9th International Electronic Conference on Synthetic Organic Chemistry. ECSOC-9
1-30 November 2005. http://www.usc.es/congresos/ecsoc/9/ECSOC9.HTM & http://www.mdpi.net/ecsoc/

[G015]

Bond, Bond-Type, and Total Linear Indices of the non-Stochastic and Stochastic Edge Adjacency Matrix. 1. Theory and QSPR Studies

Yovani Marrero Ponce[1,2]* & Francisco Torrens[2]

[1]Department 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.

[2]Institut 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 bond-level molecular descriptors based on linear maps similar to those defined in algebra theory are proposed. The k^{th} edge-adjacency matrix ($\mathbf{E}^k$) denotes the matrix of bond linear indices (non-stochastic) with respect to the canonical basis set. The k^{th} stochastic edge-adjacency matrix, $\mathbf{ES}^k$, is here proposed as a new molecular representation easily calculated from $\mathbf{E}^k$. Then, the k^{th} stochastic bond linear indices are calculated using $\mathbf{ES}^k$ as operators of linear transformations. In both cases, the bond-type formalism was developed. The k^{th} non-stochastic and stochastic bond-type linear indices values are the sum of the k^{th} non-stochastic and stochastic bond linear indices values for bonds of the same bond type, respectively. In the same way, the k^{th} non-stochastic and stochastic total (whole-molecule) linear indices are calculated by summing up the k^{th} non-stochastic and stochastic bond linear indices, correspondingly, of all bonds in the molecule. The new bond-based molecular descriptors were tested for suitability for the quantitative structure-property relationship (QSPR) by analyzing regressions of novel indices for selected physicochemical properties of octane isomers. All the found regression models are very significant from the statistical point of view and showed very good stability to data variation in leave-one-out cross-validation experiments. General performance of the new descriptors in this QSPR studies has been evaluated with respect to the well-known sets of 2D/3D molecular descriptors. From the analysis, we can conclude that the non-stochastic and stochastic bond-based (total and bond-type) linear indices have an overall good modeling capability proving their usefulness in QSPR studies. The approach described in this work appears to be a very promising structural invariant, useful not alone for QSPR/QSAR studies, but also for similarity/diversity analysis and drug discovery protocols.

Keywords: TOMOCOMD-CARDD Software, Non-Stochastic and Stochastic Bond-Based Linear Indices, Edge-Adjacency Matrix, QSPR Studies, Physicochemical Properties, Octane Isomers.

1. INTRODUCTION

In the context of new technologies for drug discovery, such as combinatorial chemistry and high-throughput screening, molecular descriptors play an important role for the analysis of molecular diversity and lead to optimization through well-established **Q**uantitative **S**tructure-**P**roperty/**A**ctivity **R**elationships (**QSPR/QSAR**) studies.[1-2] The so-called topological indices (TIs) are among the most useful molecular descriptors known nowadays.[3-5] These theoretical indices are numbers that describe the structural information of molecules through graph theoretical invariants and can be considered as structure-explicit descriptors.[6] At present, there are a great number of TIs that can be used in QSPR/QSAR studies. However, a simple inspection of the large number of TIs defined in the literature shows that many of them are computed with identical mathematical equations, by using different molecular matrices. There are two main sources of TIs, the vertex (atom)-based adjacency (**A**) and distance (**D**) matrices,[1-7] furthermore the number and diversity of the graph invariants is so wide that this makes it difficult to find general relations for the so-derived molecular fingerprints.

The edge (bond)-adjacency relationships have also been used in the generation of new TIs. Their matrix form has been considered and explicitly defined in the chemical graph theory literature, but has received very little attention in both chemical and mathematical literature. Nevertheless, in the last decade Estrada rediscovered this matrix as an important source of graph theoretical invariants useful in the generation of new molecular descriptors.[1] For instance, first the ϵ index was defined by this author[8] using the Randić-type graph-theoretical invariant. That is to say, this new index is analogous to the Randić branching index but calculated by edge degrees instead of vertex degrees.

In a second work, our research group[9] extends the edge adjacency matrix **E** in molecular graph in a 3D-**E** matrix in order to generate the so-colled topographic edge-connectivity index $\epsilon(\rho)$, also using the Randić-type graph-theoretical invariant. Later, Estrada used the same edge adjacency relationships in the generation of the a new family of TIs, spectral moments of the E-matrix.[10] The analogous concept of spectral moments of vertex-adjacency matrix had also been discussed previously by different authors.[11] Afterward, Estrada et al.[12] introduced a extended set of edge connectivity indices, ${}^{m}\epsilon_{t}(G)$, using the same way in which the branching index of Randić was extended to the series of

molecular connectivity indices. Finally, a novel graph theoretical polynomial, $P_\epsilon(G, x)$, counting the edge connectivity was introduced by the same researcher.[13] The first derivative of this polynomial evaluated for $x = 0$ is equal to the edge-connectivity index of the molecular graph. A series of edge-connectivity indices modified to include long-range bond contributions, $\epsilon^c(x)$, was obtained by this author using values of x different from zero. Such edge-adjacency relationships will be applied in the present report in order to generate a series of bond-based molecular descriptors to be used in drug design and chemoinformatic studies.

On the other hand, TIs can be classified as "global" and "local" according to the way in which they characterize the molecular structure, although most of them can be considered as global molecular fingerprints.[14] One exception in this sense is the electrotopological state (E-state) index.[15] Other "global" descriptors such as spectral moments of the edge-adjacency matrix had been redefined in local form.[14] The great success of the E-state and edge-based spectral moments in QSPR/QSAR recently stimulated us to propose and validate some novel total and local descriptors based on a topological (edge-adjacency relationships) characterization of the molecular structure. In this sense, in a manner similar to that for the atom- and atom-type level E-State, an E-State index for bonds and bond-type has been proposed. The bond-based E-State indices provided an improvement of 25% with regard to the atom-based E-State indices in the description of the boiling point of 372 alkanes, alcohols, and chloroalkanes.[15]

Recently, one of the present authors, Y. M-P, has introduced a new set of atom-level molecular descriptors of relevance to QSAR/QSPR studies and 'rational' drug design, atom linear indices $\boldsymbol{f}_k(x_i)$.[16] These local (atom and atom-type) indices are based on the calculation of linear maps in $\Re^n$ in canonical basis. The description of the significance-interpretation and the comparison to other molecular descriptors was also performed.[16] This approach describes changes along the time in the electronic distribution throughout the molecular backbone.[16-19] Specifically, the features of the k^{th} total and local linear indices were illustrated by examples of various types of molecular structures, including chain length and branching as well as content of heteroatoms, and multiple bonds.[16] Additionally, the linear independence of the atom-type linear fingerprints to 229 other 0D-3D molecular descriptors was demonstrated. In this sense, it was concluded that local

(atom-based) linear fingerprints are independent indices, which contain important structural information to be used in QSPR/QSAR and drug design studies.[16]

This -*in silico*- method has been successfully applied to the prediction of several physical, physicochemical and chemical properties of organic compounds.[17] These atom-level molecular descriptors, and their stochastic forms,[18,19] have also been useful for the selection of novel subsystems of compounds having a desired property/activity. In this sense, it was successfully applied to the virtual (computational) screening of novel anthelmintic compounds, which were then synthesized and *in vivo* evaluated on *Fasciola hepatica.*[20] Studies for the fast-track discovery of novel antibacterial and antimalarial compounds were also conducted with this theoretical approach.[18-19] In addition, the molecular linear indices have been extended to consider three-dimensional features of small/medium-sized molecules based on the trigonometric-3D-chirality-correction factor approach.[21] 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 the nucleic acid's linear indices.[22] An alternative formulation of our approach for structural characterization of proteins was also carried out recently.[23] This extended method was used to encompass protein stability studies –specifically how alanine substitution mutation on Arc repressor wild-type protein affects protein stability– by means of a combination of protein linear or quadratic indices (macromolecular fingerprints) and statistical (linear and non-linear model) methods.[23]

We propose in this paper a new local (bond and bond-type) and total molecular descriptors based on the adjacency of edges. We also propose in this paper a new matrix representation of the molecule on the "stochastic" adjacency of edges and linear indices derived from there. In addition, the correlation ability of the new descriptors is tested in a QSPR study of some physicochemical properties of octanes.

2. THEORETICAL FRAMEWORK

The basis of the extension of linear indices that will be given here is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature,[24,25] and rediscovered by Estrada as an important source of new molecular descriptors.[8-10, 12-14] In this section, we first will define the nomenclature to be used in this work, then the atom-

based molecular vector (**X**) will be redefined for bond characterization using the same approach as previously reported, and finally some new definition of bond-based non-stochastic and stochastic linear indices will be given.

2.1. Background in Edge-Adjacency Matrix

Let $G = (V, E)$ be a simple graph, with $V = \{v_1, v_2, ..., v_n\}$ and $E = \{e_1, e_2, ...e_m\}$ being the vertex- and edge-sets of G, respectively. Then G represents a molecular graph having n vertices and m edge (bonds). The edge-adjacency matrix **E** of G (likewise called bond-adjacency matrix, **B**) is a square and symmetric matrix whose elements e_{ij} are 1 if and only if edge i is adjacent to edge j.[1,10,14] Two edges are adjacent if they are incidental to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the i^{th} row (or column) of **E** is named the edge-degree of bond i, $\delta(e_i)$.[1,8,12,13]

2.2. New Edge-Relations: Stochastic Edge-Adjacency Matrix

By using the edge (bond)-adjacency relationships we can find other new relation for a molecular graph that will be introduced here. The k^{th} stochastic edge-adjacency matrix, $\mathbf{ES}^k$ can be obtained directly from $\mathbf{E}^k$. Here, $\mathbf{ES}^k = [{}^k es_{ij}]$ is a square table of order m (m = number of bonds) and the elements ${}^k es_{ij}$ are defined as follows:

$$ {}^k es_{ij} = \frac{{}^k e_{ij}}{{}^k SUM(E^k)_i} = \frac{{}^k e_{ij}}{{}^k \delta(e)_i} \qquad (1) $$

where, ${}^k e_{ij}$ are the elements of the k^{th} power of **E** and the SUM of the i^{th} row of $\mathbf{E}^k$ are named the k-order edge degree of bond i, ${}^k \delta(e)_i$. Note that the matrix $\mathbf{ES}^k$ in Eq. **1** has the property that *the sum of the elements in each row* is 1. An mxm matrix with nonnegative entries having this property is called a "**stochastic matrix**".[26]

2.3. Structural Representation Although of the Bond-Based Molecular Vector

The atom-based molecular vector (**X**) used to represent small-to-medium size organic chemicals have been explained in some detail elsewhere.[16-20,27-35] In a manner parallel to the development of **X**, we present the expansion of the bond-based molecular vector (**W**). The components (w) of **W** are numeric values, which represent a certain standard bond property (bond-label). That is to say, these weights correspond to different bond properties for organic molecules. Thus, a molecule having 5, 10, 15,..., m bonds can be represented by means of vectors, with 5, 10, 15,..., m components, belonging to the

spaces $\Re^5$, $\Re^{10}$, $\Re^{15}$,..., $\Re^m$, respectively; where m is the dimension of the real sets ($\Re^m$). This approach allows us encoding organic molecules such as 2-hydroxybut-2-enenitrile through the molecular vector $\mathbf{W} = [w_{\text{Csp3-Csp2}}, w_{\text{Csp2=Csp2}}, w_{\text{Csp2-Osp3}}, w_{\text{H-Osp3}}, w_{\text{Csp2-Csp}}, w_{\text{Csp}\equiv\text{Nsp}}]$. This vector belongs to the product space $\Re^6$.

These properties characterize each kind of bond (and bond-types) within the molecule. Diverse kinds of bond weights (w) can be used in order to codify information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distance,[36-39] standard bond dipole[36-39] or even mathematical expressions involving atomic weights such as atomic Log P,[40] surface contributions of polar atoms,[41] atomic molar refractivity,[42] atomic hybrid polarizabilities,[43] and Gasteiger-Marsilli atomic charge,[44] atomic electronegativity in Pauling scale[45] and so on. Here, we characterized each bond with the following parameter:

$$w_i = x_i/\delta_i + x_j/\delta_j \qquad (\mathbf{2})$$

which characterizes each bond. In this expression x_i can be any standard weight of the atom i bonded with atom j. δ_i is the vertex (atom) degree of atom i. The use of each scale (bond property) defines alternative molecular vectors, **W**.

2.4. Calculation of Linear Indices for Bonds, Bond-Types and the Whole Molecule

If a molecule consists of m bonds (*vector of* $\Re^m$), then the k^{th} bond linear indices for bond i in a molecule, are calculated as linear maps on $\Re^m$ (endomorphism on $\Re^m$) in canonical basis set. Specifically, the k^{th} non-stochastic and stochastic bond linear indices, $\boldsymbol{f_k}(w_i)$ and ${}^{\mathbf{s}}\boldsymbol{f_k}(w_i)$, are computed from these k^{th} non-stochastic and stochastic edge-adjacency matrices, $\mathbf{E}^{\boldsymbol{k}}$ and $\mathbf{ES}^{\boldsymbol{k}}$, as shown in Eqs. **3** and **4**, respectively:

$$f_k(w_i) = \sum_{j=1}^{m} {}^k e_{ij} w_j = [\mathrm{W'}]^k = \mathbf{E}^{\boldsymbol{k}}[\mathrm{W}] \qquad (\mathbf{3})$$

$${}^s f_k(w_i) = \sum_{j=1}^{m} {}^k es_{ij} w_j = [\mathrm{WS'}]^k = \mathbf{ES}^{\boldsymbol{k}}[\mathrm{W}] \qquad (\mathbf{4})$$

where m is the number of bonds of the molecule and w_j are the coordinates of the bond-based molecular vector (**W**) in the so-called canonical ('natural') basis. In this basis system, the coordinates of any vector **W** coincide with the components of this vector.[26,46-47] For that reason, those coordinates can be considered as weights (bond-labels) of the

edge of the molecular graph. The coefficients ${}^{k}e_{ij}$ and ${}^{k}es_{ij}$ are the elements of the k^{th} power of the matrix **E**(G) and **ES**(G), correspondingly, of the molecular graph. The defining equation (**3**) and (**4**) for $\boldsymbol{f_k}(w_i)$ and ${}^{s}\boldsymbol{f_k}(w_i)$, respectively, may be also written as the single matrix equation, where [W] is a column vector (an mx1 matrix) of the coordinates of **W** in the canonical basis of $\Re^m$. Here, $\mathbf{E}^{\boldsymbol{k}}$ and $\mathbf{ES}^{\boldsymbol{k}}$ denote the matrices of linear maps with respect to the natural basis set.

Note that both bond linear indices are defined as a linear transformation $\boldsymbol{f_k}(w_i)$ on molecular vector space $\Re^m$. This map is a correspondence that assigns a vector $\boldsymbol{f}(w)$ to every vector **W** in $\Re^m$ in such a way that:

$$f(\lambda_1 W_1 + \lambda_2 W_2) = \lambda_1 f(W_1) + \lambda_2 f(W_2) \qquad (\mathbf{5})$$

for any scalar λ_1,λ_2 and any vector W_1, W_2 in $\Re^m$.

Total (whole-molecule) bond-based non-stochastic and stochastic linear indices, $\boldsymbol{f_k}(w)$ and ${}^{s}\boldsymbol{f_k}(w)$, are calculated from local (bond) linear indices as shown in Eqs. **6** and **7**, correspondingly:

$$f_k(w) = \sum_{i=1}^{m} f_k(w_i) = [\mathrm{u}]^{\mathrm{t}}[\mathrm{W'}]^k = [\mathrm{u}]^{\mathrm{t}}\mathbf{E}^{\boldsymbol{k}}[\mathrm{W}] \qquad (\mathbf{6})$$

$${}^{s}f_k(w) = \sum_{i=1}^{m} {}^{s}f_k(w_i) = [\mathrm{u}]^{\mathrm{t}}[\mathrm{WS'}]^k = [\mathrm{u}]^{\mathrm{t}}\mathbf{ES}^{\boldsymbol{k}}[\mathrm{W}] \qquad (\mathbf{7})$$

where m is the number of bonds, and $\boldsymbol{f_k}(w_i)$ and ${}^{s}\boldsymbol{f_k}(w_i)$ are the non-stochastic and stochastic bond linear indices obtained by Eqs. **3** and **4,** respectively. Then, both total linear form, $\boldsymbol{f_k}(w)$ and ${}^{s}\boldsymbol{f_k}(w)$, can also be written in matrix form for each molecular vector $\mathbf{W} \in \Re^n$, where $[\mathrm{u}]^{\mathrm{t}}$ is an n-dimensional unitary row vector. As it can be seen, the k^{th} total linear indices (both non-stochastic and stochastic) are calculated by summing the local (bond) linear indices of all bonds in the molecule.

Finally, in addition to total and bond linear indices computed for each bond in the molecule, a local-fragment (bond-type) formalism can be developed. The k^{th} bond-type linear index of the edge-adjacency matrix is calculated by summing up the k^{th} bond linear indices of all bonds of the same bond type in the molecule. That is to say, this extension of the bond linear index is similar to the group additive schemes, in which an index

appears for each bond type in the molecule together with its contribution based on the bond linear index. Consequently, if a molecule is partitioned into Z molecular fragments, the total non-stochastic [or stochastic] linear indices can be partitioned into Z local non-stochastic [or stochastic] linear indices $\boldsymbol{f_{kL}}(w)$ [or ${}^{s}\boldsymbol{f_{kL}}(w)$], L = 1, …, Z. That is to say, the total (both non-stochastic and stochastic) linear indices of order k can be expressed as the sum of the local linear indices of the Z fragments of the same order:

$$f_k(w) = \sum_{L=1}^{Z} f_{kL}(w) \qquad (\mathbf{8})$$

$${}^{s}f_k(w) = \sum_{L=1}^{Z} {}^{s}f_{kL}(w) \qquad (\mathbf{9})$$

In the bond-type linear indices formalism, each bond in the molecule is classified into a bond-type (fragment). In this sense, bonds may be classified into bond types in terms of the characteristics of the two atoms that define the bond. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the k^{th} fragment (bond-type) linear indices provide much useful information. Thus, the development of the bond-type linear indices description provides the basis for application to a wider range of biological problems in which the local formalism is applicable without the need for superposition of a closely related set of structures.

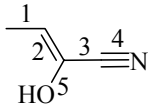
It is useful to perform a calculation on a molecule to illustrate the steps in the procedure. For this, in the next section I depict a pictorial representation of the calculus of the non-stochastic and stochastic linear indices of the bond matrix (both total and local) using a simple chemical example. In that section, I will also stand out that our approach is rather similar to the **LCBO-MO** (**L**inear **C**ombination of **B**ond **O**rbitals-**M**olecular **O**rbitals) method (e.g., for $k = 1$).[48] **LCBO-MO** is another way of forming molecular orbitals by taking linear combinations of functions associated with the different bonds in the molecule. In this sense, MOs are made up as LCBO of bonds composing the system, i.e. are written in the form,

$$\varphi_i = \sum_{j=1}^{n} c_{ij} Y_j \qquad (\mathbf{10})$$

where i is the number of the MO, φ [in our case, $\boldsymbol{f_1}(w_i)$]; j are the numbers of bond Y–orbitals (in our case, w_j); c_{ij} (in our case, ${}^1e_{ij}$ or ${}^1es_{ij}$ for non-stochastic and stochastic indices, respectively) are the numerical coefficients defining the contributions of individuals BOs to the given MO. Although the **LCAO** (**L**inear **C**ombination of **A**tom **O**rbitals) approximation has been particularly useful for the study of conjugated hydrocarbons, the **LCBO** method has been particularly applied to the calculation of properties of saturated hydrocarbons. As a saturated molecule can be considered as made up of localized bonds, it is reasonable to associate an orbital to each of the corresponding regions.[48]

2.5. Sample Calculation

The linear indices of the bond matrix are calculated in the following way. Considering the molecule of 2-hydroxybut-2-enenitrile as a simple example, we have the following labeled molecular graph and bond-based adjacency matrices (**E** and **ES**). The second (k = 2) and third (k = 3) power of these matrices and bond-based molecular vector, **W**, are also given:



$$E^0 = ES^0 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \quad E^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad E^2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 3 & 1 & 1 & 1 \\ 1 & 1 & 3 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix} \quad E^3 = \begin{bmatrix} 0 & 3 & 1 & 1 & 1 \\ 3 & 2 & 5 & 1 & 4 \\ 1 & 5 & 2 & 3 & 4 \\ 1 & 1 & 3 & 0 & 1 \\ 1 & 4 & 4 & 1 & 2 \end{bmatrix}$$

$$ES^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \quad ES^2 = \begin{bmatrix} 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.5 & 0.16 & 0.16 & 0.16 \\ 0.16 & 0.16 & 0.5 & 0 & 0.16 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0.16 & 0.16 & 0.16 & 0.16 & 0.33 \end{bmatrix} \quad ES^3 = \begin{bmatrix} 0 & 0.5 & 0.16 & 0.16 & 0.16 \\ 0.2 & 0.13 & 0.33 & 0.06 & 0.26 \\ 0.06 & 0.33 & 0.13 & 0.2 & 0.26 \\ 0.16 & 0.16 & 0.5 & 0 & 0.16 \\ 0.083 & 0.33 & 0.33 & 0.083 & 0.16 \end{bmatrix}$$

The molecule contains five localized bonds (Corresponding to five edges in the H-suppressed molecular graph). To these we will associate the five "bond orbitals" w_1, w_2, w_3, w_4, and w_5. Thus, $\mathbf{W} = [w_1, w_2, w_3, w_4, w_5] = [w_{(C\text{-}C)}, w_{(C=C)}, w_{(C\text{-}C)}, w_{(C\equiv N)}, w_{(C\text{-}O)}]$ and

each "bond orbital" can be computed by Eq. **2** using, for instance, the atomic electronegativity in Pauling scale (x)[45] as atomic weight (atom-label):

$w_1 = x_C/1 + x_C/3 = 2.55/1 + 2.55/3 = 3.4$

$w_2 = x_C/3 + x_C/4 = 2.55/3 + 2.55/4 = 1.4875$

$w_3 = x_C/4 + x_C/4 = 2.55/4 + 2.55/4 = 1.275$

$w_4 = x_C/4 + x_N/3 = 2.55/4 + 3.04/3 = 1.650833$

$w_5 = x_C/4 + x_O/1 = 2.55/4 + 3.44/1 = 4.0775$

and therefore, $\mathbf{W}$ = [3.4, 1.4875, 1.275, 1.650833, 4.0775]

Each non-stochastic and stochastic "molecular orbital" will have the form:

$$f_k(w_i) = {}^k e_{i1}w_1 + {}^k e_{i2}w_2 + {}^k e_{i3}w_3 + {}^k e_{i4}w_4 + {}^k e_{i5}w_5 \quad (\mathbf{11})$$

$$^s f_k(w_i) = {}^k es_{i1}w_1 + {}^k es_{i2}w_2 + {}^k es_{i3}w_3 + {}^k es_{i4}w_4 + {}^k es_{i5}w_5 \quad (\mathbf{12})$$

The ${}^k e_{ii}$'s and ${}^k es_{ii}$'s can be considered to measure a the attraction of an electron for a bond in the k step. The ${}^k e_{ij}$'s and ${}^k es_{ij}$'s are the terms of interaction between two bonds in the k step. The ${}^k e_{ij} = {}^k e_{ji}$ are equal by symmetry (non-oriented molecular graph). However, ${}^k es_{ij}$'s $\neq {}^k es_{ji}$'s. This is a logical result because the k^{th} es_{ij} elements are the transition probabilities with the 'electrons' moving from bond i to j at the discrete time periods t_k and it should be different in both senses. This result is in total agreement if the electronegativity of the two atom types in the bonds are taken into account.

In this way, $\mathbf{E}^k$ and $\mathbf{ES}^k$ can be seen as graph–theoretic electronic–structure models.[49] In fact, quantum chemistry starts from the fact that a molecule is made up of electrons and nuclei. The distinction here between bonded and non-bonded atoms is difficult to justify. Any two nuclei of a molecule interact directly and indirectly through the electrons present in the molecule. Only the intensity of this interaction varies on going from one pair of nuclei to another. In this sense, the electron in an arbitrary bond i can move (step-by-step) to other bonds at different discrete time periods t_k (k = 0, 1, 2, 3,…) through the chemical-bonding network. That is to say, the $\mathbf{E}^1$ and $\mathbf{ES}^1$ matrices consider the valence-bond electrons in one step and their power (k = 0, 1, 2, 3…) can be considering as an interacting–electron chemical–network model in k step. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.[49]

On the other hand, the k^{th} (k = 0-3) non-stochastic bond linear indices can be calculated for this molecule as follows:

$$f_0(w_i) = \sum_{j=1}^{5} {}^0e_{ij}w_j = \mathbf{E^0}[\mathrm{W}] = {}^0e_{i1}w_1 + {}^0e_{i2}w_2 + {}^0e_{i3}w_3 + {}^0e_{i4}w_4 + {}^0e_{i5}w_5$$

$$f_1(w_i) = \sum_{j=1}^{5} {}^1e_{ij}w_j = \mathbf{E^1}[\mathrm{W}] = {}^1e_{i1}w_1 + {}^1e_{i2}w_2 + {}^1e_{i3}w_3 + {}^1e_{i4}w_4 + {}^1e_{i5}w_5$$

$$f_2(w_i) = \sum_{j=1}^{5} {}^2e_{ij}w_j = \mathbf{E^2}[\mathrm{W}] = {}^2e_{i1}w_1 + {}^2e_{i2}w_2 + {}^2e_{i3}w_3 + {}^2e_{i4}w_4 + {}^2e_{i5}w_5$$

$$f_3(w_i) = \sum_{j=1}^{5} {}^3e_{ij}w_j = \mathbf{E^3}[\mathrm{W}] = {}^3e_{i1}w_1 + {}^3e_{i2}w_2 + {}^3e_{i3}w_3 + {}^3e_{i4}w_4 + {}^3e_{i5}w_5$$

and stochastic linear indices for each bond i can be computed for this molecule in a similar form:

$${}^sf_0(w_i) = \sum_{j=1}^{5} {}^0es_{ij}w_j = \mathbf{ES^0}[\mathrm{W}] = {}^0es_{i1}w_1 + {}^0es_{i2}w_2 + {}^0es_{i3}w_3 + {}^0es_{i4}w_4 + {}^0es_{i5}w_5$$

$${}^sf_1(w_i) = \sum_{j=1}^{5} {}^1es_{ij}w_j = \mathbf{ES^1}[\mathrm{W}] = {}^1es_{i1}w_1 + {}^1es_{i2}w_2 + {}^1es_{i3}w_3 + {}^1es_{i4}w_4 + {}^1es_{i5}w_5$$

$${}^sf_2(w_i) = \sum_{j=1}^{5} {}^2es_{ij}w_j = \mathbf{ES^2}[\mathrm{W}] = {}^2es_{i1}w_1 + {}^2es_{i2}w_2 + {}^2es_{i3}w_3 + {}^2es_{i4}w_4 + {}^2es_{i5}w_5$$

$${}^sf_3(w_i) = \sum_{j=1}^{5} {}^3es_{ij}w_j = \mathbf{ES^3}[\mathrm{W}] = {}^3es_{i1}w_1 + {}^3es_{i2}w_2 + {}^3es_{i3}w_3 + {}^3es_{i4}w_4 + {}^3es_{i5}w_5$$

The total non-stochastic linear indices can be expressed as the sum of the local (bond) linear indices for this molecule as follows:

$$f_0(w) = \sum_{i=1}^{5} f_0(w_i) = [\mathrm{u}]^t[\mathrm{W'}]^0 = [\mathrm{u}]^t\mathbf{E^0}[\mathrm{W}] = f_0(w_1) + f_0(w_2) + f_0(w_3) + f_0(w_4) + f_0(w_5)$$

$$= 3.4 + 1.4875 + 1.275 + 1.650833 + 4.0775 = 11.89083$$

$$f_1(w) = \sum_{i=1}^{5} f_1(w_i) = [\mathrm{u}]^t[\mathrm{W'}]^1 = [\mathrm{u}]^t\mathbf{E^1}[\mathrm{W}] = f_1(w_1) + f_1(w_2) + f_1(w_3) + f_1(w_4) + f_1(w_5)$$

$$= 1.4875 + 8.7525 + 7.215833 + 1.275 + 2.7625 = 21.49333$$

$$f_2(w) = \sum_{i=1}^{5} f_2(w_i) = [\mathrm{u}]^t[\mathrm{W'}]^2 = [\mathrm{u}]^t\mathbf{E^2}[\mathrm{W}] = f_2(w_1) + f_2(w_2) + f_2(w_3) + f_2(w_4) + f_2(w_5)$$

$$= 8.7525 + 11.46583 + 12.79 + 7.215833 + 15.96833 = 56.1925$$

$$f_3(w) = \sum_{i=1}^{5} f_3(w_i) = [u]^t [W']^3 = [u]^t \mathbf{E}^3 [W] = f_3(w_1) + f_3(w_2) + f_3(w_3) + f_3(w_4) + f_3(w_5)$$

$$= 11.46583 + 37.51083 + 34.65 + 12.79 + 24.25583 = 120.6725$$

The terms in the summations for calculating the total linear indices are the so-called bond linear indices. We have written these terms in the consecutive order of the bond labels in the graph. For instance, the non-stochastic bond linear indices of order 0, 1, 2 and 3 for the bond labeled as 1 are 3.4, 1.4875, 8.7525, and 11.46583, respectively.

The k^{th} total stochastic linear indices values are also the sum of the k^{th} local (bond) stochastic linear indices values for all bonds in the molecule:

$${}^s f_0(w) = \sum_{i=1}^{5} {}^s f_0(w_i) = [u]^t [WS']^0 = [u]^t \mathbf{ES}^0 [W] = {}^s f_0(w_1) + {}^s f_0(w_2) + {}^s f_0(w_3)$$

$$+ {}^s f_0(w_4) + {}^s f_0(w_5) = 3.4 + 1.4875 + 1.275 + 1.650833 + 4.0775 = 11.89083$$

$${}^s f_1(w) = \sum_{i=1}^{5} {}^s f_1(w_i) = [u]^t [WS']^1 = [u]^t \mathbf{ES}^1 [W] = {}^s f_1(w_1) + {}^s f_1(w_2) + {}^s f_1(w_3)$$

$$+ {}^s f_1(w_4) + {}^s f_1(w_5) = 1.4875 + 2.9175 + 2.405278 + 1.275 + 1.38125 = 9.466528$$

$${}^s f_2(w) = \sum_{i=1}^{5} {}^s f_2(w_i) = [u]^t [WS']^2 = [u]^t \mathbf{ES}^2 [W] = {}^s f_2(w_1) + {}^s f_2(w_2) + {}^s f_2(w_3)$$

$$+ {}^s f_2(w_4) + {}^s f_2(w_5) = 2.9175 + 1.910972 + 2.131667 + 2.405278 + 2.661389 = 12.02681$$

$${}^s f_3(w) = \sum_{i=1}^{5} {}^s f_3(w_i) = [u]^t [WS']^3 = [u]^t \mathbf{ES}^3 [W] = {}^s f_3(w_1) + {}^s f_3(w_2) + {}^s f_3(w_3)$$

$$+ {}^s f_3(w_4) + {}^s f_3(w_5) = 1.910972 + 2.500722 + 2.31 + 2.131667 + 2.021319 = 10.87468$$

3. QSPR Studies

The decisive criterion of quality for any molecular descriptor is its ability to describe structure-related properties of molecules. With this objective we developed the QSPR models to describe seven physicochemical properties of octane isomers. The use of octanes as a very suitable data set for testing topological indices has been advocated by Randić and Trinajstić.[50,51] In fact, this dataset has been used by several researchers to evaluate the modeling power of their new molecular descriptors.[13,52-58] This selection is recommended due to the most of the fact that physicochemical properties commonly studied in QSPR analyses with topological indices 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.[59] On the other hand, all topological indices are designed to have (gradual) increments with the increments 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.[13,52] 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.[13,52] 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.[60] However, when data sets of isomeric compounds are considered, typically correlations that have high correlation coefficients when molecules of different size were considered will no longer show such good linear correlation. In conclusion, if a new proposed molecular descriptor is not able to model the variation of at least one property of octanes, then it probably does not contain any useful molecular information. Moreover, octanes constituted a good set of chemicals for comparative study, since many experimental data among their physicochemical properties are available. In this sense, we analyzed the quality of the QSPR models obtained to describe the boiling point (BP), motor octane number (MON), heat of vaporization (HV), molar volume (MV), entropy (S), and heat of formation ($\Delta_f H$) of the octane isomers. In addition, regressions of octane properties based on the non-stochastic and stochastic linear indices will be compared to some regressions based on 2D (topologic) and 3D (geometric) descriptors taken from the literature.[13,52-58] Precisely, to evaluate the quality of the models based on our new bond-level chemical descriptors we have taken as the reference: 1) the models published by Randić[54-56] based on diverse topological indices such as the Wiener matrix invariants, 2) the equation published by Diudea[58] based on the SP indices, and 3) the best models obtained with a set constituted by the topological (69), WHIM (99), and GETAWAY descriptors (197).[53]

The total and local (bond-type) bond-based linear indices used to search for the best regression of the selected physicochemical properties of octanes were calculate by the

TOMOCOMD-CARDD (acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign-***C***omputer ***A***ided "***R***ational" ***D***rug ***D***esign) program.[61] This software is an interactive program for molecular design and bioinformatic research. The software was developed based on a user-friendly philosophy. That is to say, this computer graphics software shows a great efficiency of interaction with the user, without *prior* knowledge of programming skills (e.g. practicing pharmaceutic and organic chemist, teacher, university student, and so on). CARDD subprogram allows drawing the structures (drawing mode) and calculating 2D (topologic), 3D-chiral (2.5D) and 3D (geometric and topographic) non-stocahstic and stochastic molecular descriptors (calculation mode). The bond–based *TOMOCOMD-CARDD* descriptors computed in this study were the following:

1) k^{th} (k = 15) total non-stochastic bond-based linear indices not considering and considering H-atoms in the molecular graph (G) [$\boldsymbol{f}_k(w)$ and $\boldsymbol{f}_k^{H}(w)$, respectively].
2) k^{th} (k = 15) total stochastic bond-based linear indices not considering and considering H-atoms in the molecular graph (G) [$^{\$}\boldsymbol{f}_k(w)$ and $^{\$}\boldsymbol{f}_k^{H}(w)$, respectively].
3) k^{th} (k = 15) bond-type (C-H in methyl group) non-stochastic and stochastic linear indices considering H-atoms in the molecular graph (G) [$\boldsymbol{f}_{kL}^{H}(w_{C\text{-}H})$ and $^{\$}\boldsymbol{f}_{kL}^{H}(w_{C\text{-}H})$, correspondingly]. These local descriptors are calculated taken into account only one of the three bond types for carbon-hydrogen bonds ($C_{primary}$-H) that there are for octanes data.

These k^{th} total and local bond-based linear indices were used as molecular descriptors for derived QSARs. One of the difficulties with the large number of descriptors is deciding which ones will provide the best regressions, considering both goodness of fit and the chemical meaning of the regression. In addition, 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.[62-67] 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.[62-70]

The software BuildQSAR[71] was employed to perform variable selection and QSAR modeling. The mutation probability was specified as 35%. The length of the equations was set three-four 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).[71] The quality of models was also determined by examining the Leave-One-Out (LOO) cross-validation (CV) (q^2, s_{cv}).[72] In recent years, the LOO press statistics (e.g., q^2) have been used as a means of indicating predictive ability. Many authors consider high q^2 values (for instance, $q^2 > 0.5$) as an indicator or even as the ultimate proof of the high-predictive power of an QSAR model.

The best linear models found using non-stochastic and stochastic total and bond-type linear indices are presented in Table 1. For each selected property of octane isomers, the statistical information for the best regressions with 1, 2, and 3 molecular descriptors published so far are also depicted in Table 1. Together with the LOO cross-validated explained variance (q^2_{LOO}), the determination coefficient (R^2), the standard estimate of the error (s), and Fischer ratio (F) are listed. The molecular descriptor symbols are reported in eighth column, and the last column in the table contains the references of the models taken from the literature.

Table 1. Statistical Information for Best Multiple Regression Models of Selected Physicochemical Properties of Octane Isomers.

Property	Method	size	Q^2_{LOO}	R^2	s	F	Model Descriptors	Ref.
Boiling Point (BP)	NonStochastic Bond-based Linear indices	3	92.81	95.13	1.487	91.143	**BP** = 137.99 -1.47$\boldsymbol{f}_{2L}^{H}(w_{C\text{-}H})$ +0.07$\boldsymbol{f}_{4L}^{H}(w_{C\text{-}H})$ -1.51$\boldsymbol{f}_{1L}^{H}(w_{C\text{-}H})$ (**13**)	
	Stochastic Bond-based Linear indices	3	92.86	96.29	1.298	121.07	**BP** = 43.84 -17.19 $^{s}\boldsymbol{f}_{2}^{H}(w)$ +22.78 $^{s}\boldsymbol{f}_{3}^{H}(w)$ -4.18$^{s}\boldsymbol{f}_{13L}^{H}(w_{C\text{-}H})$ (**14**)	
	getaway + whim + top.	3	98.12	98.78	0.744		$^{2}\chi$ $^{2}\overline{\chi}$ $HATS_6$(p)	53
	getaway	3	97.10	98.32	0.897		$HATS_2$(v) R_4(u) R_6(v)	53
	getaway + whim + top.	2	96.62	97.58	1.013		$^{2}\chi$ $HATS_6$(p)	53
	topological	3		95.84	1.394		S^3W S^4W SJ	58
	topological	2		94.78	1.508		S^3W S^4W	58
	getaway	2	84.86	89.62	2.098		$HATS_2$(m) R^+_4(u)	53
	topological	2		81.36	2.810		WW x_1	56
	topological	1		78.85	2.90		Z	55
	getaway + whim + top.	1	66.47	74.64	3.175		$HATS_2$(m)	53
	topological	1		67.77	3.630		$^{2}\chi$ W	58
Motor Octane Number (MON)	NonStochastic Bond-based Linear indices	3	98.90	99.37	2.871	687.66	**MON** = -349.3-2.47x10^{-4} $\boldsymbol{f}_{10}^{H}(w)$ -2.33x10^{-6} $\boldsymbol{f}_{14L}^{H}(w)$ +1.33x10^{-5} $\boldsymbol{f}_{13}^{H}(w)$ (**15**)	

	Stochastic Bond-based Linear indices	3	97.73	98.51	4.424	287.16	**MON** = -243.98 +168.91 $^{s}\boldsymbol{f}_{10}{}^{H}(w)$ +29.65 $^{s}\boldsymbol{f}_{1L}{}^{H}(w_{C\text{-}H})$ -160.38 $^{s}\boldsymbol{f}_{6}{}^{H}(w)$ (**16**)	
	getaway + whim + top.	3	98.58	99.23	2.439		$^{v}I_{D}{}^{M}$ Ts $HATS_1$(m)	53
	getaway	3	97.42	98.62	3.259		$HATS_4$(u) $HATS_7$(v) R_7(p)	53
	topological	3		98.05	3.855		$S\chi^{1}W$ $\chi^{7}W$ $\chi^{3}W$	58
	getaway + whim + top.	2	96.77	97.68	4.053		Ts H_4(e)	53
	getaway	2	91.28	95.78	5.466		$HATS_7$(m) R_4(u)	53
	topological	2		95.64	5.533		$S\chi^{1}W$ $S\chi^{3}W$	58
	topological	1		95.22	5.589		$X^{7}W$	58
	getaway + whim + top.	1	90.83	92.40	7.069		Ts	53
	topological	1		91.97	7.270		I_{wD}	55
	getaway	1	85.64	88.98	8.515		REIG	53
Heat of Vaporization (HV)	NonStochastic Bond-based Linear indices	3	95.53	97.57	0.348	187.03	**HV** = 156.95 -0.63 $\boldsymbol{f}_2{}^{H}(w)$ +0.003 $\boldsymbol{f}_4(w)$ +0.05 $\boldsymbol{f}_3{}^{H}(w)$ (**17**)	
	Stochastic Bond-based Linear indices	3	96.51	97.92	0.321	220.17	**HV** = 127.48 -4.20 $^{s}\boldsymbol{f}_2{}^{H}(w)$ -5.03 $^{s}\boldsymbol{f}_{1L}{}^{H}(w_{C\text{-}H})$ +3.36 $^{s}\boldsymbol{f}_{3L}{}^{H}(w_{C\text{-}H})$ (**18**)	
	getaway + whim + top.	3	97.57	98.42	0.281		$^{0}\overline{\chi}\,{}^{3}\kappa R^{+}{}_{6}$(u)	53
	getaway	3	95.46	97.18	0.375		$HATS_6$(u) R_4(u) $R^{+}{}_{1}$(m)	53
	getaway + whim + top.	2	95.18	96.53	0.402		$^{2}\chi$ $R^{+}{}_{6}$(u)	53
	topological	3		95.65	0.459		$\chi^{1}W$ $\chi^{2}W$ $\chi^{3}W$	58
	getaway	2	93.15	94.87	0.488		$HATS_4$(u) R_6(e)	53
	topological	2		92.62	0.577		^{4}W ^{5}W	58
	topological	1		91.78	0.429		Z	55
	getaway + whim + top.	1	80.80	88.61	0.705		$^{2}\chi$	53
	getaway	1	79.74	85.70	0.790		R_2(m)	53
	topological	2		84.27	0.820		WW x_1	56
Molar Volume (MV)	NonStochastic Bond-based Linear indices	3	98.29	99.12	0.265	488.99	**MV** = 76.65 -0.23 $\boldsymbol{f}_3{}^{H}(w)$ +1.45 $\boldsymbol{f}_2{}^{H}(w)$ -0.06 $\boldsymbol{f}_2(w)$ (**19**)	
	Stochastic Bond-based Linear indices	3	87.79	92.75	0.761	55.442	**MV** = 145.71 +1.05 $^{s}\boldsymbol{f}_1(w)$ -4.01 $^{s}\boldsymbol{f}_{3L}{}^{H}(w_{C\text{-}H})$ +4.28 $^{s}\boldsymbol{f}_{2L}{}^{H}(w_{C\text{-}H})$ (**20**)	
	getaway + whim + top.	3	75.96	92.01	1.825		Ks $R^{+}{}_{6}$(u) RT^{+}(m)	53
	getaway	3	69.27	90.33	2.008		$HATS_6$(p) RT^{+}(m) R_1(v)	53
	topological	3		88.29	2.210		^{5}W ^{6}W ^{7}W	58
	getaway + whim + top.	2	54.49	84.96	2.419		$^{v}I_{D}{}^{M}$ $R^{+}{}_{6}$(u)	53
	getaway	2	45.49	81.79	2.662		$R^{+}{}_{6}$(u) R_4(v)	53
	getaway + whim + top.	1	32.66	67.61	3.437		R_6(v)	53
	topological	2		62.76	3.807		^{3}W ^{4}W	58
	topological	1		60.85	3.780		^{7}W	58

Table 1. *Cont.*

Property	**Method**	**size**	$\boldsymbol{Q^2_{LOO}}$	$\boldsymbol{R^2}$	***s***	**F**	**Model Descriptors**	**Ref.**
Entropy (S)	NonStochastic Bond-based Linear indices	1	90.33	92.48	1.277	196.67	**S** = 117.55 -0.09 $\boldsymbol{f}_2(w)$ (**21**)	
	Stochastic Bond-based Linear indices	1	91.63	93.51	1.185	230.77	**S** = 197.04 -6.53 $^{s}\boldsymbol{f}_3{}^{H}(w)$ (**22**)	
	getaway + whim + top.	3	97.17	97.96	0.711		$^{v}I_{D,deg}$ TWC $R^{+}{}_{2}$(p)	53
	getaway + whim + top.	2	96.42	97.14	0.814		$^{v}I_{D,deg}$ TWC	53
	getaway	3	93.45	95.84	1.016		I_{SH} $HATS_8$(m) R_3(v)	53
	getaway	2	92.19	94.76	1.101		I_{SH} R_3(v)	53
	getaway + whim + top.	1	89.86	92.51	1.274		R_3(v)	53
	topological	1		91.10	1.400		$\chi^{[1/2]}$	55
	topological	2		81.72	2.060		x_1 x_2	56
Heat of Formation (Δ_fH)	NonStochastic Bond-based Linear indices	2	91.30	92.23	0.371	88.975	$\boldsymbol{\Delta_f H}$ = 8.05 -0.32 $\boldsymbol{f}_2{}^{H}(w)$ +2.06x10^{-10} $\boldsymbol{f}_{15}{}^{H}(w)$ (**23**)	
	Stochastic Bond-based Linear indices	3	86.56	91.45	0.403	49.928	$\boldsymbol{\Delta_f H}$ = 23.91 -2.23 $^{s}\boldsymbol{f}_{10}(w)$ -51.58 $^{s}\boldsymbol{f}_9{}^{H}(w)$ +49.45 $^{s}\boldsymbol{f}_8{}^{H}(w)$ (**24**)	
	getaway + whim + top.	3	95.06	96.60	0.254		$HATS_5$(m) $HATS_7$(m) R_4(e)	53
	getaway + whim + top.	2	90.96	93.24	0.346		$^{2}\chi$ $HATS_2$(e)	53
	getaway	2	90.18	92.87	0.356		$HATS_7$(u) R_2(m)	53

getaway + whim + top.	1	87.18	89.34	0.421	$HATS_2$(m)	53
topological	3		87.05	0.492	$\Omega_1 \Omega_2 \Omega_3$	54
topological	2		86.86	0.478	$\Omega_1 \Omega_2$	54
topological	1		86.68	0.471	$1/^2\chi$	55
topological	2		78.70	0.570	WW x_1	56

)

As can be appreciated from the statistical parameters of regression equations in Table 1, all of the physicochemical properties were well described by bond-based linear indices. In this table we can observe that the statistical parameters for the models obtained with bond-based linear indices to describe motor octane number (MON) (Eqs **15** and **16**) and molar volume (MV) (Eqs **19** and **20**) of octanes are better than those taken from the literature. The first physicochemical property, that is, MV, is well-described exclusively by the bond-based linear indices. Note also that in the models based on the bond-level chemical linear indices, the two regressions for the heat of vaporization (HV) (Eqs **17** and **18**) are better-to-similar than the models published so far. Only the models found by us to describe boiling point (BP) (Eqs **13** and **14**), entropy (S) (Eqs **21** and **22**), and heat of formation ($\Delta_f H$) (Eqs **23** and **24**) have significant differences with the precedent models obtained by applying the selection procedure to the set given by GETAWAY descriptors plus WHIM and topological indices.

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

4. CONCLUDING REMARKS

The total and local (bond and bond-type) linear indices of the non-stochastic and stochastic edge adjacency matrices are novel sets of graph-theoretical descriptors. These indices have a series of important features that make them useful molecular descriptors to be employed in QSPR/QSAR studies, similarity/diversity analysis and drug design protocols. The correlations found by these new sets of bond-level chemical descriptors for the description of six representative physicochemical properties of octane isomers can be considered as statistically significant. The approach described in this paper appears to

be a promissory method to find *in silico* models for description of physical, chemical and biological properties. Applications of theses new descriptors in molecular property/activity modeling, similarity/diversity analysis and biosilico drug discovery will be published in subsequent papers.

ACKNOWLEDGMENTS: The author thanks the program "Estades Temporals per a Investigadors Convidats" for a fellowship to work at Valencia University. The authors acknowledges finantial support from the Spanish MEC DGI (Proyect No. CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INFO1-051 and INFRA03-047, and OCYT GRUPOS03-173).

5. REFERENCES AND NOTES

(1) Todeschini, R.; Consonni, V. (Eds.) *Handbook of Molecular Descriptors*; Wiley-VCH: Weinhein (Germany), 2000.
(2) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons: New York. 2000.
(3) Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem*. **2001**, *8*, 1699-1714.
(4) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J.; Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999.
(5) Torrens, F. Structural, Chemical Topological, Electrotopological and Electronic Structure Hypotheses. *Comb. Chem. High Throughput. Screen*. **2003**, *6*, 801-809.
(6) Randić, M. On Characterization of Chemical Structure. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 672-672.
(7) Estrada, E. Generalization of Topological Indices. *Chem Phys. Lett*. **2001**, *336*, 248-252.
(8) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 31-33.
(9) Estrada, E.; Ramírez, A. Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 837-843.
(10) 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.
(11) Marković, S.; Gutman, I. Dependence of Spectral Moments of Benzenoid Hydrocarbons on Molecular Structure. *J. Mol. Struct.* (*Theochem*). **1991**, *235*, 81-87.
(12) 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.
(13) 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.
(14) 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.
(15) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999.
(16) 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.
(17) 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.

(18) Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón Peréz, 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.
(19) Marrero-Ponce, Y.; Medina-Marrero, R.; Martinez, 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.
(20) 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.
(21) 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)
(22) 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.
(23) 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.
(24) Rouvray, D. H. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; pp 180-181.
(25) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; 2nd ed.; 1992; pp 32-33.
(26) Edwards, C.H.; Penney, D. E. *Elementary Linear Algebra*; Prentice-Hall, Englewood Cliffs: New Jersey, USA, 1988.
(27) 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.
(28) 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*. In Press (DOI: 10.1021/ci050085t).
(29) 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.
(30) 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.
(31) 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.
(32) 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.
(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.; 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.
(35) 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.
(36) Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. In Silico Studies Toward the Discovery of New Anti-HIV Nucleoside Compounds with the Use of TOPS-MODE and 2D/3D Connectivity Indices. 1. Pyrimidyl Derivatives. *J. Chem. Inf. Comput. Sci.* **2002**, *42,* 1194-1203.
(37) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for Virtual Screening and Rational Design of Anticancer Compounds*. J. Med. Chem*. **2000**, *43*, 1975-1985**.**
(38) Estrada, E.; Peña, A.; García-Domenech, R. J. Designing Sedative/Hypnotic Compounds from a Novel Substructural Graph-Theoretical Approach. *J. Comput.-Aided Mol. Des*. **1998**, *12*, 583–595.
(39) Potapov, V. M. *Stereochemistry*; Mir: Moscow, 1978.
(40) Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Perspect. Drug Discov. Des.* **2000**, *19*, 47-66.
(41) 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.
(42) 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.
(43) Millar, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533-8542.
(44) Gasteiger, J.; Marsilli, M. A New Model for Calculating Atomic Charge in Molecules. *Tetrahedron Lett.* **1978**, *34*, 3181-3184.
(45) Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: Ithaca (New York), **1939**; 2-60.
(46) Browder, A. *Mathematical Analysis. An Introduction.* 1996, Springer-Verlag, New York, Inc. pp 176-296
(47) Axler, S. *Linear Algebra Done Right.* Springer-Verlag: New York, 1996, pp 37-70.
(48) Daudel, R.; Lefebre, R.; Moser, C. *Quantum Chemistry: Methods and Applications*. Ed. Wiley, New York. 1984.
(49) Klein, D. J. Graph Theoretically Formulated Electronic–Structure Theory *Internet Electron. J. Mol. Des.* **2003**, *2*, 814–834.
(50) Randić, M.; Trinajstić, N. Viewpoint 4-Comparative Structure-Property Studies: The Connectivity Basis. *J. Mol. Struct.* (*Theochem*). **1993**, *284*, 209-221.
(51) Randić, M.; Trinajstić, N. In Search for Graph Invariants of Chemical Interest. *J. Mol. Struct.* (*Theochem*). **1993**, *300*, 551-572.
(52) 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.
(53) 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.
(54) Randić, M. Correlation of Enthalpy of Octanes with Orthogonal Connectivity Indices. *J. Mol. Struct.* (*Theochem*). **1991**, *233*, 45-59.
(55) Randić, M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chim. Acta* **1993**, *66*, 289-312.
(56) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361-367.
(57) Diudea, M. V. Walk Numbers eWM: Wiener-Type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 535-540.
(58) Diudea, M. V.; Minailiuc, O. M.; Katona, G. Molecular Topology. 26. SP Indices: Novel Connectivity Descriptors. *Rev. Roum. Chim*. **1997**, *42*, 239-249.
(59) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem*. **1991**, *7*, 155-168.

(60) 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.
(61) Marrero-Ponce Y, Romero V (2002) **TOMOCOMD** software. Central University of Las Villas. **TOMOCOMD** (**TO**pological **MO**lecular **COM**puter **D**esign) 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
(62) Goldberg, D. E. *Genetic Algorithms*, Addison Wesley, Reading, MA, 1989.
(63) Willet, P. Genetic Algorithms in Molecular Recognition and Design. *Trends Biotechnol*. **1995**, *13***,** 516–521.
(64) 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.
(65) 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.
(66) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci*. **1994,** *34*, 854-866.
(67) 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.
(68) 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.
(69) Liu, J.; Pan, D.; Tseng, Y.; Hopfinger, A. J. 4D-QSAR Analysis of a Series of Antifungal P450 Inhibitors and 3D-Pharmacophore Comparisons as a Function of Alignment. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 2170-2179.
(70) 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.
(71) De Oliveira, D. B.; Gaudio, A. C. BuildQSAR: A New Computer Program for QSAR Studies. *Quant. Struct.-Act. Relat.* **2000**, *19*, 599-601.
(72) 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.