}^k_L]. This matrix is extracted from the \mathbf{M}^k [\mathbf{S}^k] matrix and contains information

referred to the pairs of vertices (atomic nuclei) of the specific molecular fragments and also of the molecular environment in k step. The matrix $\mathbf{M}^k_L [\mathbf{S}^k_L]$ with elements ${}^k m_{ijL}$ [${}^k s_{ijL}$] is defined as follows:

$$\begin{aligned} {}^k m_{ijL} [{}^k s_{ijL}] &= {}^k m_{ij} [{}^k s_{ijL}] \text{ if both } v_i \text{ and } v_j \text{ are atomic nuclei contained within the molecular} \\ &\quad \text{fragment} \\ &= 1/2 {}^k m_{ij} [{}^k s_{ijL}] \text{ if } v_i \text{ or } v_j \text{ is an atomic nucleus contained within the molecular} \\ &\quad \text{fragment but not both} \\ &= 0 \text{ otherwise} \end{aligned}$$

(15)

These local analogues can also be expressed in matrix form by the expressions:

$$b_L(\bar{x}, \bar{y}) = [X]^T \mathbf{M}^k_L [Y] \quad (16)$$

$${}^s b_L(\bar{x}, \bar{y}) = [X]^T \mathbf{S}^k_L [Y] \quad (17)$$

It should be remarked that the scheme above follows the spirit of a Mulliken population analysis.⁶⁵ It should be also pointed out that for every partitioning of a molecule into Z molecular fragments there will be Z local molecular fragment matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrix $\mathbf{M}^k [\mathbf{S}^k]$ can be correspondingly partitioned into Z local matrices $\mathbf{M}^k_L [\mathbf{S}^k_L]$, $L = 1, \dots, Z$, and the k^{th} power of matrix $\mathbf{M} [\mathbf{S}]$ is exactly the sum of the k^{th} power of the local Z matrices. In this way, the total non-stochastic and stochastic bilinear indices are the sum of the non-stochastic and stochastic bilinear indices, respectively, of the Z molecular fragments:

$$b(\bar{x}, \bar{y}) = \sum_{L=1}^Z b_{kL}(\bar{x}, \bar{y}) \quad (18)$$

$${}^s b(\bar{x}, \bar{y}) = \sum_{L=1}^Z {}^s b_{kL}(\bar{x}, \bar{y}) \quad (19)$$

Atomic, group and atom-type bilinear fingerprints are specific cases of local bilinear indices. Atomic bilinear indices, $b_{kL}(\bar{x}_i, \bar{y}_i)$ and ${}^s b_{kL}(\bar{x}_i, \bar{y}_i)$, can be computed for each atom i in the molecule and contain electronic and topological structural information from all other atoms within the structure. The atom-level bilinear indices values for the common scaffold atoms can be directly used as variables in seeking a QSPR/QSAR model as long as these atoms are numbered in the same way in all molecules in the database.

In addition, the atom-type bilinear indices can also be calculated. In the same way as atom-type E-state values,¹¹ for all data sets (including those with a common skeletal

core as well as those with very diverse structures), these novel local MDs provide much useful information. That is, this approach provides the basis for application to a wider range of problems to which the atomic bilinear indices formalism is applicable without the need for superposition.^{66,67} For this reason the present method represents a significant advantage over traditional QSAR methods. The atom-type bilinear descriptors are calculated by adding the k^{th} atomic bilinear indices for all atoms of the same type in the molecule. This atom type index lends itself to use in a group additive-type scheme in which an index appears for each atom type in the molecule. In the atom-type bilinear indices formalism, each atom in the molecule is classified into an atom type (fragment), such as -F, -OH, =O, -CH₃, and so on.^{11,66,67} That is to say, each atom in the molecule is categorized according to a valence-state classification scheme including the number of attached H-atoms.¹¹ The atom-type descriptors combine three important aspects of structural information: 1) Collective electron and topologic accessibility to the atoms of the same type (for each structural feature: atom or hybrid group such as -Cl, =O, -CH₂-, etc), 2) presence/absence of the atom type (structural features), and 3) count of the atoms in the atom-type sets.

Finally, these local MDs can be calculated by a chemical (or functional) group in the molecule, such as heteroatoms (O, N and S in all valence states and including the number of attached H-atoms), hydrogen bonding (H-bonding) to heteroatoms (O, N and S in all valence states), halogen atoms (F, Cl, Br and I), all aliphatic carbon chains (several atom types), all aromatic atoms (aromatic rings), and so on. The group-level bilinear indices are the sum of the individual atom-level bilinear indices for a particular group of atoms. For all data set structures, the k^{th} group-based bilinear indices provide important information for QSAR/QSPR studies.

2.6. Sample Calculation

It is useful to perform a calculation on a molecule to illustrate the effect of structure on atomic and global bilinear indices values. For this we use the 3-mercapto-pyridine-4-carbaldehyde molecule. The labeled (atom numbering) molecular structure of this chemical and the non-stochastic and stochastic (atom-level, group and atom-type as well as total) atom-based bilinear indices are shown at the Table 3.

Table 3 (click here)

3. COMPUTATIONAL STRATEGIES

All computations were carried out on a PC Pentium-4 2.0 GHz. The TOMOCOMD program for Windows package developed in our laboratory was used for computing the molecular descriptors for the dataset of compounds. This software is an interactive program for molecular design and bioinformatic research.⁶⁸ It is composed of four subprograms; each one of them allows both drawing the structures (drawing mode) and calculating molecular 2D/3D descriptors (calculation mode). The modules are named CARDD (Computed-Aided 'Rational' Drug Design), CAMPS (Computed-Aided Modeling in Protein Science), CANAR (Computed-Aided Nucleic Acid Research) and CABPD (Computed-Aided Bio-Polymers Docking). In the present report, we outline salient features concerned with only one of these subprograms, CARDD and with the calculation of non-stochastic and stochastic 2D atom-based bilinear indices.

The main steps for the application of the present method in QSAR/QSPR and drug design can be summarized briefly in the following algorithm: 1) Draw the molecular structure for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol belonging to the different groups in the periodic table of the elements; 2) Use appropriate weights in order to differentiate the atoms in the molecule. The weights used in this work are those previously proposed for the calculation of the DRAGON descriptors,^{51,69,70} i.e., atomic mass (M), atomic polarizability (P), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (E). The values of these atomic labels are shown in Table 2;^{51,52,69,70} 3) compute the total and local (atomic, group and atom-type) non-stochastic and stochastic bilinear indices. It can be carried out in the software calculation mode, where one can select the atomic properties and the descriptor family before calculating the molecular indices. This software generates a table in which the rows correspond to the compounds, and columns correspond to the atom-based (both total and local) bilinear maps or other MDs family implemented in this program; 4) Find a QSPR/QSAR equation by using several multivariate analytical techniques, such as multilinear regression analysis (MRA), neural networks, linear discrimination analysis, and so on. That is to say, we can find a quantitative relation between a property **P** and the bilinear fingerprints having, for instance, the following appearance,

$$\mathbf{P} = a_0\mathbf{b}_0(x,y) + a_1\mathbf{b}_1(x,y) + a_2\mathbf{b}_2(x,y) + \dots + a_k\mathbf{b}_k(x,y) + c \quad (20)$$

where \mathbf{P} is the measured property, $\mathbf{b}_k(x,y)$ are the k^{th} non-stochastic total bilinear indices, and the a_k 's are the coefficients obtained by the MRA; 5) test the robustness and predictive power of the QSPR/QSAR equation by using internal cross-validation techniques.

The atom-based *TOMOCOMD-CARDD* MDs computed in this study were the following:

- i) k^{th} ($k = 15$) total (global) non-stochastic bilinear indices not considering and considering H-atoms in the molecule [$\mathbf{b}_k(x,y)$ and $\mathbf{b}_k^{\text{H}}(x,y)$, respectively].
- ii) k^{th} ($k = 15$) total (global) stochastic bilinear indices considering H-atoms in the molecule [${}^s\mathbf{b}_k^{\text{H}}(x,y)$].
- iii) k^{th} ($k = 15$) group (methyl group, $-\text{CH}_3$) non-stochastic and stochastic bilinear indices considering H-atoms in the molecule. [$\mathbf{b}_{kL(\text{CH}_3)}^{\text{H}}(x,y)$ and ${}^s\mathbf{b}_{kL(\text{CH}_3)}^{\text{H}}(x,y)$, respectively]. These local descriptors are calculated taken into account only one of the three bond types for carbon-hydrogen bonds ($\text{C}_{\text{primary}}\text{-H}$) that there are for octanes data.

4. DATA SETS

The data sets for this study were taken from Consonni et al.,⁶⁹ because the data of physicochemical properties of octane isomers have been carefully selected for testing MDs. Octanes constituted a good set of chemicals for comparative study, since many experimental data among their physicochemical properties and QSPR studies are available.^{69,71-76} In this sense, we analyzed the quality of the obtained QSPR models to describe the boiling point (BP), motor octane number (MON), heat of vaporization (HV), molar volume (MV), entropy (S), and heat of formation ($\Delta_f\text{H}$) of the octane isomers.

The use of octanes as a very suitable data set for testing TIs have been advocated by Randić and Trinajstić.^{77,78} This selection is recommended due to the most of the fact that physicochemical properties commonly studied in QSPR analyses with TIs are interrelated for data sets of compounds with different molecular weights, for instance for alkanes with two to nine carbon atoms. These correlations are not necessarily observed when the same indices are used in isomeric data sets of compounds, such as the octane data set. In addition, these properties are hardly interrelated when octanes are used as a data set.⁷⁴ On the other hand; all the TIs are designed to have (gradual)

increments with the increase in the molecular weight. By this way, if we do the present study by using a series of compounds having different molecular weights, we will find “false” interrelations between the indices by an overestimation of the size effects inherent to these descriptors.⁷¹ The same is also valid when the QSPR model is to be obtained. It is not difficult to find “good” linear correlations between TIs and physicochemical properties of alkanes in data sets with great size variability.⁷¹ In fact, the simple use of the number of vertices in the molecular graph produced regression coefficients greater than 0.97 for most of the physicochemical properties of C2-C9 alkanes studied by Needham et al.⁷⁹ However, when data sets of isomeric compounds are considered, correlations that typically have high correlation coefficients when molecules of different sizes are considered will no longer show such good linear correlation. In conclusion, if a new proposed MD is not able to model the variation of at least one property of octane isomers, then it probably does not contain any useful molecular information.

5. STATISTICAL METHOD

The k^{th} total and local atom-based bilinear indices were used as molecular descriptors for derived QSPRs. One of the difficulties with the large number of descriptors is deciding which ones will provide the best regressions. Furthermore, as testing a large number of all possible combinations of variables would be a tedious task and time-consuming procedure, we have used a genetic algorithm (GA) input selection.⁸⁰⁻⁸⁵

GAs are a class of algorithms inspired by the process of natural evolution in which species having a high fitness under some conditions can prevail and survive to the next generation; the best species can be adapted by crossover and/or mutation in the search for better individuals. Genetic function approximation (GFA), a combination of GA and the linear polynomials, higher-order polynomials, splines (multivariate adaptive regression splines algorithm), or other non-linear functions, provides multiple models with high predictive ability.⁸⁰⁻⁸⁷

The BuildQSAR⁸⁸ software was employed to perform variable selection and QSPR modeling. The mutation probability was specified as 35%. The length of the equations was set for three terms and a constant. The population size was established as 100. The GA with an initial population size of 100 rapidly converged (200 generations) and reached an optimal QSAR model in a reasonable number of GA generations. The search

for the best model can be processed in terms of the highest correlation coefficient (R) or F-test equations (Fisher-ratio's p -level [$p(F)$]), and the lowest standard deviation equations (s).⁸⁸ The quality of models was also determined by examining the Leave-One-Out (LOO) cross-validation (CV) (q^2 , s_{cv}).⁸⁹

6. QSPR APPLICATION OF PHYSICOCHEMICAL PROPERTIES OF OCTANE ISOMERS

The decisive criterion of quality for any MD is its ability to describe structure-related properties of molecules. In order to illustrate the applicability of the novel TOMOCOMD-CARDD descriptors, we performed the QSPR models to describe six physicochemical properties of octane isomers. In this sense, regressions of octane properties based on the non-stochastic and stochastic bilinear indices will be compared to some regressions based on 2D (topologic) and 3D (geometric) descriptors taken from the literature.^{69,71-76} Precisely, to evaluate the quality of the models based on our novel atom-based MDs we have taken as references: 1) the models published by Randić⁷²⁻⁷⁴ based on diverse TIs such as the Wiener matrix invariants, 2) the equation published by Diudea⁷⁶ based on the SP indices, and 3) the best models obtained from a set constituted by the topological (69), WHIM (99), and GETAWAY descriptors (197).⁶⁹

The best linear models found using non-stochastic and stochastic total and atom-type bilinear indices are presented in Table 4, together with the statistical information for the best regressions with 1, 2, and 3 MDs published so far^{69,71-76} for each selected property of octane isomers.

Table 4 (click here)

As it can be appreciated from the statistical parameters of regression equations in Table 4, all of the physicochemical properties were well described by atom-based bilinear indices. In this table we can observe that the statistical parameters for the models obtained with atom-level bilinear indices to describe heat of vaporization (HV) (Eqs 25 and 26), molar volume (MV) (Eqs 27 and 28) and heat of

formation ($\Delta_f H$) (Eqs **31** and **32**) of octanes are better than those taken from the literature using 2D and 3D MDs. The second physicochemical property, that is MV, is well described exclusively by the bond-based linear indices (both non-stochastic and stochastic ones). In addition, the obtained QSPR model (Eq. **32**) for description of $\Delta_f H$ using only one stochastic linear index [$^{V-Es}b_5^H(x,y)$] showed better statistical results and predictive power than the equations developed from 2D and 3D (and their combination) MDs.

It should be pointed out that in the models based on the bond-level chemical linear indices, both regressions for the motor octane number (MON) (Eqs **23** and **24**) are better-to-similar than the models published so far. Only the models found by us to describe boiling point (BP) (Eqs **21** and **22**) and entropy (S) (Eqs **29** and **30**) have significant differences with the preceding models obtained by applying the selection procedure to the set given by GETAWAY descriptors plus WHIM and topological indices. However, these properties were described with our approach better than using several TIs.

According to the obtained QSPR results, it is possible to conclude that the novel MDs encode some useful molecular information different from that of previously proposed descriptors. Moreover, they are quite diverse among themselves being able to describe well the variation of different properties of octanes.

7. CONCLUSIONS

The approach described in this report appears to be a prominent method to find quantitative models for description of physical, thermodynamic, or biological properties. The novel MDs proposed here have been shown to have some interesting features, such as:

- i. Their functional definitions are based on well-known and accepted algorithms and formulas in mathematics. These novel atom-based molecular descriptors are based on a bilinear map similar to those defined in linear algebra. The atom- and group-level as well as atom-type formalism will permit to expedite investigation of molecular mechanisms and rational design of molecules at the local level.

These local bilinear indices together with global ones are now added as a new set of MDs to the significant arsenal of whole-molecule indices.

- ii. They can be easily and quickly calculated. That is, the calculation is simple and straightforward, requiring only the 2D information.
- iii. They show good prediction power in physicochemical-properties modeling. The obtained QSPR models presented here yield a good statistical account of the octane data, giving both sound direct statistical as well as the press statistics. Further, it was clearly demonstrated that prediction ability was higher than those showed by other 2D/3D well-known sets of MDs.

Despite these positive features of non-stochastic and stochastic atom-based bilinear indices, additional study has to be done to further investigate their meaning and behaviour with respect to the structural features of the molecules. The applications of the present method to QSPR/QSAR and drug-design studies as well as similarity/diversity analysis of several classes of organic compounds are now in progress and will be the subject of a future publication.

Acknowledgements: One of the authors (M-P. Y) thanks the program 'Estades Temporals per a Investigadors Convidats' for a fellowship to work at Valencia University. F. T. acknowledges financial support from the Spanish MEC DGI (Project No.CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173).

8. REFERENCES AND NOTES

- (1) Galvez, J. Prediction of Molecular Volume and Surface of Alkanes by Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1231-1239.
- (2) Seybold, P., G.; May, M.; Bagal, U., A. Molecular Structure-Property Relationships. *J. Chem. Educ.* **1987**, *64*, 575-581.
- (3) Karelson, M. *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons: New York. 2000.
- (4) Todeschini, R.; Consonni, V. (Eds.) *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim (Germany), 2000.
- (5) Randić, M. On Characterization of Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672-672.
- (6) Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1699-1714.
- (7) *Topological Indices and Related Descriptors in QSAR and QSPR*, Devillers, J.; Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999.
- (8) Torrens, F. Structural, Chemical Topological, Electrotological and Electronic Structure Hypotheses. *Comb. Chem. High Throughput Screen.* **2003**, *6*, 801-809.
- (9) Galvez, J.; Garcia, R.; Salabert, M., T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520-525.

- (10) Rouvray, D. H. Should We Have Designs on Topological Indexes? *Stud. Phys. Theor. Chem.* **1983**, *28* (*Chem. Appl. Topol. Graph Theory*), 159-177.
- (11) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999.
- (12) Estrada, E. Generalization of Topological Indices. *Chem Phys. Lett.* **2001**, *336*, 248-252.
- (13) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31-33.
- (14) Estrada, E.; Ramírez, A. Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837-843.
- (15) Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844-849.
- (16) Marković, S.; Gutman, I. Dependence of Spectral Moments of Benzenoid Hydrocarbons on Molecular Structure. *J. Mol. Struct. (Theochem)*. **1991**, *235*, 81-87.
- (17) Estrada, E.; Guevara, N.; Gutman, I. Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Application. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 428-431.
- (18) Estrada, E. Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long-Range Bond Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042-1048.
- (19) Estrada, E.; Molina, E. Novel Local (Fragment-Based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design. *J. Mol. Graphics Mod.* **2001**, *20*, 54-64.
- (20) Ramos de, A., R.; González Díaz, H.; Molina, R.; González, M., P.; Uriarte, E. Stochastic-based Descriptors Studying Peptides Biological Properties: Modeling the Bitter Tasting Threshold of Dipeptides. *Bioorg. Med. Chem.* **2004**, *12*, 4815-4822.
- (21) González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernández, S. I.; Morales, A.; Serrano, H., S.; González, J.; Ramos de, A., R. Markovian Chemicals "in silico" Design (MARCH-INSIDE), a Promising Approach for Computer Aided Molecular Design II: Experimental and Theoretical Assessment of a Novel Method for Virtual Screening of Fasciolicides. *J. Mol. Mod.* **2002**, *8*, 237-245.
- (22) González-Díaz, H.; Gia, O.; Uriarte, E.; Hernández, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J., A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L., A.; Cabrera, M. A. Markovian Chemicals "in silico" Design (MARCH-INSIDE), a Promising Approach for Computer-Aided Molecular Design I: Discovery of Anticancer Compounds. *J. Mol. Mod.* **2003**, *9*, 395-407.
- (23) González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H., S.; Ramos de, A., R. Simple Stochastic Fingerprints Towards Mathematical Modelling in Biology and Medicine. 1. The Treatment of Coccidiosis. *Bull. Math. Biol.* **2004**, *66*, 1285-1311.
- (24) Marrero-Ponce, Y. Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Definition, Significance-Interpretation and Application to QSAR Analysis of Flavone Derivatives as HIV-1 Integrase Inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010-2026.
- (25) Marrero-Ponce, Y. Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* **2003**, *8*, 687-726.
- (26) Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix" and Their Stochastic Forms: A Novel Approach for Virtual Screening and *in silico* Discovery of New Lead Paramphistomicide Drugs-like Compounds. *J. Mol. Struct. (Theochem)*. **2005**, *717*, 67-79.
- (27) Marrero-Ponce, Y.; Castillo-Garit, J., A. 3D-Chiral Atom, Atom-type, and Total Non-Stochastic and Stochastic Molecular Linear Indices and their Applications to Central Chirality Codification. *J. Comput.-Aided Mol. Design*. In Press (DOI: DO00017575)
- (28) Edwards, C., H.; Penney, D., E. *Elementary Linear Algebra*; Prentice-Hall, Englewood Cliffs: New Jersey, USA, 1988.
- (29) Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón Pérez, M.; García Sánchez, R. Non-Stochastic and Stochastic Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to "*in silico*" Studies for the Rational Discovery of New Antimalarial Compounds. *Bioorg. Med. Chem.* **2005**, *13*, 1293-1304.
- (30) Marrero-Ponce, Y.; Iyarreta-Veitía, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Ávila, P. E.; Kirchgatter, K.; Machado, Y. Ligand-Based Virtual Screening and *in silico* Design of New Antimalarial Compounds Using Non-Stochastic and Stochastic Total and Atom-type Quadratic Maps. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 1082-1100.
- (31) Marrero-Ponce, Y.; Medina-Marrero, R.; Martínez, Y.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. A. Non-Stochastic and Stochastic Linear Indices of the Molecular Pseudograph's Atom Adjacency Matrix: A Novel Approach for Computational *-in silico-* Screening and "Rational" Selection of New Lead Antibacterial Agents. *J. Mol. Mod.* In Press.
- (32) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; Sánchez, A. M.; Torrens, F.; Castro, E. A. Atom, Atom-Type and Total

- Molecular Linear Indices as a Promising Approach for Bioorganic & Medicinal Chemistry: Theoretical and Experimental Assessment of a Novel Method for Virtual Screening and Rational Design of New Lead Anthelmintic. *Bioorg. Med. Chem.* 2005, 13, 1005-1020.
- (33) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a Novel Approach for Computer-Aided "Rational" Drug Design: I. Theoretical and Experimental Assessment of a Promising Method for Computational Screening and *in silico* Design of New Anthelmintic Compounds. *J. Comput.-Aided Mol. Des.* 2004, 18, 615-633.
- (34) Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. Atom, Atom-type, and Total Non-Stochastic and Stochastic Quadratic Fingerprints: A Promising Approach for Modeling of Antibacterial Activity. *Bioorg. Med. Chem.* 2005, 13, 2881-2899.
- (35) Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R.; García Sánchez, R. N. A Linear Discrimination Analysis Based Virtual Screening of Trichomonacidal Lead-Like Compounds. Outcomes of *in silico* Studies Supported by Experimental Results. *Bioorg. Med. Chem Lett.* 2005, 17, 3838-3843.
- (36) Marrero-Ponce, Y. Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Molecular Descriptors and QSPR/QSAR Applications. *Bioorg. Med. Chem.* 2004, 12, 6351-6369.
- (37) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro E. Atom, Atom-Type and Total Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to QSPR/QSAR Studies of Organic Compounds. *Molecules.* 2004, 9, 1100-1123.
- (38) Marrero-Ponce, Y.; Cabrera, M., A.; Romero, V.; Ofori, E.; Montero, L. A. Total and Local Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix". Application to Prediction of Caco-2 Permeability of Drugs. *Int. J. Mol. Sci.* 2003, 4, 512-536.
- (39) Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. A New Topological Descriptors Based Model for Predicting Intestinal Epithelial Transport of Drugs in Caco-2 Cell Culture. *J. Pharm. Pharmaceut. Sci.* 2004, 7, 186-199.
- (40) Marrero-Ponce, Y.; Cabrera, M. A.; Romero-Zaldivar, V.; Bermejo, M.; Siverio, D.; Torrens, F. Prediction of Intestinal Epithelial Transport of Drug in (Caco-2) Cell Culture from Molecular Structure using 'in silico' Approaches During Early Drug Discovery. *Internet Electronic J. Mol. Des.* 2005, 4, 124-150.
- (41) Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. 3D-Chiral Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix" and their Application to Central Chirality Codification: Classification of ACE Inhibitors and Prediction of σ -Receptor Antagonist Activities. *Bioorg. Med. Chem.* 2004, 12, 5331-5342.
- (42) Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. Nucleic Acid Quadratic Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix". Modeling of Footprints after the Interaction of Paromomycin with the HIV-1 Ψ -RNA Packaging Region. *Int. J. Mol. Sci.* 2004, 5, 276-293.
- (43) Marrero-Ponce, Y.; Castillo-Garit, J.A.; Nodarse, D. Linear Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix" as a Promising Approach for Bioinformatics Studies. 1. Prediction of Paromomycin's Affinity Constant with HIV-1 Ψ -RNA Packaging Region. *Bioorg. Med. Chem.* 2005, 13, 3397-3404.
- (44) Marrero-Ponce, Y.; Medina, R.; Castro, E. A.; de Armas, R.; González, H.; Romero, V.; Torrens, F. Protein Quadratic Indices of the "Macromolecular Pseudograph's α -Carbon Atom Adjacency Matrix". 1. Prediction of Arc Repressor Alanine-mutant's Stability. *Molecules* 2004, 9, 1124-1147.
- (45) Marrero-Ponce, Y.; Medina-Marrero, R.; Castillo-Garit, J. A.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. Protein Linear Indices of the "Macromolecular Pseudograph's α -Carbon Atom Adjacency Matrix" in Bioinformatics. 1. Prediction of Protein Stability Effects of a Complete Set of Alanine Substitutions in Arc Repressor. *Bioorg. Med. Chem.* 2005, 13, 3003-3015.
- (46) Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Perspect. Drug Discov. Des.* 2000, 19, 47-66.
- (47) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* 2000, 43, 3714-3717.
- (48) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* 1987, 27, 21-35.
- (49) Millar, K., J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* 1990, 112, 8533-8542.
- (50) Gasteiger, J.; Marsilli, M., A New Model for Calculating Atomic Charge in Molecules. *Tetrahedron Lett.* 1978, 34, 3181-3184.

- (51) Kier L., B; Hall, L., H. *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press: Letchworth, U. K, 1986.
- (52) Pauling, L. *The Nature of Chemical Bond*, Cornell University Press: Ithaca (New York), 1939; 2-60.
- (53) Rouvray, D., H. In *Chemical Applications of Graph Theory*, Balaban, A. T., Ed.; Academic Press: London, 1976; pp 180-181.
- (54) Trinajstić, N. *Chemical Graph Theory*, CRC Press: Boca Raton, FL, 1983; 2nd ed.; 1992; pp 32-33.
- (55) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986.
- (56) Estrada, E.; Patlewicz, G. On the Usefulness of Graph-theoretic Descriptors in Predicting Theoretical Parameters. Phototoxicity of Polycyclic Aromatic Hydrocarbons (PAHs). *Croat. Chim. Acta.*, 2004, 77, 203-211.
- (57) Klein, D., J. Graph Theoretically Formulated Electronic–Structure Theory. *Internet Electron. J. Mol. Des.* 2003, 2, 814–834.
- (58) Jacobson, N. *Basic Algebra I*. 2nd Edition. Ed. W.H. Freeman and Company, New York. 1985, pp. 343-361.
- (59) Riley, K., F.; Hobson, M., P.; Vence, S., J. *Mathematical Methods for Physics and Engineering*. Ed. Cambridge University Press. 1998, pp 228-236.
- (60) Hernández, E. *Álgebra y Geometría*. Ed. Universidad Autonoma de Madrid, Madrid. 1987, pp 521-544.
- (61) de Burgos-Román, J. *Álgebra y Geometría Cartesiana*. Ed. McGraw-Hill Interamericana de España, 2^{da} Edición. 2000, pp. 208-246.
- (62) de Burgos-Román, J. *Curso de Álgebra y Geometría*. Ed. Alambra Longman, Madrid. 1994, pp. 638-684.
- (63) Werner, G. *Linear Algebra*. 4th Edition. Ed. Springer-Verlag, New York. 1981, pp. 261-288.
- (64) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* 1991, 7, 155-168.
- (65) Walker, P., D.; Mezey, P., G. Molecular Electron Density Lego Approach to Molecule Building. *J. Am. Chem. Soc.* 1993, 115, 12423-12430.
- (66) Hall, L., H.; Story, C., T. Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* 1996, 36, 1004-1014.
- (67) Gough, J., D.; Hall, L., H. Modeling Antileukemic Activity of Carboquinones with Electrotopological State and Chi Indices. *J. Chem. Inf. Comput. Sci.* 1999, 39, 356-361.
- (68) Marrero-Ponce Y, Romero V (2002) TOMOCOMD software. Central University of Las Villas. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es
- (69) Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* 2002, 42, 682-692.
- (70) Todeschini, R.; Gramatica, P. New 3D Molecular Descriptors: The WHIM Theory and QSAR Applications. *Perspect. Drug Dis. Des.* 1998, 9-11, 355–380.
- (71) Estrada, E.; Rodríguez, L. Edge-Connectivity Indices in QSPR/QSAR Studies. 1. Comparison to Other Topological Indices in QSPR Studies. *J. Chem. Inf. Comput. Sci.* 1999, 39, 1037-1041.
- (72) Randić, M. Correlation of Enthalpy of Octanes with Orthogonal Connectivity Indices. *J. Mol. Struct. (Theochem)*. 1991, 233, 45-59.
- (73) Randić, M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chim. Acta* 1993, 66, 289-312.
- (74) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci.* 1994, 34, 361-367.
- (75) Diudea, M., V. Walk Numbers eWM: Wiener-Type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci.* 1996, 36, 535-540.
- (76) Diudea, M., V.; Minailiuc, O. M.; Katona, G. Molecular Topology. 26. SP Indices: Novel Connectivity Descriptors. *Rev. Roum. Chim.* 1997, 42, 239-249.
- (77) Randić, M.; Trinajstić, N. Viewpoint 4-Comparative Structure-Property Studies: The Connectivity Basis. *J. Mol. Struct. (Theochem)*. 1993, 284, 209-221.
- (78) Randić, M.; Trinajstić, N. In Search for Graph Invariants of Chemical Interest. *J. Mol. Struct. (Theochem)*. 1993, 300, 551-572.
- (79) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of Alkanes. *J. Am. Chem. Soc.* 1988, 110, 4186-4194.
- (80) Goldberg, D. E. *Genetic Algorithms*, Addison Wesley, Reading, MA, 1989.
- (81) Willet, P. Genetic Algorithms in Molecular Recognition and Design. *Trends Biotechnol.* 1995, 13, 516–521.
- (82) So, S. S.; Karplus, M. Evolutionary Optimization in Quantitative Structure-Activity Relationship: An Application of Genetic Neural Networks. *J. Med. Chem.* 1996, 39, 1521–1530.
- (83) So, S. S.; Karplus, M. Three-Dimensional Quantitative Structure-Activity Relationship from Molecular Similarity Matrices and Genetic Neural Networks. *J. Med. Chem.* 1997, 40, 4347–4359.

- (84) Rogers, D.; Hopfinger, A., J. Application of Genetic Function Approximation to Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854-866.
- (85) Hopfinger, A., J.; Wang, S.; Tokarski, J., S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509-10524.
- (86) Senese, C., L.; Hopfinger, A., J. Receptor-Independent 4D-QSAR Analysis of a Set of Norstatine Derived Inhibitors of HIV-1 Protease. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1297-1307.
- (87) Senese, C., L.; Hopfinger, A., J. A Simple Clustering Technique to Improve QSAR Model Selection and Predictivity: Application to a Receptor Independent 4D-QSAR Analysis of Cyclic Urea Derived Inhibitors of HIV-1 Protease. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2180-2193.
- (88) De Oliveira, D., B.; Gaudio, A., C. BuildQSAR: A New Computer Program for QSAR Studies. *Quant. Struct.-Act. Relat.* **2000**, *19*, 599-601.
- (89) Wold, S.; Erikson, L. *Statistical Validation of QSAR Results. Validation Tools*; In *Chemometric Methods in Molecular Design*, van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995, p. 309-318.