12th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-12) 1-30 November 2008
http://www.usc.es/congresos/ecsoc/12/ECSOC12.htm & http://www.mdpi.org/ecsoc-12

[G0009]

# Bond-Extended Stochastic and Non-Stochastic Bilinear Indices. 1. QSPR/QSAR Applications to the Description of Properties/Activities of Small-Medium Size Organic Compounds

**Yovani Marrero-Ponce,[a,b,c]* Eugenio R. Martínez,[a,d] Gerardo M. Casañola-Martín,[a] Facundo Pérez-Giménez,[c] Yunaimy Echevería Díaz,[a] & José E. Rodriguez Borges[e]**

[a]Unit of Computer-Aided Molecular "*Biosilico*" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

[b]Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, Polígon la Coma s/n, E-46071 Valencia, Spain.

[c]Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain.

[d]Department of Chemistry, Faculty of Chemistry-Pharmacy. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

[e]Associação para o desenvolvimento da Faculdade de Ciências do Porto, Rua do Campo Alegre, 687, 4169-007, Porto. Portugal

*_**Corresponding author:**_

**Fax:** 53-42-281130 [or 53-42-281455] (Cuba) and 963543156 (València)
**Phone:** 53-42-281192 [or 53-42-281473] (Cuba) and 963543156 (València)
**Cell**: 610028990
***e*-mail**: yovanimp@uclv.edu.cu , ymarrero77@yahoo.es or ymponce@gmail.com
**URL**: http://www.uv.es/yoma/

**Running head:** *Bond-Extended Bilinear Indices & QSPR Applications…*

## ABSTRACT

Bond-extended stochastic and non-stochastic bilinear indices are introduced in this paper as novel bond-level molecular descriptors (MDs). These novel totals (whole-molecule) MDs are based on a bilinear maps (forms) similar to use defined in linear algebra. The proposed non-stochastic indices try to match molecular structure provided by the molecular topology by using the $k^{th}$ Edge(Bond)-Adjacency Matrix ($\mathbf{E}^k$, designed here as non-stochastic **E** matrix). The stochastic parameters are computed by using the $k^{th}$ stochastic edge-adjacency matrix, $\mathbf{ES}^k$, as matrix operators of bilinear transformations. This new edge (bond)-adjacency relationships can be obtained directly from $\mathbf{E}^k$ and can be consider like a new matrix-transformation strategic to obtain new relation for a molecular graph. In both set of MDs, chemical information is codified by using different pair combinations of atomic weightings (in this case four atomic-labels: atomic mass, polarizability, van der Waals volume, and electronegativity). In addition, a local-fragment (bond-type) formalism was also developed. The $k^{th}$ bond-type bilinear indices are calculated by summing the $k^{th}$ bond bilinear indices of all bonds of the same bond type in the molecules. The new set of MDs can be easily and quickly calculate in our *in house* software ***TOMOCOMD-CARDD*** (***TO***pological ***MO***lecular ***COM***puter **D**esign ***C****omputer-****A****ided –****R****ational– ****D****rug ****D****esign*). The reported application and utilization of these MDs for predictive capability correlations of structure with physicochemical and pharmacology properties are reviewed. Three *benchmark* datasets have been used to evaluate the QSPR/QSAR behavior of the new bond-level ***TOMOCOMD-CARDD*** MDs. We developed the QSPR models to describe several physicochemical properties of octane isomers (**FIRST CASE**) and, to analyze of the boiling point of 28 alkyl-alcohols (**SECOND CASE**) and to examine of the specific rate constant (log *k*), the partition coefficient (log *P*), as well as the antibacterial activity of 34 derivatives of 2-furylethylenes (**THIRD CASE**). For these three rounds, the quantitative models found are significant from a statistical point of view and permit a clear interpretation of the studied properties in terms of the structural features of molecules. A leave-out-out cross-validation procedure revealed that the regression models had a good predictability. The comparison with other approaches reveals good performance of the method proposed. Therefore, it is clearly demonstrated that this suitability is higher than that shown by other 2D/3D well-known sets of MDs. The approach described here appears to be a very promising structural invariant, useful for QSPR/QSAR studies and shown to provide an excellent alternative or guides for discovery and optimization of new *lead* compounds, reducing the time and cost of traditional procedure.

**Keywords:** Octane Isomers, Alkyl-alcohol, 2-Furylethylene, Physicochemical Properties, Antibacterial Activity, QSPR/QSAR Model, Multiple Linear Regression, Linear Drisciminant Analysis, Edge-Adjacency Matrix, Stochastic Edge-Adjacency Matrix, Non-Stochastic and Stochastic Bond-Based Bilinear Indices, *TOMOCOMD-CARDD* Software.

## Introduction

In any process of molecular modelling (e.g., QSPR/QSAR studies, ligand-based virtual screening, and so on), the need for a representation of molecular structure is critical and its role is significant to find appropriate predictive models. An information-rich representation which is rapidly computed and readily manipulated is essential.[1] This is the case of the so-called topological (as well as topo-chemical) indices (TIs), which are among the most useful *molecular descriptors* (MDs) known nowadays.[2-6] TIs are "numerical values associated with chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity or biological activity" and these MDs are parameters derived from graph-theoretical invariants.[5,7,8] That is, they are numbers calculated from a molecular graph representing a molecule, which does not depend on the numbering of the graph vertices or edges.

Several TIs have been introduced to date. A compilation by Todeschini and Consonni systematizes more than 1600 MDs for small-molecule drug discovery.[9] There are two main sources of TIs, the Distance (**D**) and adjacency (**A**) matrices, but the number and diversity of graph invariants is so wide that this makes it difficult to find general relations for the indices so derived.[10] However, some of these MDs are redundant in some way or have topics in common. For instance, many researchers define TIs for graphs by using vector-matrix-vector (VMV) procedures, a fact that indicates significant similarities between these systems.[6,10] Indeed, the first TIs ever defined in a chemical context, the Wiener index (W)[11] can be calculate by using the same mathematical formalism (VMV)[12,13] that invariants based on the sum of vertex degrees products, e.g., Randic index;[14] although there is no apparent relation between these invariants. In fact, several other TIs can be written in a VMV form and these include:[9,13,15]

$$W = \frac{1}{2}\sum_{i}\sum_{j} d_{ij} = \frac{1}{2}\left(uDu^{T}\right) \tag{1}$$

$$M_1 = \sum_{i}\delta_i^2 = vAu^{T} \tag{2}$$

$$M_2 = \sum_{k=1}^{m}(\delta_i\delta_j)_k = \frac{1}{2}\left(vAv^{T}\right) \tag{3}$$

$$H = \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n} d_{ij}{}^{-k} = \frac{1}{2}\left(uD^{-k}u^{T}\right) \tag{4}$$

$$\chi = \sum_{k=1}^{m}(\delta_i\delta_j)_k^{-0.5} = \frac{1}{2}\left(v'Av'^{T}\right) \tag{5}$$

$$\chi^{v} = \sum_{k=1}^{m}(\delta_i^{v}\delta_i^{v})_k^{-0.5} = \frac{1}{2}\left(v''Av''^{T}\right) \tag{6}$$

$$J = \frac{1}{2}\sum_{k=1}^{m}(s_i s_j)_k^{-0.5} = \frac{1}{2}C\left(d'Ad'^{T}\right) \tag{7}$$

All of the vectors and matrices used in equations 1-7 have been exhaustively explained in the literature and the reader can find further details there.[5,6,9,15] Recently, new MDs also can been expressed by means of VMV form, such as the invariants reported by González-Díaz *et al.* in MARCH-INSIDE software, but it are based on Markov Chains theory.[6,16]

More recently, one of present author (M–P, Y) also proposed new MDs by using a more elaborate approach in terms of algebraic space; witch from matrix point of view can also be expressed like VMV.[17-22] Initially, three sets of MDs, *namely atom-based non-stochastic and stochastic linear, quadratic and bilinear indices*, were introduced; and they have been defined in analogy to the quadratic, linear and bilinear mathematical maps.[19,21-23] The calculation of these novel sets of atom-level MDs can be carried out employing our *in house* ***TOMOCOMD-CARDD*** (acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign ***C***_omputer-_***A***_ided –_***R***_ational–_ ***D***_rug_ ***D***_esign_)[24] program, witch is an interactive program for molecular design and bioinformatics research, developed upon the base of a user-friendly philosophy. The last defined MDs and introduced for QSAR and drug design studies were the so-called bilinear indices.[21,25-27] These MDs are based on the calculation of bilinear maps (bilinear form) in $\Re^{n}$ in canonical basis sets. The computation of the non-stochastic and stochastic bilinear indices is develop by using the $k^{\text{th}}$ "nonstochastic and stochastic graph–theoretical electronic-density matrices" $\mathbf{M}^{\boldsymbol{k}}$ and $\mathbf{S}^{\boldsymbol{k}}$, correspondingly, as matrices of the mathematical forms.[28-30] These matricial operators are graph-theoretical electronic-structure models, like the ''extended Hückel MO model.'' The $\mathbf{M}^{1}$ matrix considers all valence-bond electrons ($\sigma$- and $\pi$-networks) in one step, and their power $k$ ($k = 0, 1, 2, 3,...$) can be considered as an interacting-electronic chemical-network in step $k$. The present approach is based on a simple model for the intramolecular (stochastic) movement of all outer-shell electrons. The theoretical scaffold of these atom-based MDs and their use to represent small-to-medium size organic chemicals as well as QSAR and drug design studies has been explained in some detail elsewhere.[21,25-27]

On the other hand, relevant atom-based ***TOMOCOMD-CARDD*** MDs that also can be formulated via VMV procedure are the non-stochastic and stochastic quadratic indices,[17,22,31-33] which were recently extended to bond(edge)-adjacency relationship codification, depicting better results than atom-level counterpart in several QSPR/QSAR studies.[34] These new extended local (bond and bond-type) and total quadratic indices have also been useful for the selection of novel molecular *subsystems* having a desired property/activity. In this sense, they were successfully applied to the virtual screening (computational discovery) of novel trichomonacidals and tyrosinase inhibitors.[34,35] Thus it is desirable to also to extend the already defined atom-based (atom-level) bilinear indices to bilinear index for bonds and bond-type as well as for whole molecule.

In fact, the edge (bond)-adjacency relationships have also been used in the generation of new TIs. In the last decade Estrada rediscovered this matrix as an important source of graph theoretical invariants useful in the generation of new MDs.[9] For instance, the $\epsilon$[36] [as well as $\epsilon(\rho)$[37] and ${}^{m}\epsilon_{t}(G)$[38]], $\mu$[39] and $P_{\epsilon}(G, x)$[40] [$\epsilon^{c}(x)$] indices were re-defined by this author[36-40] by using the Randić-type graph-theoretical invariant, spectral moment (trace) of E-matrix, edge-connectivity formalism [and long-range MDs for $x$ different from zero], respectively. That is to say, these new indices are analogous to the *tradicional* MDs but calculated by edge(bond)-relationship instead of vertex(atom)-relationship. In the present report, such edge-adjacency relationships will be applied, in order to generate a series of bond-based bilinear indices to be used in drug design and chemoinformatic studies.

Other relevant MDs have been extended to bonds codification. For example, in a manner similar to that for the atom- and atom-type level E-State,[41] an E-State index for bonds and bond-type has been proposed.[42] 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.[42] Other important example to bond-extend approach, in this case in quantum chemistry, is **LCBO-MO** (**L**inear **C**ombination of **B**ond **O**rbitals-**M**olecular **O**rbitals). **LCBO-MO** is another way of forming molecular orbitals by taking linear combinations of functions associated with the different bonds in the molecule.[43] To this effect, MOs are made up as LCBO of bonds composing the system, i.e. they are written in the form

$$\varphi_i = \sum_{j=1}^{m} c_{ij}\psi_j \qquad \textbf{(8)}$$

where, *i* is the number of the **MO** $\varphi$, *j* is the number of bond $\psi$–orbitals and $c_{ij}$ are the numerical coefficients defining the contributions of individual **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 chiefly applied to the calculation of the 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 region.[43] The great success of the bond-based MDs in QSPR/QSAR and drug design recently stimulated us to propose and validate some novel total and local parameters based on a topological (edge-adjacency relationships) characterization of the molecular structure.

Here we will present new topological (topo-chemical) indices, namely local (bond and bond-type) and total bilinear indices based on the adjacency of edges. We also propose in this report a new matrix representation of the molecule on the "stochastic" adjacency of edges and bilinear indices derived from there. The ability of the new MDs is tested in a QSPR/QSAR studies of some physicochemical properties of octanes as well as in the description and prediction of the boiling point of 28 alkyl-alcohols and to the modelling of the specific rate constant (log *k*), partition coefficient (log *P*), as well as the antibacterial activity of 34 derivatives of 2-furylethylenes. The comparison with other approaches (2D/3D edge- and vertices-based connectivity indices, total and local spectral moments, and quantum-chemical descriptors as well as E-state/biomolecular encounter parameters) will be develop in order to see the behavior of our novel method in this QSPR/QSAR studies with regard to most of the MDs reported until now.

The structure of this paper will be as follows: 1) An outline and definition of our procedures, with its peculiar mathematical properties, will be described in second section, 2) Later, the **Material and Methods** (database selection, computational approach and chemometric studies) will be present in third section, 3) The correlation equations for the properties and molecules selected, as well as the statistical considerations on the obtained results will be developed in fourth section and, 4) Finally, conclusions will be presented in the last section.

## Theoretical Formalism

The basis of the extension of bilinear indices, which will be given here, is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature,[44,45] and rediscovered by Estrada as an important source of new MDs.[36-40]

## BACKGROUND IN EDGE-ADJACENCY MATRIX AND NEW EDGE-RELATIONS: STOCHASTIC 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 $m$ edges (bonds). The edge-adjacency matrix $\mathbf{E}$ of $G$ (likewise called bond adjacency matrix, $\mathbf{B}$) is a symmetric square matrix whose elements $e_{ij}$ are 1 if and only if edge $i$ is adjacent to edge $j$.[39] Two edges are adjacent if they are incident to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the $i^{\text{th}}$ row (or column) of $\mathbf{E}$ is named the edge degree of bond $i$, $\delta(e_i)$.[36-40]

By using the edge (bond)-adjacency relationships we can find other new relation for a molecular graph that will be introduced here. The $k^{\text{th}}$ 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 matrix of order $m$ ($m$ = number of bonds) and the elements ${}^k es_{ij}$ are defined as follow:

$$ {}^{k}es_{ij} = \frac{{}^{k}e_{ij}}{{}^{k}SUM(E^{k})_{i}} = \frac{{}^{k}e_{ij}}{{}^{k}\delta(e)_{i}} \qquad \textbf{(9)} $$

where ${}^{k}e_{ij}$ are the elements of the $k^{\text{th}}$ power of **E** and the SUM of the $i^{\text{th}}$ row of $\mathbf{E}^{k}$ is named the $k$-order edge degree of bond $i$, ${}^{k}\delta(e_i)$. Notice that the matrix $\mathbf{ES}^{k}$, defined on Eq. **1,** has the property that *the sum of the elements in each row* is 1. Such an *m*x*m* matrix, with nonnegative entries having this property, is called a "stochastic matrix".[46-48]

## CHEMICAL INFORMATION AND BOND-BASED MOLECULAR VECTOR

The atom-based molecular vector ($\bar{x}$) used to represent small-to-medium size organic chemicals has been explained in some detail elsewhere.[19,21-23] In a way parallel to the development of $\bar{x}$, we present the extension to the bond-based molecular vector ($\bar{w}$). The components ($w_i$) of $\bar{w}$ are numerical values, which represent a certain standard bond property (bond label). Therefore, 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 set ($\Re^{m}$). This approach allows encoding organic molecules, such as 2-hydroxybut-2-enenitrile through the molecular vector $\bar{w}$ = [$w_{\text{Csp3-Csp2}}$, $w_{\text{Csp2=Csp2}}$, $w_{\text{Csp2-Osp3}}$, $w_{\text{H-Osp3}}$, $w_{\text{Csp2-Csp}}$, $w_{\text{Csp}\equiv\text{Nsp}}$]. This vector belongs to the product space $\Re^{6}$.

These properties characterize each kind of bond (and bond-type) within the molecule. Diverse kinds of bond weights ($w_i$) can be used to encode information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distances[49,50] and standard bond dipoles,[49,50] or even mathematical expressions involving atomic weights such as atomic electronegativities in Pauling's scale.[51] Here we characterized each bond with the following parameter:

$$w = \frac{x_i}{\delta_i} + \frac{x_j}{\delta_j} \qquad \textbf{(10)}$$

In this expression, $x_i$ can be any standard weight of the atom $i$ bonded to atom $j$. The $\delta_i$ is the vertex (atom) degree of atom $i$. Thus, chemical information can be codified by means of two different molecular vectors, for instance, $\overline{w} = [w_1,\ldots,w_n]$ and $\overline{u} = [u_1,\ldots,u_n]$, $\overline{w} \neq \overline{u}$**,** which is possible if different weighting schemes are used.

**Table 1 comes about here (see end of the document)**

In the present report, we characterized each bond with mathematical expressions (Eq. 10) involving the following parameters: atomic masses (M), van der Waals volumes (V), atomic polarizabilities (P), and atomic electronegativities (K) in Mulliken scale.[21] The values of these atomic labels are shown in Table 1. From this weighting scheme, six combinations (or twelve permutations if $\overline{w}_{\mathrm{M}}$-$\overline{u}_{\mathrm{V}} \neq \overline{w}_{\mathrm{V}}$-$\overline{u}_{\mathrm{M}}$) of molecular vectors ($\overline{w}, \overline{u}; \overline{w} \neq \overline{u}$) can be computed, namely: $\overline{w}_{\mathrm{M}}$-$\overline{u}_{\mathrm{V}}$, $\overline{w}_{\mathrm{M}}$-$\overline{u}_{\mathrm{P}}$, $\overline{w}_{\mathrm{M}}$-$\overline{u}_{\mathrm{K}}$, $\overline{w}_{\mathrm{V}}$-$\overline{u}_{\mathrm{P}}$, $\overline{w}_{\mathrm{V}}$-$\overline{u}_{\mathrm{K}}$, and $\overline{w}_{\mathrm{P}}$-$\overline{u}_{\mathrm{K}}$. Here, the symbols $\overline{w}_{\mathrm{Y}}$-$\overline{u}_{\mathrm{Z}}$ are used, where the subscripts Y and Z mean two atomic properties from our weighting scheme and a hyphen (-) expresses the pair. In order to illustrate the latter, an example will be described in a section below.

## DEFINITION OF BILINEAR FORMS: A MATHEMATICAL POINT OF VIEW

In mathematics, a *bilinear form* in a real vector space $\Re^m$ is a mapping $b: \Re^m x\, \Re^m \to \Re$, which is linear in both arguments.[21,25-27] That is to say, this function satisfies the following axioms for any scalar $\alpha$ and any choice of vectors $\overline{v}, \overline{w}, \overline{v}_1, \overline{v}_2, \overline{w}_1$ and $\overline{w}_2$ from $\Re^m$:

**i.** $b(\alpha\overline{v}, \overline{w}) = b(\overline{v}, \alpha\overline{w}) = \alpha b(\overline{v}, \overline{w})$

**ii.** $b(\overline{v}_1+\overline{v}_2,\overline{w})=b(\overline{v}_1,\overline{w})+b(\overline{v}_2,\overline{w})$

**iii.** $b(\overline{v},\overline{w}_1+\overline{w}_2)=b(\overline{v},\overline{w}_1)+b(\overline{v},\overline{w}_2)$ **(11)**

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

Let *V* be a subset of the real vector space $\Re^m\,/\,V\subset\Re^m$, and $\{\overline{e}_1,\overline{e}_2,...,\overline{e}_m\}$ the canonical basis set of $\Re^m$. This basis set permits us to write in unambiguous form any vectors $\overline{w}$ and $\overline{u}$ of *V*, where: $(w^1,w^2,...,w^m)\in V$ and $(u^1,u^2,...,u^m)\in V$ are the coordinates of the vectors $\overline{w}$ and $\overline{u}$, respectively. Namely:

$$\overline{w}=\sum_{i=1}^{m}w^i e_i \text{ and } \overline{u}=\sum_{j=1}^{m}u^j\overline{e}_j \qquad \mathbf{(12)}$$

Subsequently, applying properties **i.**, **ii.**, **iii.** (Eqs. **3**) we thus obtain:

$$b(\overline{w},\overline{u})=b(\sum_{i=1}^{m}w^i\overline{e}_i,\sum_{j=1}^{m}u^j\overline{e}_j)=\sum_{i=1}^{m}\sum_{j=1}^{m}w^iu^jb(\overline{e}_i,\overline{e}_j) \qquad \mathbf{(13)}$$

If we take the $a_{ij}$ as the *m*x*m* scalars $b(\overline{e}_i,\overline{e}_j)$ i.e.

$$a_{ij}=b(\overline{e}_i,\overline{e}_j),\quad i=\overline{1,m}\wedge j=\overline{1,m} \qquad \mathbf{(14)}$$

Then:

$$b(\overline{w},\overline{u})=\sum_{i,j}^{m}a_{ij}w^iu^j=\overline{w}^t\bullet\mathbf{A}\bullet\overline{u}=\begin{bmatrix}w^1 & ... & w^m\end{bmatrix}\bullet\begin{bmatrix}a_{11} & ... & a_{1m}\\ ... & ... & ...\\ a_{m1} & ... & a_{mm}\end{bmatrix}\bullet\begin{bmatrix}u^1\\ \vdots\\ u^m\end{bmatrix} \qquad \mathbf{(15)}$$

As it can be seen, the equation defined for *b* can be written in matrix form (see Eq. **15**), where $\overline{u}$ is a column vector (an *n*x1 matrix) of the coordinates of $\overline{u}$ in a basis set of $\Re^m$, and $\overline{w}^t$ (a *1*x*m* matrix) is the transpose of $\overline{w}$, where $\overline{w}$ is a column vector (an *m*x*1* matrix) of the coordinates of $\overline{w}$ in the same basis set of $\Re^m$.

Finally, we introduce the formal definition of symmetric bilinear form. Let $\Re^m$ be a real vector space and $b$ a bilinear function in $\Re^m$. The bilinear function $b$ is called symmetric if $b(\overline{w},\overline{u}) = b(\overline{u},\overline{w}), \quad \forall\, \overline{w},\overline{u} \in \Re^m$.[21,25-27] Then:

$$b(\overline{w},\overline{u}) = \sum_{i,j}^{n} a_{ij} w^i u^j = \sum_{i,j}^{n} a_{ji} w^j u^i = b(\overline{u},\overline{w}) \qquad (\mathbf{16})$$

## GLOBAL BOND-BASED BILINEAR INDICES

If a molecule consists of $m$ bonds (*vectors of* $\Re^m$), then the $k^{\text{th}}$ total bilinear indices are calculated as bilinear maps (bilinear forms) on $\Re^m$, in a canonical basis set. Specifically, the $k^{\text{th}}$ total non-stochastic and stochastic bond bilinear indices, $b_k(\overline{w},\overline{u})$ and ${}^sb_k(\overline{w},\overline{u})$, are computed from these $k^{\text{th}}$ non-stochastic and stochastic edge adjacency matrices, $\mathbf{E}^k$ and $\mathbf{ES}^k$, as shown in Eqs. **17** and **18**, correspondingly:

$$b_k(\overline{w},\overline{u}) = \sum_{i=1}^{m}\sum_{j=1}^{m} {}^ke_{ij}\, w^i u^j = \overline{w}^t \bullet \mathbf{E}^k \bullet \overline{u} \qquad (\mathbf{17})$$

$${}^sb_k(\overline{w},\overline{u}) = \sum_{i=1}^{m}\sum_{j=1}^{m} {}^kes_{ij}\, w^i u^j = \overline{w}^t \bullet \mathbf{ES}^k \bullet \overline{u} \qquad (\mathbf{18})$$

where $m$ is the number of bonds of in the molecule, and $w^1,\ldots,w^m$ and $u^1,\ldots,u^m$ are the coordinates of the bond-based molecular vectors $\overline{w}$ and $\overline{u}$, respectively, in the canonical basis set of $\Re^m$**.** These coordinates will in turn coincide with the vector's components, namely, $w_1,\ldots,w_m$ and $u_1,\ldots,u_m$, respectively.[21,25-27] Therefore, those coordinates can be considered as weights (bond labels) of the molecular graph's edge. The coefficients ${}^ke_{ij}$ and ${}^kes_{ij}$ are the elements of the $k^{\text{th}}$ power of the matrices **E**(G) and **ES**(G), correspondingly, of the molecular pseudograph. The defining equations (Eqs. **17** and **18**) for $b_k(\overline{w},\overline{u})$ and ${}^sb_k(\overline{w},\overline{u})$, respectively, may also be written as the single matrix equation (see Eqs. **17** and **18**), where $\overline{u}$ is a column vector (an $m$x$1$ matrix) of the coordinates of $\overline{u}$ in the canonical basis set of $\Re^m$, and $\overline{w}^t$ is the transposed of $\overline{w}$, where $\overline{w}$ is a column vector (an $m$x$1$ matrix) of the coordinates of $\overline{w}$ in the canonical basis of $\Re^m$. Here, $\mathbf{E}^k$ and $\mathbf{ES}^k$ denote the matrices of bilinear maps with regard to the natural basis set.

It should be remarked that non-stochastic and stochastic bilinear indices are symmetric and non-symmetric bilinear forms, respectively. Therefore, if in the weighting schemes, M and V are used as weights to compute theses molecular descriptors (MDs), two different sets of stochastic bilinear indices, ${}^{\text{M-V}s}b_k^{\text{H}}(\overline{w},\overline{u})$ and ${}^{\text{V-M}\,s}b_k^{\text{H}}(\overline{w},\overline{u})$ [because in this case $\overline{w}_{\text{M}} - \overline{u}_{\text{V}} \neq \overline{w}_{\text{V}} - \overline{u}_{\text{M}}$] can be obtained and only one group of non-stochastic bilinear indices ${}^{\text{M-V}s}b_k^{\text{H}}(\overline{w},\overline{u}) = {}^{\text{V-M}s}b_k^{\text{H}}(\overline{w},\overline{u})$ [because in this case $\overline{w}_{\text{M}} - \overline{u}_{\text{V}} = \overline{w}_{\text{V}} - \overline{u}_{\text{M}}$) ] can be calculated.

## LOCAL BOND-BASED BILINEAR INDICES

Finally, in addition to total bond-based bilinear indices, computed for the whole molecule, some local-fragmental (bond and bond-type as well as group-type) formalisms can be developed. These descriptors are termed as *local* non-stochastic and stochastic bilinear indices, $b_{kL}(\overline{w},\overline{u})$ and ${}^{s}b_{kL}(\overline{w},\overline{u})$, respectively. The definition of these descriptors is as follows:

$$b_{kL}(\overline{w},\overline{u}) = \sum_{i=1}^{m}\sum_{j=1}^{m} {}^{k}e_{ijL}\, w^i u^j = \overline{w}^t \bullet \mathbf{E}_L^k \bullet \overline{u} \qquad \textbf{(19)}$$

$${}^{s}b_{kL}(\overline{w},\overline{u}) = \sum_{i=1}^{m}\sum_{j=1}^{m} {}^{k}es_{ijL}\, w^i u^j = \overline{w}^t \bullet \mathbf{ES}_L^k \bullet \overline{u} \qquad \textbf{(20)}$$

where $m$ is the number of bonds, and ${}^{k}e_{ijL}$ [${}^{k}es_{ijL}$] is the $k^{\text{th}}$ element of the row "$i$" and column "$j$" of the local matrix $\mathbf{E}^k_L$ [$\mathbf{ES}^k_L$]. This matrix is extracted from the $\mathbf{E}^k$ [$\mathbf{ES}^k$]

matrix; it contains information referred to the edges (bonds) of the specific molecular fragments and also of the molecular environment, in $k$ steps. The matrix $\mathbf{E}^k_L$ [$\mathbf{ES}^k_L$] with elements ${}^k e_{ijL}$ [${}^k es_{ijL}$] is defined as follows:

${}^k e_{ij\mathrm{L}}$ [${}^k es_{ij\mathrm{L}}$] $= {}^k e_{ij}$ [${}^k es_{ij\mathrm{L}}$] *if* both $e_i$ and $e_j$ edges (bonds) are contained within the molecular fragment

$= ½\, {}^k e_{ij}$ [${}^k es_{ij\mathrm{L}}$] *if* either $e_i$ or $e_j$ is contained within the molecular fragment

$= 0$, otherwise **(21)**

It is important to highlight that the scheme above follows the spirit of the Mulliken population analysis.[52] It should be also remarked that for every partition of a molecule into Z molecular fragments there will be Z local molecular fragmental matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrices $\mathbf{E}^{\boldsymbol{k}}$ [$\mathbf{ES}^{\boldsymbol{k}}$] can be correspondingly partitioned into Z local matrices $\mathbf{E}^k_L$ [$\mathbf{ES}^k_L$], L = 1,... Z, and the $k$th power of matrix **E** [**ES**] is exactly the sum of the $k$th power of the local Z matrices. Therefore, the total (both non-stochastic and stochastic) bond-based bilinear indices are the sum of the local non-stochastic and stochastic bond-based bilinear indices, respectively, of the Z molecular fragments:

$$b_k(\overline{w},\overline{u}) = \sum_{L=1}^{Z} b_{kL}(\overline{w},\overline{u}) \qquad \textbf{(22)}$$

$$^{s}b_k(\overline{w},\overline{u}) = \sum_{L=1}^{Z} {}^{s}b_{kL}(\overline{w},\overline{u}) \qquad \textbf{(23)}$$

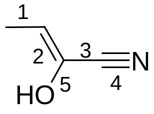
Bond and bond-type bilinear fingerprints are specific cases of local bond-based bilinear indices. The $k$th bond-type bilinear indices of the edge-adjacency matrix are calculated by adding the $k$th bond bilinear indices for all the bonds of the same type in the molecule. That is to say, this extension of the bond bilinear index is similar to group additive schemes, in which an index appears for each bond type in the molecule together with its contribution based on the bond bilinear index.

In the bond-type bilinear-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 the data sets, including those with a common molecular scaffold as well as those with rather diverse structure, the $k^{th}$ fragment (bond-type) bilinear indices provide much useful information. Thus, the development of the bond-type bilinear 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 we show a representation of the computation of the non-stochastic and stochastic bilinear indices of the bond matrix (both total and local) by using a simple chemical example.

## SAMPLE CALCULATION

The bilinear indices of the bond matrix are calculated in the following way. By considering the molecule of 2-hydroxybut-2-enenitrile as a simple example, we have the following labelled molecular graph and bond-based adjacency matrices (**E** and **ES**). The second ($k = 2$) and third ($k = 3$) powers of these matrices as well as the bond-based molecular vectors, $\bar{w}$ and $\bar{u}$ are also given:



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

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

The molecule contains five localized bonds (corresponding to the five edges in the H-suppressed molecular graph). With these, we shall associate the five “bond orbitals” $w_1$, $w_2$, $w_3$, $w_4$, and $w_5$. Thus, $\overline{w} = [w_1, w_2, w_3, w_4, w_5] = [w_{(\mathrm{C\text{-}C})}, w_{(\mathrm{C{=}C})}, w_{(\mathrm{C\text{-}C})}, w_{(\mathrm{C{\equiv}N})}, w_{(\mathrm{C\text{-}O})}]$ and each “bond orbital” can be computed by Eq. **2** by using, for instance, the atomic electronegativity in Pauling scale ($x$)[51] 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, $\overline{w} = [3.4, 1.4875, 1.275, 1.650833, 4.0775]$.

Besides, other vector, $\overline{u}$, can be calculated in the same way as $\overline{w}$, but using other property, for example the atomic mass[53] as atomic weight (atom-label):

$u_1 = y_C/1 + y_C/3 = 12.01/1 + 12.01/3 = 16.013333$

$u_2 = y_C/3 + y_C/4 = 12.01/3 + 12.01/4 = 7.005833$

$u_3 = y_C/4 + y_C/4 = 12.01/4 + 12.01/4 = 6.0050$

$u_4 = y_C/4 + y_N/3 = 12.01/4 + 14.01/3 = 7.6725$

$u_5 = y_C/4 + y_O/1 = 12.01/4 + 16.00/1 = 19.0025$

and therefore, $\overline{u} = [16.013333, 7.005833, 6.0050, 7.6725, 19.0025]$.

Each non-stochastic or stochastic total bilinear index will have the form:

$$\begin{aligned}b_k(\overline{w},\overline{u}) = &+ {}^ke_{11}w^1u^1 + {}^ke_{12}w^1u^2 + {}^ke_{13}w^1u^3 + {}^ke_{14}w^1u^4 + {}^ke_{15}w^1u^5 + {}^ke_{21}w^2u^1 + {}^ke_{22}w^2u^2\\ &+ {}^ke_{23}w^2u^3 + {}^ke_{24}w^2u^4 + {}^ke_{25}w^2u^5 + {}^ke_{31}w^3u^1 + {}^ke_{32}w^3u^2 + {}^ke_{33}w^3u^3 + {}^ke_{34}w^3u^4\\ &+ {}^ke_{35}w^3u^5 + {}^ke_{41}w^4u^1 + {}^ke_{42}w^4u^2 + {}^ke_{43}w^4u^3 + {}^ke_{44}w^4u^4 + {}^ke_{45}w^4u^5 + {}^ke_{51}w^5u^1\\ &+ {}^ke_{52}w^5u^2 + {}^ke_{53}w^5u^3 + {}^ke_{54}w^5u^4 + {}^ke_{55}w^5u^5 = \sum_{i=1}^{5} {}^ke_{ii}w^iu^i + 2\sum_{(i,j=1,i\neq j)}^{5} {}^ke_{ij}w^iu^j\end{aligned}$$

$$\begin{aligned}{}^sb_k(\overline{w},\overline{u}) = &+ {}^kes_{11}w^1u^1 + {}^kes_{12}w^1u^2 + {}^kes_{13}w^1u^3 + {}^kes_{14}w^1u^4 + {}^kes_{15}w^1u^5 + {}^kes_{21}w^2u^1\\ &+ {}^kes_{22}w^2u^2 + {}^kes_{23}w^2u^3 + {}^kes_{24}w^2u^4 + {}^kes_{25}w^2u^5 + {}^kes_{31}w^3u^1 + {}^kes_{32}w^3u^2\\ &+ {}^kes_{33}w^3u^3 + {}^kes_{34}w^3u^4 + {}^kes_{35}w^3u^5 + {}^kes_{41}w^4u^1 + {}^kes_{42}w^4u^2 + {}^kes_{43}w^4u^3\\ &+ {}^kes_{44}w^4u^4 + {}^kes_{45}w^4u^5 + {}^kes_{51}w^5u^1 + {}^kes_{52}w^5u^2 + {}^kes_{53}w^5u^3 + {}^kes_{54}w^5u^4\\ &+ {}^kes_{55}w_5u_5 = \sum_{i=1}^{5} {}^ke_{ii}w^iu^i + \sum_{(i,j=1)}^{5} {}^ke_{ij}w^iu^j\end{aligned}$$

The elements ${}^ke_{ii}$ and ${}^kes_{ii}$ can be considered a measure of the attraction of a bond for an electron in the *k* step while the elements ${}^ke_{ij}$ and ${}^kes_{ij}$ are the terms of interaction between two bonds in the *k* step. In addition, ${}^ke_{ij} = {}^ke_{ji}$ are equal by symmetry (non-oriented molecular graph). However, ${}^kes_{ij} \neq {}^kes_{ji}$. 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 they should be different in the two senses. This result is in total agreement if the electronegativities 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–theoretical electronic–structure models.[54] Actually, 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 $t_k$ ($k = 0, 1, 2, 3,\ldots, n$) through the chemical-bonding network. Therefore, the $\mathbf{E}^1$ and $\mathbf{ES}^1$ matrices consider the valence-bond electrons in one step and their power ($k = 0, 1, 2, 3,\ldots, n$) can be considered as an interacting–electron chemical–network model in *k* steps. This model can be seen as intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.[54]

On the other hand, the $k^{\text{th}}$ ($k = \overline{0,3}$) non-stochastic total bilinear indices can be expressed as the sum of the local (bond) bilinear indices for this molecule as follows:

$$b_0(\overline{w},\overline{u}) = \sum_{L=1}^{5} b_{0L}(\overline{w},\overline{u}) = b_{01}(\overline{w},\overline{u}) + b_{02}(\overline{w},\overline{u}) + b_{03}(\overline{w},\overline{u}) + b_{04}(\overline{w},\overline{u}) + b_{05}(\overline{w},\overline{u})$$

$$= 54.44533 + 10.42118 + 7.656375 + 12.66602 + 77.48269 = 162.6716$$

$$b_1(\overline{w},\overline{u}) = \sum_{L=1}^{5} b_{1L}(\overline{w},\overline{u}) = b_{11}(\overline{w},\overline{u}) + b_{12}(\overline{w},\overline{u}) + b_{13}(\overline{w},\overline{u}) + b_{14}(\overline{w},\overline{u}) + b_{15}(\overline{w},\overline{u})$$

$$= 23.81983 + 61.16852 + 43.13707 + 9.847846 + 52.77304 = 190.7463$$

$$b_2(\overline{w},\overline{u}) = \sum_{L=1}^{5} b_{2L}(\overline{w},\overline{u}) = b_{21}(\overline{w},\overline{u}) + b_{22}(\overline{w},\overline{u}) + b_{23}(\overline{w},\overline{u}) + b_{24}(\overline{w},\overline{u}) + b_{25}(\overline{w},\overline{u})$$

$$= 139.8138 + 80.10137 + 76.67535 + 55.48246 + 304.0172 = 656.0901$$

$$b_3(\overline{w},\overline{u}) = \sum_{L=1}^{5} b_{3L}(\overline{w},\overline{u}) = b_{31}(\overline{w},\overline{u}) + b_{32}(\overline{w},\overline{u}) + b_{33}(\overline{w},\overline{u}) + b_{34}(\overline{w},\overline{u}) + b_{35}(\overline{w},\overline{u})$$

$$= 183.0889 + 262.1182 + 207.3626 + 98.6209 + 462.3363 = 1213.527$$

The terms in the summations for calculating the total bilinear indices are the so-called local (bond) bilinear indices, which, in turn, have been written in the consecutive order of the bond labels in the graph. For instance, the non-stochastic bond bilinear indices of orders *0*, *1*, *2* and *3* for the bond labelled as 1 are 54.44533, 23.81983, 139.8138 and 183.0889, respectively.

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

$$^{s}b_0(\overline{w},\overline{u}) = \sum_{L=1}^{5} {}^{s}b_{0L}(\overline{w},\overline{u}) = {}^{s}b_{01}(\overline{w},\overline{u}) + {}^{s}b_{02}(\overline{w},\overline{u}) + {}^{s}b_{03}(\overline{w},\overline{u}) + {}^{s}b_{04}(\overline{w},\overline{u}) + {}^{s}b_{05}(\overline{w},\overline{u})$$

$$= 54.44533 + 10.42118 + 7.656375 + 12.66602 + 77.48269 = 162.6716$$

$$^{s}b_1(\overline{w},\overline{u}) = \sum_{L=1}^{5} {}^{s}b_{1L}(\overline{w},\overline{u}) = {}^{s}b_{11}(\overline{w},\overline{u}) + {}^{s}b_{12}(\overline{w},\overline{u}) + {}^{s}b_{13}(\overline{w},\overline{u}) + {}^{s}b_{14}(\overline{w},\overline{u}) + {}^{s}b_{15}(\overline{w},\overline{u})$$

$$= 15.87989 + 30.70998 + 19.72389 + 6.587033 + 22.01199 = 94.91277$$

$$^{s}b_2(\overline{w},\overline{u}) = \sum_{L=1}^{5} {}^{s}b_{2L}(\overline{w},\overline{u}) = {}^{s}b_{21}(\overline{w},\overline{u}) + {}^{s}b_{22}(\overline{w},\overline{u}) + {}^{s}b_{23}(\overline{w},\overline{u}) + {}^{s}b_{24}(\overline{w},\overline{u}) + {}^{s}b_{25}(\overline{w},\overline{u})$$

$$= 39.46198 + 14.31402 + 14.48064 + 14.93603 + 58.66773 = 141.8604$$

$$^{s}b_3(\overline{w},\overline{u}) = \sum_{L=1}^{5} {}^{s}b_{3L}(\overline{w},\overline{u}) = {}^{s}b_{31}(\overline{w},\overline{u}) + {}^{s}b_{32}(\overline{w},\overline{u}) + {}^{s}b_{33}(\overline{w},\overline{u}) + {}^{s}b_{34}(\overline{w},\overline{u}) + {}^{s}b_{35}(\overline{w},\overline{u})$$

$$= 23.20039 + 22.578 + 17.14819 + 13.09528 + 40.77731 = 116.7992$$

## Material and Methods

### DATABASE SELECTION

Three *benchmark* datasets have been used to evaluate the QSPR/QSAR behavior of the new bond-level ***TOMOCOMD-CARDD*** MDs. With this objective in mind, we developed the QSPR models to describe several physicochemical properties of octane isomers (**FIRST CASE**)[55,56] and, to analyze of the boiling point of 28 alkyl-alcohols (**SECOND CASE**)[42,57] and to examine of the specific rate constant (log *k*), the partition coefficient (log *P*), as well as the antibacterial activity of 34 derivatives of 2-furylethylenes (**THIRD CASE**).[57,58]

The use of octanes (**FIRST CASE**), as a very suitable data set for testing TIs, has been advocated by Randić and Trinajstić and at present is consider by International

Academy of Mathematical Chemistry like a benchmark database for comparison among old (well-know) and new MDs.[59,60] In fact, this dataset has been used by several researchers to evaluate the modeling power of their new MDs.[40,56,61-66] This selection is recommended, because the 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 octane's are used as a data set.[67] On the other hand; all TIs are designed to have (gradual) increments with the increments in the molecular weight. In 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.[40,61] 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.[40,61] 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.[68] 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.

In order to illustrate the possibilities of our approach in the QSPR studies of heteroatomic molecules, we have selected the boiling point of 28 alkyl-alcohols

(**SECOND CASE**) to be investigated.[42,57] This data set was firstly studied by Kier and Hall[42] using E-state/biomolecular encounter parameters and later by Estrada and Molina[57] employing the local spectral moments of the edge adjacency matrix. This heteromolecules-based database is composed by 28 alkyl-alcohols, 14 are primary, 6 secondary and 8 tertiary, for which the boiling point (Bp) has been reported previously. Alcohols constituted a good set of chemicals for comparative study, because it is a isomeric data set, which are heteroatomic compounds and boiling point not only depend of *gradual* variation of molecular weight, but also of H-bonding capacity and R-group type. Additionally, QSPR studies are available for comparison purpose.[42,57]

The second heteromolecule-based database that will be studied here consists of by a set of 34 2-furylethylene derivatives (**THIRD CASE**), early studied by using total and local spectral moments, 2D/3D connectivity indices (vertex and edge ones) and two quantum-chemical descriptors.[57,58] These chemicals have different substituents at position 5 of the furan ring, as well as at the β position of the exocyclic double bond.[69] The values of the *n*-octanol/water partition coefficient (log $P$) and rate constant (log $k$) (for nucleophilic addition of the mercaptoacetic acid) of these compounds have been experimentally determined and reported in the literature.[69] The antibacterial activity of these compounds was determined as the inverse of the concentration $C$ that produces 50% of growth inhibition in *E. coli* at six different times and reported as log(1/$C$).[69] This antibacterial activity was used to classify furylethylenes into two groups by Estrada and Molina.[58] The group of active compounds is composed of those substance having values of log(1/$C$) < 3, while the rest form the group of inactive molecules. In this study, we also taken into account a series of nine new 2-furylethylenes, used by Estrada and Molina[58] as external prediction (test) set. These compounds have an $NO_2$ group at position $R_3$ and a Br or I at positions $R_1$ and/or $R_2$. All these compounds showed antibacterial activity in different assays.[70]

## COMPUTATIONAL APPROACH

**TOMOCOMD** is an interactive program for molecular design and bioinformatic research.[71] It consists of four subprograms; each one of them allows drawing the structures (drawing mode) and calculating molecular 2D/3D (calculation mode) descriptors. 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, salient features are outlined concerning with only one of these subprograms, **CARDD,** and with the calculation of non-stochastic and stochastic 2D bond-based bilinear indices.

***Work Methodology.*** The main steps for the application of the present method in QSAR/QSPR and drug design can be briefly summarized in the following algorithm: 1) Draw the molecular pseudographs for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbols belonging to the different groups in the periodic table of the elements, 2) Use appropriated atomic properties in order to weight and differentiate the molecular bonds. In this study, the used properties are those previously proposed for the calculation of the DRAGON descriptors,[9,53,56] i.e., atomic mass (M), atomic polarizability (P), atomic Sanderson electronegativity (K), van der Waals atomic volume (V).The values of these atomic labels are shown in Table 1. In order to calculate the required weights, it is used the mathematical expression given by Eq. 10, which involves atomic weights. 3) Compute the total and local (bond and bond-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 to calculate the molecular indices. This software generates a table, in which the rows correspond to the compounds, and columns correspond to the total and local bond-based linear indices or other family of molecular descriptors implemented in this program. 4) Find a QSPR/QSAR equation by using several multivariate analytical techniques, such as multilinear regression analysis (MRA), neural networks (NN), linear discrimination analysis (LDA), and so on. That is to say, we can find a quantitative relation between an activity A and the bilinear indices having, for instance, the following appearance, A = $a_0b_0(\overline{w},\overline{u}) + a_1b_1(\overline{w},\overline{u}) + a_2b_2(\overline{w},\overline{u}) + \ldots + a_kb_k(\overline{w},\overline{u}) + c$, where A is the measured activity, $b_k(\overline{w})$ are the $k^{th}$ total bond-based bilinear indices, and the $a_k$ and c are the coefficients obtained by the linear regression analysis. 5) Test the robustness and predictive power of the QSPR/QSAR equation by using internal (cross-validation) and external (using a test set and an external predicting set) validation techniques

The common (for the three QSPR/QSAR experiments) bond–based ***TOMOCOMD-CARDD*** MDs, computed in this study, were the following:

1) $k^{th}$ $\left(k=\overline{0,15}\right)$ total non-stochastic bond-based bilinear indices, not considering and considering H-atoms in the molecular graph (G) [$^{Y\text{-}Z}b_k(\overline{w},\overline{u})$ and $^{Y\text{-}Z}b_k^{H}(\overline{w},\overline{u})$], respectively.
2) $k^{th}$ $\left(k=\overline{0,15}\right)$ total stochastic bond-based bilinear indices, not considering and considering H-atoms in the molecular graph (G) [$^{Y\text{-}Zs}b_k(\overline{w},\overline{u})$ and $^{Y\text{-}Zs}b_k^{H}(\overline{w},\overline{u})$], respectively.

In addition, for every case we computed other more specific total or local ***TOMOCOMD-CARDD*** MDs, such as:

i) **CASE 1**

3) $k^{\text{th}}$ $\left(k=\overline{0,15}\right)$ bond-type (C-H in methyl group) non-stochastic and stochastic bilinear indices considering H-atoms in the molecular graph (G) [${}^{\text{Y-Z}}b_{kL}{}^{\text{H}}(\overline{w}_{\text{C-H}},\overline{u}_{\text{C-H}})$ and ${}^{\text{Y-Zs}}b_{kL}{}^{\text{H}}(\overline{w}_{\text{C-H}},\overline{u}_{\text{C-H}})$, correspondingly]. These local descriptors are calculated taken into account only one of the three bond-types for carbon-hydrogen bonds ($\text{C}_{\text{primary}}$-H).

b) **CASE 2**

ii) $k^{\text{th}}$ $\left(k=\overline{0,15}\right)$ local (bond-type: O in hydroxyl group) non-stochastic [and stochastic] bilinear indices considering and not-considering H-atoms and not in the molecular graph (G) {${}^{\text{Y-Z}}b_{kL}{}^{\text{H}}(\overline{w}_{\text{O}},\overline{u}_{\text{O}})$ [${}^{\text{Y-Zs}}b_{kL}{}^{\text{H}}(\overline{w}_{\text{O}},\overline{u}_{\text{O}})$] and ${}^{\text{Y-Z}}b_{kL}(\overline{w}_{\text{O}},\overline{u}_{\text{O}})$ [${}^{\text{Y-Zs}}b_{kL}(\overline{w}_{\text{O}},\overline{u}_{\text{O}})$], in the same way}

iii) **CASE 3**

5) $k^{\text{th}}$ $\left(k=\overline{0,15}\right)$ bond-type (group = heteroatoms: S, N, O) non-stochastic [and stochastic] bilinear indices, not considering and considering H-atoms in the molecular graph (G) {${}^{\text{Y-Z}}b_{kL\text{E}}(\overline{w}_{\text{E}},\overline{u}_{\text{E}})$ [${}^{\text{Y-Zs}}b_{kL\text{E}}(\overline{w}_{\text{E}},\overline{u}_{\text{E}})$] and ${}^{\text{Y-Z}}b_{kL\text{E}}{}^{\text{H}}(\overline{w}_{\text{E}},\overline{u}_{\text{E}})$ [${}^{\text{Y-Zs}}b_{kL\text{E}}{}^{\text{H}}(\overline{w}_{\text{E}},\overline{u}_{\text{E}})$]}, respectively. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors character.

6) $k^{\text{th}}$ $\left(k=\overline{0,15}\right)$ bond-type {group = exocyclic double bond [atoms 2, 6 and 7, or for the bonds defined by these atoms ($C_2$-$C_6$ and $C_6$-$C_7$)]} non-stochastic [and stochastic] bilinear indices not considering and considering H-atoms in the molecular graph (G), respectively. These atoms are those involved in the exocyclic double bond of the 2-furylethylene and these are the "target" of the nucleophilic attack by the thiol (mercapto) group.[57]

## CHEMOMETRIC STUDIES

The whole set of new MDs were used as independent variables for derived QSPRs/QSARs by using multiple linear regression (MLR) technique. The STATISTICA[72] software was employed to perform variable selection and QSPR modeling. 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). The quality of models was also determined by examining the Leave-One-Out (LOO) cross-validation (CV) ($q^2$, $s_{cv}$).[73,74] In recent years, the LOO press statistics (e.g., $q^2$) have been used as a means of indicating predictive ability. Many authors[73-75] 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 a QSAR model.

## Results and Discussion

## CASE 1. PHYSICOCHEMICAL PROPERTIES OF OCTANE ISOMERES

In the first experiment, the octane isomers and several of their physicochemical properties was analyzed. However, to evaluate the quality of the models based on our new bond-level chemical descriptors we have taken as the reference only six physicochemical properties selected in the previous study[56] Therefore, 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. The regressions of octane properties, based on the non-stochastic and stochastic bond/based bilinear indices, will be compared to some regressions based on 2D (topological/topo-chemical) and 3D (geometrical) MDs, taken from the literature.[56]

The best linear models, found using non-stochastic and stochastic total and bond-type bilinear indices are presented in Table 2. For each selected property of octane isomers, the statistical information for the best regressions with 1, 2, and 3 MDs published so far[56] are also depicted in Table 2. Together with the LOO cross-validation-explained variance ($q^2_{\mathrm{LOO}}$), the determination coefficient ($R^2$), the standard deviation the error ($s$), and Fischer ratio (F) are listed. The MD symbols are reported in eighth column, and the last column contains the references of the models taken from the literature.

**Table 2 comes about here (see end the document)**

As can be appreciated from the statistical parameters of regression equations in Table 2, all of the physicochemical properties were well described by bond-based bilinear indices. In this table we can observe that the statistical parameters for the models, obtained with bond-based bilinear indices to describe motor octane number (MON) (Eqs. **26** and **27**), molar volume (MV) (Eqs. **30** and **31**) and heat of vaporization (HV) (Eqs. **28** and **29**) of octanes, are better than those taken from the literature. The physicochemical property MV is well-described ***exclusively*** by the bond-based bilinear indices. It is remarkable, that for all properties the bond-level bilinear indices are better than atom-based MDs. Notice also that in the models based on the bond-level chemical bilinear indices, the two regressions for the boiling point (BP) (Eqs. **24** and **25**) are similar than the models published so far.[56]

Only the models found by us to describe entropy (S) (Eqs. **32** and **33**), and heat of formation ($\Delta_f$H) (Eqs. **34** and **35**) have significant differences with the precedent models obtained by applying the selection procedure to the set given by GETAWAY descriptors plus WHIM and all the TIs. However, in the later property, the atom-based bilinear indices showed the best results.

According to the obtained QSPR results, it is possible to conclude that the new MDs 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 in different properties of octanes.

Finally, we also model others properties of this set of molecules. The obtained QSPRs together with the statistical parameters are depicting in Table 3. As can see, all properties were well described by using three-parameters in every model. Always, the non-stochastic indices were better than stochastic form.

**Table 3 comes about here (see end the document)**

## CASE 2. BOILING POINT OF 28 ALKYL-ALCOHOLS

The boiling point (Bp) of a set of 28 alkyl-alcohols (see Table 4) compiled by Kier and Hall[42] was examined using the new bond-based bilinear indices. The statistical information for the best regressions with 2, 3, and 4 parameters are depicted in Tables 5.

**Tables 4 come about here (see end the document)**

Tables 5 depict the obtained QSPR models by using non-stochastic and stochastic bond-level MDs, employing the whole weight scheme (best combination of MDs computed with four atomic properties). As can be seen, all the two-, three- and four-parameter regression equations showed rather-to-good behaviour for the description of Bp of alkyl-alcohols. It should be remarked that the two-parameter regression models had rather good predictive power (see $\boldsymbol{q}_{\mathrm{LOO}}$). In fact, the inclusion of a new MD (best three-parameter regression models) not depicted a *dramatic* improvement of statistical $R^2$, $s$, and $\boldsymbol{q}_{\mathrm{LOO}}$ values from the two-parameter equations. Similarly, the best combination of four-parameter correlation with Bp was searched to give the greatest statistical parameters values. That is, as long as the statistical parameters are concerned, any combination of two, three and four parameters can give more or less similar results. Maybe, at this level of $R^2$ values, is better to develop the statistical analysis of the obtained models, taken into consideration the LOO-press statistics, specifically $s_{cv}$.

**Tables 5 come about here (see end the document)**

In concluding, the best linear regression model obtained to describe the Bp of these chemicals, by using non-stochastic and stochastic bond-based linear indices is given below, respectively:

$$\mathbf{Bp}\ (\mathbf{°C}) = 69.302\ (\pm 2.149) - 8.31\text{x}10^{-3}(\pm\ 4.36\text{x}10^{-4})\ {}^{MV}\boldsymbol{b}_1(w,v)\ -0.802\ (\pm\ 0.032)\ {}^{MP}\boldsymbol{b}_{2L}^{H}(w,v)_E +\ 19.730\ (\pm 0.211)\ {}^{PK}\boldsymbol{b}_0^{H}(w,v)\ +9.356\text{x}10^{-7}(\pm 1.672\text{x}10^{-7})\ {}^{PK}\boldsymbol{b}_{12L}(w,v)_E \qquad \mathbf{(36)}$$

N = 28   R² = 0.999   $q^2 = 0.998$   s = 1.1°C   $s_{CV} = 1.3°C$   F(4,23) = 4174.7   $p < 0.0001$

$$\mathbf{Bp}\ (\mathbf{°C}) = 45.859\ (\pm\ 1.757) + 1.856\ (\pm\ 0.075)\ {}^{MVe}\boldsymbol{b}_4^{H}(w,v)\ -\ 17.549\ (\pm\ 0.353)\ {}^{MKe}\boldsymbol{b}_3^{H}(w,v)\ +0.138\ (\pm 0.012)\ {}^{MKe}\boldsymbol{b}_1(w,v) + 0.534\ (\pm\ 0.041)\ {}^{PVe}\boldsymbol{b}_0(w,v) \qquad \mathbf{(37)}$$

N = 28  R² = 0.999   $q^2 = 0.998$   s = 1.1°C   $s_{CV} = 1.2°C$   F(4,23) = 4404.8   $p < 0.0001$

where N is the number of compounds, R² is the determination coefficient, s is the standard deviation of the regression, $q^2$ ($s_{CV}$) is the square correlation coefficient (standard deviation) obtained from the LOO cross-validation procedure, and *F* is the Fisher ratio.

The values of experimental and calculated values of the Bp for the data set (both models) are given in Table 4, and the linear relationships between them are illustrated in Figures 1 and 2, respectively.

**Figure 1 and 2 come about here (see end the document)**

These models (Eqs. **36** and **37**) explain more than 99% of the variance of the experimental Bp values. Similar-to-inferior equations were reported by Estrada and Molina,[57] and Kier and Hall[42] by using spectral moment and E-state as MDs, correspondingly. The statistical parameters of the best equations obtained by those authors are given in Table 6 (together with others results obtained with different atom- and bond-based ***TOMOCOMD-CARDD*** MDs[34,76]). The reported models explain more than 98% and 92% of the variance of the experimental Bp values, respectively. Unfortunately, the authors (Estrada and Molina,[57] and Kier and Hall[42]) did report the result of the cross-validation. However, it is remarkable that our models explain a greater percentage of the variance of the experimental Bp values than the previously developed models do, showing a decrease in the standard error of 73.81% and 81.03%, with regard to the results previously achieved by Estrada and Molina[57] and Kier and Hall),[42] Respectively. Table 6 summarizes the statistical parameters yielded by all these approaches. As can be observed in this Table, the result obtained by using other atom- and bond-based ***TOMOCOMD-CARDD*** MDs were also included. The QSPR model derived with bond-level bilinear indices showed better-to-similar results that obtained by some of the present authors en previously studies.

**Table 6 come about here (see end the document)**

## CASE 3. 34 2-FURYLETHYLENE DERIVATIVES

**Partition Coefficient (log *P*) and Rate Constant (log *k*).** It has been clear, from structure-activity relationship studies, that the lipophilicity and the nucleophilic addition of the thiol groups of some enzymes to the exocyclic double bond of 2-furylethylene derivatives are critical for the development of their antibacterial activity.[69,77] The log *P* and log *k* of nucleophilic addition of the mercaptoacetic acid to the exocyclic double bond has an important role in the understanding of the biological behavior of these 2-furylethylene derivatives. Consequently, we shall study these parameters to compare the possibilities of molecular bond-based bilinear indices in QSPRs, and to compare these results to those obtained by Estrada and Molina,[57,58] by using topological (total and local spectral moment and 2D connectivity indices), topographic and quantum chemical descriptors. This experiment will also permit us to compare the present result with the one achieved by using some atom-based ***TOMOCOMD-CARDD*** MDs. The MDs, included in these equations, clearly pointed to the identification of the reaction centers involved in the studied chemical interaction.[57,58] That is to say, the $k^{th}$ local spectral moments calculated for the atoms 2, 6 and 7, or for the bonds defined by these atoms ($C_2$-$C_6$ and $C_6$-$C_7$) were selected as the most significant ones. These atoms are those involved in the exocyclic double bond of the 2-furylethylene and these are the "target" of the nucleophilic attack by the thiol (mercapto) group. Taking into account this logical result, we also calculated the $k^{th}$ local bilinear indices for these atoms (bonds $C_2$-$C_6$ and $C_6$-$C_7$). The best models obtained, by using these bond-type bilinear indices as MDs, together with their statistical parameters, are given below:

$$\begin{aligned}\textbf{Log k} = {} & 8.626(\pm 0.802) + 6.332\text{x}10^{-3}(\pm 5.77\text{x}10^{-4})\,{}^{MK}\boldsymbol{b}^{H}{}_{2L}(w,v)_{E} \\ & - 1.90\text{x}10^{-4}(\pm 2.29\text{x}10^{-5})\ {}^{MK}\boldsymbol{b}^{H}{}_{5L}(w,v)_{E} - 0.0495(\pm 0.0042)\,{}^{MK}\boldsymbol{b}_{3L}(w,v)_{C2\text{-}C6} \\ & + 0.0511(\pm 0.0029)\,{}^{VK}\boldsymbol{b}^{\mathbf{H}}{}_{4L}(w,v)_{C6\text{-}C7} - 0.0207(\pm 0.0013)\,{}^{VK}\boldsymbol{b}^{\mathbf{H}}{}_{5L}(w,v)_{C6\text{-}C7} \\ & + 1.34\text{x}10^{-4}(\pm 1.01\text{x}10^{-5})\,{}^{VK}\boldsymbol{b}^{\mathbf{H}}{}_{8L}(w,v)_{C6\text{-}C7} + 0.019(\pm 0.002)\,{}^{VK}\boldsymbol{b}_{3L}(w,v)_{C6\text{-}C7} \qquad \textbf{(38)}\end{aligned}$$

$N = 34 \quad R^2 = 0.985 \quad s = 0.198 \quad q^2 = 0.964 \quad s_{cv} = 0.270 \quad F(7,26) = 242.04$

$$\begin{aligned}\textbf{Log k} = {} & 5.946(\pm 0.741) - 0.4085(\pm 0.0344)\,{}^{MVe}\boldsymbol{b}^{\boldsymbol{H}}{}_{6L}(w,v)_{C2\text{-}C6} \\ & + 0.3917(\pm 0.0340)\,{}^{MVe}\boldsymbol{b}^{\boldsymbol{H}}{}_{10L}(w,v)_{C2\text{-}C6} + 0.0101\,(\pm 0.0056)\,{}^{MVe}\boldsymbol{b}_{1L}(w,v)_{C2\text{-}C6}\end{aligned}$$

$$+\ 1.6242(\pm0.1547)^{\mathrm{MPe}}\boldsymbol{b}^{\boldsymbol{H}}{}_{2\mathrm{L}}(w,v)_{\mathrm{C6\text{-}C7}} - 0.2421(\pm0.0311)^{\mathrm{VPe}}\boldsymbol{b}_{5\mathrm{L}}(w,v)_{\mathrm{C2\text{-}C6}}$$

$$+\ 0.0737(\pm0.0172)\,^{\mathrm{VKe}}\boldsymbol{b}_{1\mathrm{L}}(w,v)_{\mathrm{C2\text{-}C6}} - 8.1139(\pm1.1287)\,^{\mathrm{PKe}}\boldsymbol{b}^{\boldsymbol{H}}{}_{13\mathrm{L}}(w,v)_{\mathrm{C2\text{-}C6}} \qquad \mathbf{(39)}$$

$N = 34$ $R^2 = 0.987$ $s = 0.178$ $q^2 = 0.971$ $s_{cv} = 0.240$ $F(7,26) = 293.74$

$$\textbf{Log P} = 1.1708(\pm0.2120) + 1.99\times10^{-3}(\pm1.46\times10^{-4})\,^{\mathrm{MV}}\boldsymbol{b}_0(w,v)$$

$$- 0.0172(\pm0.0016)\,^{\mathrm{MV}}\boldsymbol{b}_{1\mathrm{L}}(w,v)_E - 2.79\times10^{-4}(\pm5.55\times10^{-5})\,^{\mathrm{MP}}\boldsymbol{b}^{\mathbf{H}}{}_4(w,v)$$

$$-4.00\times10^{-3}(\pm3.12\times10^{-4})\,^{\mathrm{MP}}\boldsymbol{b}^{\mathbf{H}}{}_4(w,v)_E + 1.092\times10^{-3}(\pm1.13\times10^{-4})\,^{\mathrm{MP}}\boldsymbol{b}^{\mathbf{H}}{}_{5\mathrm{L}}(w,v)_E$$

$$+0.2030\,(\pm0.0202)\,^{\mathrm{MP}}\boldsymbol{b}_{1\mathrm{L}}(w,v)_E + 0.0150(\pm0.0010)\,)\,^{\mathrm{MK}}\boldsymbol{b}^{\mathbf{H}}{}_{1\mathrm{L}}(w,v)_E \qquad \mathbf{(40)}$$

$N = 34$ $R^2 = 0.970$ $s = 0.140$ $q^2 = 0.910$ $s_{cv} = 0.216$ $F(7,26) = 119.14$

$$\textbf{Log P} = 0.4413(\pm0.1247) - 0.0031(\pm0.0003)\,^{\mathrm{MVe}}\boldsymbol{b}_2(w,v) + 0.0474(\pm0.0028)\,^{\mathrm{MPe}}\boldsymbol{b}_0(w,v)$$

$$- 0.4157(\pm0.0237)\,^{\mathrm{MPe}}\boldsymbol{b}^{\mathbf{H}}{}_{6\mathrm{L}}(w,v)_E + 0.0327(\pm0.0025)\,)\,^{\mathrm{MKe}}\boldsymbol{b}^{\mathbf{H}}{}_0(w,v)$$

$$+ 0.0945\,(\pm0.0064)\,^{\mathrm{MKe}}\boldsymbol{b}^{\mathbf{H}}{}_{5\mathrm{L}}(w,v)_E + 0.0386(\pm0.0058)\,^{\mathrm{VPe}}\boldsymbol{b}_{5\mathrm{L}}(w,v)_E$$

$$- 0.0250(\pm0.0026)\,^{\mathrm{VKe}}\boldsymbol{b}_{3\mathrm{L}}(w,v)_E \qquad \mathbf{(41)}$$

$N = 34$ $R^2 = 0.981$ $s = 0.110$ $q^2 = 0.949$ $s_{cv} = 0.162$ $F(7,26) = 197.61$

The structures of these 34 furylethylene derivatives are given in Table 7. The observed and calculated values of log *P* and log *k* are show in Tables 8 and 9, respectively.

**Tables 7-9 come about here (see end the document)**

These equations, obtained by using non-stochastic (stochastic) bilinear indices, explained 98.5% (98.7%) and 97.0% (98.1%) of the variance of log *k* and log *P,* respectively. These statistics are rather better than those previously obtained (see Table 10 for more details).[57,58]

**Table 10 comes about here (see end the document)**

The LOO cross-validation procedure was used in order to assess the predictive ability of the developed models. Using this approach, models **38**, **39**, **40** and **41** had a LOO $q^2$ of 0.964, 0.971, 0.971 and 0.910, respectively. These values of $q^2$ ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models.[73] Therefore, the equations obtained with the vertex- and edge-connectivity indices, with the topographic descriptors, and with the quantum-chemical indices showed smaller predictive abilities ($s_{cv}$ of 0.247, 0.176, and 0.370, respectively) than Eqs. **40** ($s_{cv} = 0.216$) and **41** ($s_{cv} = 0.162$), achieved with the total and local bond-based bilinear indices, respectively, for description of the Log $P$ values (see Table 10 for more details). Unfortunately, the authors[57] of the previous work did report the result of the LOO cross-validation experiment for log $k$. However, in Table 10 it can be easily observed that our obtained model, Eq. **38** (Eq. **39**) explains a greater percentage of the variance of the experimental Log $k$ values than the previously developed models do, showing decreases in the standard error of 70.93%, **69.77%,** *38.13%* and 32.14% (73.86%, **72.82%**, *44.38%* and 38.19%), with regard to the results previously achieved by Estrada and Molina,[57] by means of connectivity indices (both 2D and 3D as well as edge- and vertex-based), **total (global) spectral moment (sum of the trace of the bond matrix)**, *local (fragment) spectral moment* (partial sum of the trace of the bond matrix) and quantum chemical descriptors, respectively. As can be observed in the same Table, the result obtained by using other atom- and bond-based ***TOMOCOMD-CARDD*** MDs were also included. The QSPR model derived with bond-level bilinear indices showed better-to-similar results than the ones obtained by some of the present authors en early publications.[33,34,76]

On the other hand, in the different form as to the Bp description, the stochastic indices are similar-to-better than non-stochastic ones to describe both Log $k$ and Log $P$

of furylethylene derivatives. This is an interesting finding, because this result showed that maybe for some properties/molecules non-stochastic descriptors are the best selection but not for other properties/activities. Here, it is important to outstand that all the obtained equations have a great representation of local indices, mainly of molecular fragments ($C_2$-$C_6$ and $C_6$-$C_7$) and group-type (heteroatoms). The distribution was significantly slight different for the two properties. The description of Log *P* takes the calculus on the heteroatoms like most relevant variables, while Log *k* sees the $C_2$-$C_6$ and and $C_6$-$C_7$ indices as most important ones. This is a logical result, if we take into account that these atoms are those involved in the exocyclic double bond of the 2-furylethylene, and that these are the "target" of the nucleophilic attack by the thiol (mercapto) group. Nevertheless, the total MDs included in the achieved models also indicate that the best description of the properties will be obtained by using a combination the global and local features of every chemical included in the analysis. From this point of view, it is of great importance to have atom- or bond-level as well as total molecular indices in the molecular space, to do a better description than the one obtained by using separately the local and global sets of MDs. The indices that appear more frequently in to the final equation are the short-to-medium range ($k$ = 2-5) ones. This showed that the interaction of electron (bonds) of every atomic nucleus in ±2-5 steps, as well as the molecular environment of exocyclic double bond (in ±2-5 steps), determined the chemical reactivity of fuerylethylenes. In fact, the pair electronegativity (K)-Atomic mass (M) can be selected like the best *combine* atomic-label of all the employed ones, having more representation in the selected/included indices in the final equations. Polarizability (P)-M was the second more relevant vertex-weight combination. Nonetheless, the whole weighting-schedule, included in every model, showed that the used adequate combinations of chemical-labels are rather important, in order to predict properties and activities of different natures.

**Antibacterial Activity.** Finally, LDA will be used here to obtain a classification model of 2-furylethylene compounds according to their antibacterial activity. The two best models obtained classification is given below, together with the statistical parameters of LDA:

$$\begin{aligned}\boldsymbol{Class} = {} & -19.757 - 7.02\text{x}10^{-5}\,{}^{\mathrm{VK}}\boldsymbol{b}^{\mathbf{H}}_{7\mathrm{L}}(w,v)_{E} + 0.0482\,{}^{\mathrm{VK}}\boldsymbol{b}_{1\mathrm{L}}(w,v)_{E} \\ & + 1.16\text{x}10^{-5}\,{}^{\mathrm{VK}}\boldsymbol{b}^{\mathbf{H}}_{10\mathrm{L}}(w,v)_{\mathrm{C6\text{-}C7}}\end{aligned} \qquad (\mathbf{42})$$

$N = 34$ $\lambda = 0.260$ $D^2 = 11.05$ $F(3,30) = 28.42$ $p<0.0001$

$$\boldsymbol{Class} = -20.00 - 0.0236\,{}^{\mathrm{MVe}}\boldsymbol{b}^{\mathbf{H}}_{10}(w,v) + 5.6203\,{}^{\mathrm{MPe}}\boldsymbol{b}^{\mathbf{H}}_{14\mathrm{L}}(w,v)_{\mathrm{C6\text{-}C7}} + 0.1827\,{}^{\mathrm{VPe}}\boldsymbol{b}_{1\mathrm{L}}(w,v)_{E} \quad (\mathbf{43})$$

$N = 34$ $\lambda = 0.268$ $D^2 = 10.59$ $F(3,30) = 27.25$ $p<0.0001$

where $\lambda$ is Wilks' statistic, $D^2$ is the square Mahalanobis distance, and F is the Fisher ratio. The statistical analysis showed that it exists appropriate discriminatory power to differentiate between the two respective groups. The calculation of percentages of good classification in the data set and external prediction set permitted us to carry out the assessment of the models.

Model **42** classified correctly 97.10% of the compounds in the training data set, misclassifying only 1 out of 34 compounds. The percentage of false actives in this data set was only 2.94%, i.e., 1 inactive compound was classified as active from 34 cases. Conversely, no compound from the group of actives was misclassified as inactive one (0.00% of misclassification). In contrast, Model **43** depicted 100.0% and 90.0% of good classification in the active and inactive groups, with an overall accuracy of 94.12 % in the training data set (32/34). Both equations showed a 100% of prediction in the test set (see the bottom of Table 10 and 11).

**Table 10 comes about here (see end the document)**

The statistical analysis of the three models, previously obtained by using 2D and 3D connectivity and quantum chemical descriptors, showed quite similar results. In this case, the overall accuracy of the three models was 91.2%, 94.1%, and 88.2%, respectively (see Table 10 for more details).[58] The improvement in the statistical parameters of our models (Eqs. **42** and **43**) compared the one using 2D and 3D connectivity indices as well as quantum chemical descriptors is easily detected by the decrease in the Wilks' $\lambda$ parameter and an increase in the Mahalanobis square distance. Finally, the present result compared quite favorably us to the other atom-based ***TOMOCOMD-CARDD*** descriptors (Table 10).

## Final Conclusions

The total and local (bond and bond-type) bilinear 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 MDs useful to be employed in QSPR/QSAR studies, similarity/diversity analysis and drug-design protocols. The functional definitions of these new set of molecular parameters are based on well-known and accepted algorithms and formulas in mathematics. Namely, these novel bond-based MDs are based on a bilinear map similar to those defined in linear algebra. The bond- and 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. The correlations found by these new sets of bond-level chemical descriptors for the examination of physicochemical properties of octane isomers (**FIRST CASE**), Bp of alkyl-alcohols (**SECOND CASE**), as well as the specific rate constant (log *k*), partition coefficient (log *P*), and the antibacterial activity of 34 derivatives of 2-furylethylenes (**THIRD CASE**) can be considered as statistically significant and permit a clear interpretation of the studied properties in terms of the structural features of molecules. The comparison with other approaches reveals good performance of the method proposed. Therefore, it is clearly demonstrated that this suitability is higher than that shown by other 2D/3D well-known sets of MDs. 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. In fact, several promissory results have been achieved with the use of these novel total and atom-level MDs in *computational* drug discovery of new trichomonacidals.[35]

## ***ACKNOWLEDGMENTS***

One of the authors (M-P. Y.) thanks the program '*Estades Temporals per a Investigadors Convidats*' for a fellowship to work at Valencia University (2008). CAMD-BIR Unit thanks are given to the Research Project called: "Strengthening postgraduate education and research in Pharmaceutical Sciences". This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium. Finally, but not least important, M-P. Y thanks also support from Spanish MEC (Project Reference: SAF2006-04698) as well as Dr. Klein for send me their papers about graph theory.

## References

1. Galvez, J. J. Chem. Inf. Comput. Sci. 2003, 43, 1231-1239.
2. Katritzky, A. R.; Gordeeva, E. V. J Chem Inf Comput Sci 1993, 33, 835-857.
3. Balaban, A. T. SAR QSAR Environ. Res. 1998, 8, 1-21.
4. Devillers, J. Curr. Opin. Drug Discovery Dev. 2000, 3.
5. Estrada, E.; Uriarte, E. Curr Med Chem 2001, 8, 1573-1588.
6. González-Díaz, H.; Santana, L. M.; Ubeira, F.; Uriarte, E. Proteomics 2008, 8, 0000–0000.
7. Balaban, A. T. From Chemical Graphs to Three-Dimensional Geometry; Plenum Press: New York, 1997.
8. Rouvray, D. H. Chemical Applications of Graph Theory; Academic Press: London, 1976.
9. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH: Germany, 2000.
10. Estrada, E. Chem Phys. Lett. 2001, 336, 248-252.
11. Wiener, H. J. Am. Chem. Soc. 1947, 69, 17.
12. Estrada, E.; Rodríguez, L.; Gutierrez, A. Comm. Math. Chem. (MATCH) 1997, 35, 145.
13. Estrada, E.; Rodríguez, L. Comm. Math. Chem. (MATCH) 1997, 35, 157.
14. Randić, M. J. Am. Chem. Soc. 1975, 97, 6609.
15. Devillers, J. B., A. T, Ed. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach Amsterdam, The Netherlands, 1999.
16. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, 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. J. Mol. Mod. 2003, 9, 395-407.
17. Marrero-Ponce, Y. Molecules 2003, 8, 687-726.
18. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. J. Mol. Mod. 2006, 12, 255–271.
19. Marrero-Ponce, Y. Bioorg. Med. Chem. 2004, 12, 6351-6369.
20. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Castro, E. A.; Torrens, F.; Rotondo, R. J. Math. Chem. 2008, DOI 10.1007/s10910-008-9386-3.
21. Marrero-Ponce, Y.; Torrens, F.; García-Domenech, R.; Ortega-Broche, S. E.; Romero Zaldivar, V. J. Math. Chem. 2008, DOI 10.1007/s10910-10008-19389-10910.
22. Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. J. Mol. Struct. (Theochem) 2005, 717, 67-79.
23. Marrero-Ponce, Y. J Chem Inf Comput Sci 2004, 44, 2010-2026.
24. Marrero-Ponce, Y.; Romero, V.; Central University of Las Villas: Santa Clara, Villa Clara, 2002.
25. Marrero-Ponce, Y.; Meneses-Marcel, A.; Catillo-Garit, J. A.; Machado-Tugores, Y.; Escario, J. A.; Gómez-Barrio, A.; Montero Pereira, D.; Nogal-Ruiz, J. J.; Arán, V. J.; Martínez-Fernández, A. R.; Torrens, F.; Rotondo, R. Bioorg. Med. Chem. 2006, 14, 6502–6524.
26. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. J. Mol. Graph. Modell. 2007, 26, 32-47.
27. Marrero-Ponce, Y.; Khan, M. T. H.; Casañola-Martín, G. M.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F.; Rotondo, R. ChemMedChem 2007, 2, 449 – 478.
28. Jacobson, N. Basic Algebra I.; Ed. W.H. Freeman and Company: New York 1985.

29. Riley, K. F.; Hobson, M. P.; Vence, S. J. Mathematical Methods for Physics and Engineering; Ed. Cambridge University Press, 1998.
30. de Burgos-Román, J. Álgebra y Geometría Cartesíana; Ed. McGraw-Hill Interamericana de España, 2000.
31. Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. J Chem Inf Model 2005, 45, 1082-1100.
32. Marrero-Ponce, Y.; Díaz, H. G.; Romero, V.; Torrens, F.; Castro, E. A. Bioorg. Med. Chem. 2004, 12, 5331-5342.
33. Marrero-Ponce, Y. Bioorg. Med. Chem. 2004, 12, 6351-6369.
34. Marrero-Ponce, Y.; Torrens, F.; Alvarado, Y. J.; Rotondo, R. J. Comput. Aided Mol. Des. 2006, 20, 685–701.
35. Rivera-Borroto, O. M.; Marrero-Ponce, Y.; Meneses-Marcel, A.; Escario, J. A.; Gómez-Barrio, A.; A., A. V.; Martins-Alho, M. A.; Montero Pereira, D.; Nogal, J. J.; Torrens, F.; Ibarra-Velarde, F.; Vera Montenegro, V.; Huesca-Guille, A.; Rivera, N.; Vogel, V. QSAR Comb. Sci. 2008, DOI: 10.1002/qsar.200610165.
36. Estrada, E. J Chem Inf Comput Sci 1995, 35, 31-33.
37. Estrada, E.; Ramirez, A. J Chem Inf Comput Sci 1996, 36, 837-843.
38. Estrada, E.; Guevara, N.; Gutman, I. J Chem Inf Comput Sci 1998, 38, 428-431.
39. Estrada, E. J Chem Inf Comput Sci 1996, 36, 844-849.
40. Estrada, E. J Chem Inf Comput Sci 1999, 39, 1042-1048.
41. Kier, L. B.; Hall, L. H. Pharm Res 1990, 7, 801-807.
42. Kier, L. B.; Hall, L. H. Molecular Structure Description. The Electrotopological State; Academic Press: New York, 1999.
43. Daudel, R.; Lefebre, R.; Moser, C. Quantum Chemistry: Methods and Applications; Wiley: New York, 1984.
44. Rouvray, D. H.; Balaban, A. T. In Chemical Applications of Graph Theory; Eds.: Academic Press: London, 1976.
45. Trinajstić, N. Chemical Graph Theory; CRC Press: Boca Raton, FL., 1983.
46. Edwards, C. H.; Penney, D. E. Elementary Linear Algebra; Prentice-Hall, Englewood Cliffs: New Jersey, USA 1988.
47. Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. J. Mol. Struct. (Theochem) 2005, 717, 67-79.
48. Gonzales-Diaz, H.; Gia, O.; Uriarte, E.; Hernadez, 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. J Mol Model (Online) 2003, 9, 395-407.
49. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. J Med Chem 2000, 43, 1975-1985.
50. Potapov, V. M. Stereochemistry; Mir: Moscow, 1978.
51. Pauling, L. The Nature of Chemical Bond; Cornell University Press Ithaca (New York), 1939.
52. Walker, P. D.; Mezey, P. G. J. Am. Chem. Soc. 1993, 115, 12423.
53. Todeschini, R.; Gramatica, P. Perspect. Drug Dis. Des. 1998, 9-11, 355–380.
54. Klein, D. J. Internet Electronics J. Mol. Des. 2003, 2, 814-834.
55. International Academy of Mathematical Chemistry, http://michem.disat.unimib.it/chm/Bulletin/IAMC/index.html.
56. Consonni, V.; Todeschini, R.; Pavan, M. J Chem Inf Comput Sci 2002, 42, 682-692.
57. Estrada, E.; Molina, E. J Mol Graph Model 2001, 20, 54-64.
58. Estrada, E.; Molina, E. J Chem Inf Comput Sci 2001, 41, 791-797.
59. Randić, M.; Trinajstić, N. J. Mol. Struct. (Theochem) 1993, 284 209.

60. Randić, M.; Trinajstić, N. J. Mol. Struct. (Theochem) 1993, 300, 551-572.
61. Estrada, E.; Rodríguez, L. J. Chem. Inf. Comput. Sci. 1999, 39, 1037-1041.
62. Randić, M. J Mol Struct (Theochem) 1991, 233, 45-59.
63. Randić, M. Croat. Chim. Acta 1993, 66, 289-312.
64. Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. J. Chem. Inf. Comput. Sci. 1994, 34, 361-367.
65. Diudea, M. V. J. Chem. Inf. Comput. Sci. 1996, 36, 535-540.
66. Diudea, M. V.; Minailiuc, O. M.; Katona, G. Rev. Roum. Chim. 1997, 42, 239-249.
67. Randić, M. J. Math. Chem. 1991, 7, 155-168.
68. Needham, D. E.; Wei, I.-C.; Seybold, P. G. J. Am. Chem. Soc. 1988, 110, 4186-4194.
69. Balaz, S.; Sturdik, E.; Rosenberg, M.; Augustin, J.; Skara, B. J Theor Biol 1988, 131, 115-134.
70. Blondeau, J. M.; Castañedo, N.; Gonzalez, O.; Medina, R.; Silveira, E. Antimicrob. Agents Chemother. 1999, 11, 1663-1669.
71. Marrero-Ponce, Y.; Romero, V.; Central University of Las Villas: Santa Clara, Villa Clara, 2002.
72. ; StatSoft Inc, 2001.
73. Golbraikh, A.; Tropsha, A. J Mol Graph Model 2002, 20, 269-276.
74. Wold, S.; Erikson, L. In Chemometric Methods in Molecular Design; van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995, p 309-318.
75. Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. J Med Chem 2004, 47, 2356-2364.
76. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. Molecules 2004, 9, 1100-1123.
77. Sturdik, E.; Drobnica, L.; Balaz, S. Coll. Czch. Chem. Comm. 1985, 50, 470-480.

# ANNEXES

## (Tables, Figures and Schemes to be Inserted in the Main Text)

**Table 1.** Values of the Atom Weights Used for Linear Indices Calculation.

| ID | Atomic Mass | VdW[a] Volume ($\text{Å}^3$) | Mulliken Electronegativity | Polarizability ($\text{Å}^3$) |
|---|---|---|---|---|
| H | 1.01 | 6.709 | 2.592 | 0.667 |
| B | 10.81 | 17.875 | 2.275 | 3.030 |
| C | 12.01 | 22.449 | 2.746 | 1.760 |
| N | 14.01 | 15.599 | 3.194 | 1.100 |
| O | 16.00 | 11.494 | 3.654 | 0.802 |
| F | 19.00 | 9.203 | 4.000 | 0.557 |
| Al | 26.98 | 36.511 | 1.714 | 6.800 |
| Si | 28.09 | 31.976 | 2.138 | 5.380 |
| P | 30.97 | 26.522 | 2.515 | 3.630 |
| S | 32.07 | 24.429 | 2.957 | 2.900 |
| Cl | 35.45 | 23.228 | 3.475 | 2.180 |
| Fe | 55.85 | 41.052 | 2.000 | 8.400 |
| Co | 58.93 | 35.041 | 2.000 | 7.500 |
| Ni | 58.69 | 17.157 | 2.000 | 6.800 |
| Cu | 63.55 | 11.494 | 2.033 | 6.100 |
| Zn | 65.39 | 38.351 | 2.223 | 7.100 |
| Br | 79.90 | 31.059 | 3.219 | 3.050 |
| Sn | 118.71 | 45.830 | 2.298 | 7.700 |
| I | 126.90 | 38.792 | 2.778 | 5.350 |

[a] VdW: van der Waals

**Table 2.** Statistical Information for Best Multiple Regression Models of Selected Physicochemical Properties of Octane Isomers.

| **Property** | **Method** | **size** | $Q^2_{LOO}$ | $R^2$ | $s$ | **F** | **Equation or Model Descriptors** | **Ref.** |
|---|---|---|---|---|---|---|---|---|
| | 2D Bond- and Atom-based *bilinear indices* | | | | | | | |
| Boiling Point (BP) | NonStochastic Bond-based Bilinear Indices | 3 | 97.50 | 98.49 | 0.828 | 304.73 | $\mathbf{BP} = 194.62 - 0.319\,{}^{MP}b_{0l}\,(w,v)_{C\text{-}CH3} + 4.8\text{x}10^{-3}\,{}^{MK}b_{4l}^{\;H}\,(w,v)_{C\text{-}CH3} - 9.0\text{x}10^{-2}\,{}^{PV}b_{1}\,(w,v)$ | **(24)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 96.15 | 98.21 | 0.901 | 256.21 | $\mathbf{Bp} = 120.53 + 0.010\,{}^{MV}b_{7}\,(w,v) - 103.20\,{}^{MP}b_{5}^{\;H}\,(w,v) + 438.08\,{}^{PK}b_{4}^{\;H}\,(w,v)$ | **(25)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 96.1 | 97.7 | 1.021 | 198.64 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 92.7 | 96.23 | 1.314 | 118.18 | | |
| | Best Models From Literature by Using 2D and 3D MDs | | | | | | | |
| | getaway + whim + top. | 3 | 98.12 | 98.78 | 0.744 | | ${}^{2}\chi\ {}^{2}\overline{\chi}\ \ HATS_{6}(\mathrm{p})$ | 56 |
| | getaway | 3 | 97.10 | 98.32 | 0.897 | | $HATS_{2}(\mathrm{v})\ \mathrm{R}_{4}(\mathrm{u})\ \mathrm{R}_{6}(\mathrm{v})$ | |
| | getaway + whim + top. | 2 | 96.62 | 97.58 | 1.013 | | ${}^{2}\chi\ \ HATS_{6}(\mathrm{p})$ | |
| | topological | 3 | | 95.84 | 1.394 | | $\mathrm{S}^{3}\mathrm{W}\ \mathrm{S}^{4}\mathrm{W}\ \mathrm{SJ}$ | |
| | topological | 2 | | 94.78 | 1.508 | | $\mathrm{S}^{3}\mathrm{W}\ \mathrm{S}^{4}\mathrm{W}$ | |
| | getaway | 2 | 84.86 | 89.62 | 2.098 | | $HATS_{2}(\mathrm{m})\ \mathrm{R}^{+}{}_{4}(\mathrm{u})$ | |
| | topological | 2 | | 81.36 | 2.810 | | $\mathrm{WW}\ \mathrm{x}_{1}$ | |
| | topological | 1 | | 78.85 | 2.90 | | Z | |
| | getaway + whim + top. | 1 | 66.47 | 74.64 | 3.175 | | $HATS_{2}(\mathrm{m})$ | |
| | topological | 1 | | 67.77 | 3.630 | | ${}^{2}\chi\ \mathrm{W}$ | |
| | 2D Bond- and Atom-based *bilinear indices* | | | | | | | |
| Motor Octane Number (MON) | NonStochastic Bond-based Bilinear Indices | 3 | 99.03 | 99.43 | 3.224 | 756.64 | $\mathbf{MON} = -514.96 - 6.88\text{x}10^{-9}\,{}^{MV}b_{15}^{\;H}\,(w,v) + 1.80\text{x}10^{-6}\,{}^{PK}b_{14}^{\;H}\,(w,v) - 5.10\text{x}10^{-2}\,{}^{VK}b_{1l}^{\;H}\,(w,v)_{C\text{-}CH3}$ | **(26)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 99.01 | 99.40 | 3.251 | 721.86 | $\mathbf{MON} = -\ 94.87 - 5.846\,{}^{MV}b_{3l}^{\;H}\,(w,v)_{C\text{-}CH3} - 5.121\,{}^{MV}b_{4l}^{\;H}\,(w,v)_{C\text{-}CH3} + 670.561\,{}^{PK}b_{10l}^{\;H}\,(w,v)_{C\text{-}CH3}$ | **(27)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 98.4 | 99.2 | 3.327 | 511.07 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 97.6 | 98.6 | 4.319 | 301.58 | | |
| | Best Models From Literature by Using 2D and 3D MDs | | | | | | | |
| | getaway + whim + top. | 3 | 98.58 | 99.23 | 2.439 | | ${}^{\mathrm{v}}I_{D}^{\;M}\ \mathrm{Ts}\ HATS_{1}(\mathrm{m})$ | 56 |
| | getaway | 3 | 97.42 | 98.62 | 3.259 | | $HATS_{4}(\mathrm{u})\ HATS_{7}(\mathrm{v})\ \mathrm{R}_{7}(\mathrm{p})$ | |
| | topological | 3 | | 98.05 | 3.855 | | $\mathrm{S}\chi^{1}\mathrm{W}\ \chi^{7}\mathrm{W}\ \chi^{3}\mathrm{W}$ | |
| | getaway + whim + top. | 2 | 96.77 | 97.68 | 4.053 | | $\mathrm{Ts}\ H_{4}(\mathrm{e})$ | |
| | getaway | 2 | 91.28 | 95.78 | 5.466 | | $HATS_{7}(\mathrm{m})\ \mathrm{R}_{4}(\mathrm{u})$ | |
| | topological | 2 | | 95.64 | 5.533 | | $\mathrm{S}\chi^{1}\mathrm{W}\ \mathrm{S}\chi^{3}\mathrm{W}$ | |
| | topological | 1 | | 95.22 | 5.589 | | $\mathrm{X}^{7}\mathrm{W}$ | |
| | getaway + whim + top. | 1 | 90.83 | 92.40 | 7.069 | | Ts | |
| | topological | 1 | | 91.97 | 7.270 | | $\mathrm{I}_{\mathrm{wD}}$ | |
| | getaway | 1 | 85.64 | 88.98 | 8.515 | | REIG | |

**Table 2**. *Cont.*

| Property | Method | size | $Q^2_{LOO}$ | $R^2$ | s | F | Equation or Model Descriptors | Ref. |
|---|---|---|---|---|---|---|---|---|
| 2D Bond- and Atom-based *bilinear indices* | | | | | | | | |
| Heat of Vaporization (HV) | NonStochastic Bond-based Bilinear Indices | 3 | 98.70 | 99.32 | 0.043 | 685.46 | $\mathbf{HV} = 13.88 - 8.0\mathrm{x}10^{-4}\,{}^{MV}b_{1}\,(w,v) + 2.33\mathrm{x}10^{-2}\,{}^{MV}b_{1l}^{H}\,(w,v)_{C\text{-}CH3} - 1.03\mathrm{x}10^{-3}\,{}^{MV}b_{0l}\,(w,v)_{C\text{-}CH3}$ | **(28)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 98.50 | 99.07 | 0.046 | 497.64 | $\mathbf{HV} = 12.08 - 0.031\,{}^{MV}b_{0l}^{H}\,(w,v)_{C\text{-}CH3} - 0.104\,{}^{MK}b_{1}^{H}\,(w,v) + 1.299\,{}^{PK}b_{2l}^{H}\,(w,v)_{C\text{-}CH3}$ | **(29)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 98.4 | 98.9 | 0.233 | 294.01 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 97.0 | 98.4 | 0.279 | 249.01 | | |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | | | |
| | getaway + whim + top. | 3 | 97.57 | 98.42 | 0.281 | | ${}^{0}\overline{\chi}\ {}^{3}\kappa\ \mathrm{R}^{+}_{6}(\mathrm{u})$ | 56 |
| | getaway | 3 | 95.46 | 97.18 | 0.375 | | $HATS_{6}(\mathrm{u})\ \mathrm{R}_{4}(\mathrm{u})\ \mathrm{R}^{+}_{1}(\mathrm{m})$ | |
| | getaway + whim + top. | 2 | 95.18 | 96.53 | 0.402 | | ${}^{2}\chi\ \mathrm{R}^{+}_{6}(\mathrm{u})$ | |
| | topological | 3 | | 95.65 | 0.459 | | $\chi^{1}\mathrm{W}\ \chi^{2}\mathrm{W}\ \chi^{3}\mathrm{W}$ | |
| | getaway | 2 | 93.15 | 94.87 | 0.488 | | $HATS_{4}(\mathrm{u})\ \mathrm{R}_{6}(\mathrm{e})$ | |
| | topological | 2 | | 92.62 | 0.577 | | ${}^{4}\mathrm{W}\ {}^{5}\mathrm{W}$ | |
| | topological | 1 | | 91.78 | 0.429 | | Z | |
| | getaway + whim + top. | 1 | 80.80 | 88.61 | 0.705 | | ${}^{2}\chi$ | |
| | getaway | 1 | 79.74 | 85.70 | 0.790 | | $\mathrm{R}_{2}(\mathrm{m})$ | |
| | topological | 2 | | 84.27 | 0.820 | | $\mathrm{WW}\ \mathrm{x}_{1}$ | |
| 2D Bond- and Atom-based *bilinear indices* | | | | | | | | |
| Molar Volume (MV) | NonStochastic Bond-based Bilinear Indices | 3 | 98.60 | 99.24 | 0.306 | 565.03 | $\mathbf{MV} = 214.22 + 2.2\mathrm{x}10^{-3}\,{}^{MV}b_{0l}\,(w,v)_{C\text{-}CH3} - 0.198\,{}^{MP}b_{2}^{H}\,(w,v) + 0.257\,{}^{MK}b_{1}^{H}\,(w,v)$ | **(30)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 97.83 | 99.01 | 0.282 | 434.35 | $\mathbf{MV} = 242.186 - 2.773\,{}^{PV}b_{2}^{H}\,(w,v) + 0.018\,{}^{PV}b_{0l}\,(w,v)_{C\text{-}CH3} + 0.231\,{}^{VK}b_{1}^{H}\,(w,v)$ | **(31)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 98.2 | 99.1 | 0.273 | 461.48 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 98.7 | 99.1 | 0.264 | 491.33 | | |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | | | |
| | getaway + whim + top. | 3 | 75.96 | 92.01 | 1.825 | | $\mathrm{Ks}\ \mathrm{R}^{+}_{6}(\mathrm{u})\ RT^{+}(\mathrm{m})$ | 56 |
| | getaway | 3 | 69.27 | 90.33 | 2.008 | | $HATS_{6}(\mathrm{p})\ RT^{+}(\mathrm{m})\ R_{1}(\mathrm{v})$ | |
| | topological | 3 | | 88.29 | 2.210 | | ${}^{5}\mathrm{W}\ {}^{6}\mathrm{W}\ {}^{7}\mathrm{W}$ | |
| | getaway + whim + top. | 2 | 54.49 | 84.96 | 2.419 | | ${}^{\mathrm{v}}I_{D}^{M}\ \mathrm{R}^{+}_{6}(\mathrm{u})$ | |
| | getaway | 2 | 45.49 | 81.79 | 2.662 | | $\mathrm{R}^{+}_{6}(\mathrm{u})\ R_{4}(\mathrm{v})$ | |
| | getaway + whim + top. | 1 | 32.66 | 67.61 | 3.437 | | $\mathrm{R}_{6}(\mathrm{v})$ | |
| | topological | 2 | | 62.76 | 3.807 | | ${}^{3}\mathrm{W}\ {}^{4}\mathrm{W}$ | |
| | topological | 1 | | 60.85 | 3.780 | | ${}^{7}\mathrm{W}$ | |

**Table 2.** *Cont.*

| **Property** | **Method** | **size** | **$Q^2_{LOO}$** | **$R^2$** | ***s*** | **F** | **Equation or Model Descriptors** | **Ref.** |
|---|---|---|---|---|---|---|---|---|
| 2D Bond- and Atom-based *bilinear indices* | | | | | | | | |
| Entropy (S) | NonStochastic Bond-based Bilinear Indices | 3 | 90.60 | 95.22 | 1.387 | 93.020 | **S** $= 112.54 + 5.50\text{x}10^{-6}\,{}^{MV}b_{8l}\,(w,v)_{C\text{-}CH3} - 4.26\text{x}10^{-4}\,{}^{MK}b_{6l}\,(w,v)_{C\text{-}CH3} - 4.04\text{x}10^{-8}\,{}^{PK}b_{14}\,(w,v)$ | **(32)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 89.50 | 94.53 | 1.465 | 80.74 | **S** $= 109.0 – 4.30\,{}^{MK}b_{4}{}^{H}\,(w,v) + 0.184\,{}^{VK}b_{0}(w,v) + 0.128\,{}^{VK}b_{1l}{}^{H}\,(w,v)_{C\text{-}CH3}$ | **(33)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 88.5 | 91.1 | 1.320 | 180.96 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 90.9 | 95.3 | 1.08 | 94.12 | | |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | | | |
| | getaway + whim + top. | 3 | 97.17 | 97.96 | 0.711 | | $^{v}I_{D,deg}$ TWC $R^{+}_{2}$(p) | 56 |
| | getaway + whim + top. | 2 | 96.42 | 97.14 | 0.814 | | $^{v}I_{D,deg}$ TWC | |
| | getaway | 3 | 93.45 | 95.84 | 1.016 | | $I_{SH}$ $HATS_{8}$(m) $R_{3}$(v) | |
| | getaway | 2 | 92.19 | 94.76 | 1.101 | | $I_{SH}$ $R_{3}$(v) | |
| | getaway + whim + top. | 1 | 89.86 | 92.51 | 1.274 | | $R_{3}$(v) | |
| | topological | 1 | | 91.10 | 1.400 | | $\chi^{[1/2]}$ | |
| | topological | 2 | | 81.72 | 2.060 | | $x_1$ $x_2$ | |
| 2D Bond- and Atom-based *bilinear indices* | | | | | | | | |
| Heat of Formation ($\Delta_f$H) | NonStochastic Bond-based Bilinear Indices | 3 | 93.40 | 96.05 | 0.321 | 112.63 | $\mathbf{\Delta_f H} = - 46.30 – 1.20\text{x}10^{-4}\,{}^{MV}b_{5l}{}^{H}\,(w,v)_{C\text{-}CH3} - 6.55\text{x}10^{-6}\,{}^{MK}b_{7}w,v) + 1.16\text{x}10^{-9}\,{}^{PV}b_{14l}{}^{H}\,(w,v)_{C\text{-}CH3}$ | **(34)** |
| | Stochastic Bond-based Bilinear Indices | 3 | 91.28 | 94.51 | 0.369 | 80.45 | $\mathbf{\Delta_f H} = - 123.85 + 7.151\,{}^{MP}b_{3}{}^{H}\,(w,v) - 1.364\,{}^{MK}b_{1}{}^{H}\,(w,v) - 0.053\,{}^{PV}b_{0l}\,(w,v)_{C\text{-}CH3}$ | **(35)** |
| | NonStochastic Atom-based Bilinear Indices | 3 | 97.3 | 98.4 | 0.053 | 292.84 | | 21 |
| | Stochastic Atom-based Bilinear indices | 3 | 96.6 | 97.61 | 0.062 | 632.63 | | |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | | | |
| | getaway + whim + top. | 3 | 95.06 | 96.60 | 0.254 | | $HATS_{5}$(m) $HATS_{7}$(m) $R_{4}$(e) | 56 |
| | getaway + whim + top. | 2 | 90.96 | 93.24 | 0.346 | | $^{2}\chi$ $HATS_{2}$(e) | |
| | getaway | 2 | 90.18 | 92.87 | 0.356 | | $HATS_{7}$(u) $R_{2}$(m) | |
| | getaway + whim + top. | 1 | 87.18 | 89.34 | 0.421 | | $HATS_{2}$(m) | |
| | topological | 3 | | 87.05 | 0.492 | | $\Omega_1\,\Omega_2\,\Omega_3$ | |
| | topological | 2 | | 86.86 | 0.478 | | $\Omega_1\,\Omega_2$ | |
| | topological | 1 | | 86.68 | 0.471 | | $1/^{2}\chi$ | |
| | topological | 2 | | 78.70 | 0.570 | | WW $x_1$ | |

**Table 3.** Results of Regression-Equations (Tree-Variables) for other Physical Properties of Octanes by Using (both Total and Local) Bond-based Bilinear Indices.

| **Property**[a] | Equations and their Statistical Parameter | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **size** | **$R^2$** | **S** | **$q^2$** | **$S_{cv}$** | **F** | **Equation** |
| ΔHc | non-stochastic | 3 | 96.09 | 0.274 | 93.87 | 0.312 | 114.82 | $\Delta Hc = 1226.40 - 9.578\text{x}10^{-5}\ {}^{MP}b_{7l}^{H}\ (w,v)_{C\text{-}CH3} - 1.924\text{x}10^{-6}\ {}^{MK}b_{8}\ (w,v) + 3.232\text{x}10^{-10}\ {}^{PV}b_{15l}^{H}\ (w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 94.72 | 0.319 | 91.73 | 0.362 | 83.77 | $\Delta Hc = 1148.98 + 7.134\ {}^{MP}b_{3}^{H}\ (w,v) - 1.367\ {}^{MK}b_{1}^{H}\ (w,v) - 0.053\ {}^{PV}b_{0l}\ (w,v)_{C\text{-}CH3}$ |
| TSA | non-stochastic | 3 | 96.37 | 2.986 | 94.32 | 3.420 | 124.03 | $TSA = 1488.29 + 1.258\text{x}10^{-4}\ {}^{MV}b_{4}\ (w,v - 5.301\ {}^{MP}b_{0}\ (w,v) - 0.144\ {}^{MP}b_{11}\ (w,v)$ |
| | stochastic | 3 | 95.64 | 3.700 | 92.50 | 3.886 | 102.51 | $TSA = 1475.95 + 10.055\ {}^{MP}b_{15}^{H}\ (w,v) - 5.956\ {}^{MP}b_{0}\ (w,v) - 0.150\ {}^{VK}b_{1l}\ (w,v)_{C\text{-}CH3}$ |
| AF | non-stochastic | 3 | 99.78 | 0.002 | 99.70 | 0.002 | 2150.0 | $AF = 0.470 - 1.548\text{x}10^{-8}\ {}^{PK}b_{10l}^{H}\ (w,v)_{C\text{-}CH3} + 7.13\text{x}10^{-4}\ {}^{PK}b_{2l}\ (w,v)_{C\text{-}CH3} - 4.2\text{x}10^{-5}\ {}^{VK}b_{2}\ (w,v)$ |
| | stochastic | 3 | 99.78 | 0.002 | 99.68 | 0.002 | 2194.5 | $AF = 1.494 - 3.699\text{x}10^{-}2\ {}^{MK}b_{3}^{H}\ (w,v) - 1.056\text{x}10^{-2}\ {}^{MV}b_{15}^{H}\ (w,v) + 5.62\text{x}10^{-4}\ {}^{MK}b_{14l}\ (w,v)_{C\text{-}CH3}$ |
| $R^2$ | non-stochastic | 3 | 95.28 | 0.043 | 99.91 | 0.048 | 94.40 | $R^2 = 15.146 - 0.055\ {}^{MP}b_{0}\ (w,v) + 1.97\text{x}10^{-3}\ {}^{MK}b_{4}^{H}\ (w,v) - 7.98\text{x}10^{-3}\ {}^{PV}b_{3}^{H}\ (w,v)$ |
| | stochastic | 3 | 98.28 | 0.017 | 99.98 | 0.020 | 649.55 | $R^2 = 2.490 - 0.002\ {}^{VK}b_{1l}^{H}\ (w,v)_{C\text{-}CH3} + 0.353\ {}^{VK}b_{3l}^{H}\ (w,v)_{C\text{-}CH3} - 0.355\ {}^{VK}b_{5l}^{H}\ (w,v)_{C\text{-}CH3}$ |
| MR | non-stochastic | 3 | 99.39 | 0.016 | 98.78 | 0.020 | 709.0 | $MR = 42.92 - 2.79\text{x}10^{-3}\ {}^{MP}b_{4}^{H}\ (w,v) + 4.6\text{x}10^{-4}\ {}^{MV}b_{5}^{H}\ (w,v) + 1.96\text{x}10^{-2}\ {}^{MK}b_{1}^{H}\ (w,v)$ |
| | stochastic | 3 | 99.37 | 0.015 | 98.68 | 0.021 | 692.85 | $MR = 43.24 - 3.84\ \text{x}10^{-3}\ {}^{MV}b_{1}^{H}\ (w,v) - 0.190\ {}^{MK}b_{2}^{H}\ (w,v) + 1.71\text{x}10^{-3}\ {}^{PV}b_{2}(w,v)$ |
| Log P | non-stochastic | 3 | 96.96 | 0.010 | 93.10 | 0.013 | 149.24 | $Log\ P = 3.843 + 5.03\text{x}10^{-5}\ {}^{VK}b_{3}\ (w,v) - 1.07\text{x}10^{-3}\ {}^{VK}b_{0l}\ (w,v)_{C\text{-}CH3} - 3.80\text{x}10^{-5}\ {}^{VK}b_{3l}\ (w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 99.85 | 0.002 | 99.69 | 0.003 | 3174.7 | $Log\ P = 3.378 - 2.645\text{x}10^{-3}\ {}^{MV}b_{0l}^{H}\ (w,v)_{C\text{-}CH3} + 9.548\text{x}10^{-6}\ {}^{MV}b_{15l}\ (w,v)_{C\text{-}CH3} + 3.273\text{x}10^{-2}\ {}^{PV}b_{1}^{H}\ (w,v)$ |
| [b]ΔHv | non-stochastic | 3 | 99.27 | 0.190 | 98.90 | 0.209 | 639.14 | $\Delta Hv = 121.63 + 9.7\text{x}10^{-3}\ {}^{MP}b_{3l}^{H}\ (w,v)_{C\text{-}CH3} + 0.024\ {}^{MK}b_{4}^{H}\ (w,v) - 0.106\ {}^{PV}b_{3}^{H}\ (w,v)$ |
| | stochastic | 3 | 98.44 | 0.279 | 97.10 | 0.345 | 294.80 | $\Delta Hv = 67.182 + 0.0116\ {}^{PV}b_{1}\ (w,v) - 19.848\ {}^{PK}b_{3}^{H}\ (w,v) + 1.337\ {}^{VK}b_{2}^{H}\ (w,v)$ |
| d | non-stochastic | 3 | 99.04 | 0.003 | 93.39 | 0.007 | 483.30 | $d = 0.7542 - 2.1\text{x}10^{-5}\ {}^{MV}b_{1}\ (w,v) + 1.83\text{x}10^{-4}\ {}^{MV}b_{0l}^{H}\ (w,v)_{C\text{-}CH3} + 3.495\ {}^{MP}b_{15l}\ (w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 88.83 | 0.011 | 23.98 | 0.024 | 37.14 | $d = 0.7813 + 2.023\ {}^{MV}b_{2l}^{H}\ (w,v)_{C\text{-}CH3} - 7.65\text{x}10^{-5}\ {}^{MV}b_{0l}\ (w,v)_{C\text{-}CH3} -$ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | $9.79\text{x}10^{-4}\ {}^{MK}b_{12l}\,(w,v)_{C\text{-}CH3}$ |
| CT | non-stochastic | 3 | 96.45 | 2.020 | 94.36 | 2.311 | 127.02 | $CT = -244.458 + 1.076\ {}^{MK}b_2^H\,(w,v) - 1.86\text{x}10^{-3}\ {}^{PK}b_6(w,v) - 0.075\ {}^{VK}b_{0l}\,(w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 89.62 | 3.457 | 67.70 | 5.529 | 40.29 | $CT = 25.818 + 1.369\ {}^{MV}b_2^H\,(w,v) - 0.405\ {}^{MV}b_{0l}^H\,(w,v)_{C\text{-}CH3} + 3.34\text{x}10^{-2}\ {}^{MV}b_{5l}\,(w,v)_{C\text{-}CH3}$ |
| CP | non-stochastic | 3 | 97.07 | 0.271 | 94.15 | 0.347 | 154.48 | $CP = -32.742 - 2.53\text{x}10^{-7}\ {}^{MV}b_8(w,v) + 0.1097\ {}^{MK}b_2^H\,(w,v) - 0.040\ {}^{VK}b_{0l}\,(w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 89.52 | 0.513 | 60.00 | 0.910 | 39.86 | $CP = 12.10 - 4.54\text{x}10^{-3}\ {}^{MV}b_{3l}\,(w,v)_{C\text{-}CH3} + 0.460\ {}^{PV}b_5^H\,(w,v) - 2.43\text{x}10^{-2}\ {}^{VK}b_{1l}\,(w,v)_{C\text{-}CH3}$ |
| ${}^{c}\Delta H_f$ | non-stochastic | 3 | 81.20 | 0.405 | 98.73 | 0.471 | 20.16 | $\Delta H_f = 11.423 - 6.71\text{x}10^{-2}\ {}^{PV}b_1^H\,(w,v) - 1.803\text{x}10^{-2}\ {}^{PV}b_{2l}^H\,(w,v)_{C\text{-}CH3} + 1.0\text{x}10^{-3}\ {}^{PV}b_{5l}^H\,(w,v)_{C\text{-}CH3}$ |
| | stochastic | 3 | 79.87 | 0.419 | 98.67 | 0.482 | 18.52 | $\Delta H_f = -157.35 - 13.598\ {}^{MP}b_{6l}^H\,(w,v)_{C\text{-}CH3} + 0.535\ {}^{7\,MK}b_0(w,v) + 58.206\ {}^{PK}b_{3l}^H\,(w,v)_{C\text{-}CH3}$ |

[a] ΔHc: combustion enthalpy; TSA: Total Surface Area ; AF: Pitzer's Acentric Factor; $R^2$: the mean radius squared; MR: Molar Refraction; Log P: octanol-water partition coefficient; [b] ΔHv: no standard heat vaporization; d: density; CT: critical temperature; CP: critical pressure; [c] $\Delta H_f$ : standard enthalpy formation.

**Table 4.** Experimental and Predicted Values of the Boiling Point of 28 Alkyl-Alcohols R-OH Used in This Study.

| **Alcohol-R** | **Obsd.**[a] | **Pred.**[b] | **Res$_{VC-LOO}$**[c] | **Pred.**[d] | **Res$_{VC-LOO}$**[e] |
|---|---|---|---|---|---|
| $(CH_3)_2CH$- | 82,3 | 82.4 | -0.15 | 82.6 | -0.47 |
| $CH_3CH_2CH_2$- | 97,2 | 98.2 | -1.31 | 98.0 | -1.09 |
| $CH_3(CH_2)_3$- | 117,7 | 117.5 | 0.20 | 117.5 | 0.25 |
| $CH_3CH(CH_3)CH_2$- | 107,8 | 106.0 | 2.00 | 107.0 | 0.88 |
| $CH_3CH_2C(CH_3)_2$- | 102,4 | 102.1 | 0.35 | 102.8 | -0.45 |
| $CH_3CH_2CH_2CH(CH_3)$- | 119,3 | 119.6 | -0.33 | 119.2 | 0.12 |
| $CH_3CH(CH_3)CH_2CH_2$- | 131,1 | 130.9 | 0.17 | 129.6 | 1.67 |
| $CH_3CH_2CH(CH_3)CH_2$- | 128,0 | 129.1 | -1.24 | 128.3 | -0.34 |
| $CH_3(CH_2)_4$- | 137,9 | 136.9 | 1.14 | 138.0 | -0.09 |
| $CH_3C(CH_3)_2CH(CH_3)$- | 120,4 | 119.1 | 1.86 | 120.3 | 0.11 |
| $CH_3(CH_2)_2C(CH_3)_2$- | 121,1 | 121.1 | 0.03 | 122.2 | -1.32 |
| $(CH_3CH_2)_2C(CH_3)$- | 122,4 | 123.3 | -1.10 | 122.2 | 0.29 |
| $CH_3CH_2C(CH_3)_2CH_2$- | 136,5 | 137.2 | -0.78 | 137.4 | -1.34 |
| $CH_3CH(CH_3)CH_2CH(CH_3)$- | 131,6 | 133.0 | -1.60 | 130.6 | 1.18 |
| $CH_3CH(CH_3)CH(CH_3CH_2)$- | 126,5 | 127.3 | -1.14 | 127.9 | -1.56 |
| $CH_3CH(CH_3)CH(CH_3)CH_2$- | 144,5 | 144.5 | 0.04 | 141.7 | 3.25 |
| $CH_3CH_2CH_2CH(CH_3)CH_2$- | 149,0 | 148.1 | 0.93 | 149.4 | -0.46 |
| $CH_3(CH_2)_5$- | 157,6 | 156.3 | 1.52 | 157.2 | 0.46 |
| $(CH_3CH(CH_3))_2CH$- | 138,7 | 137.0 | 2.78 | 137.9 | 0.96 |
| $CH_3CH(CH_3)CH_2CH(CH_3)CH_2$- | 159,0 | 161.5 | -2.81 | 162.1 | -3.61 |
| $(CH_3CH_2)_3C$- | 142,0 | 143.6 | -2.38 | 141.8 | 0.25 |
| $CH_3(CH_2)_6$- | 176,4 | 175.6 | 0.92 | 176.3 | 0.15 |
| $(CH_3CH_2CH_2)_2(CH_3)C$- | 161,0 | 161.3 | -0.34 | 161.4 | -0.44 |
| $(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C$- | 163,0 | 161.7 | 1.50 | 162.2 | 0.91 |
| $CH_3CH(CH_3)CH_2(CH_2)_4$- | 188,0 | 189.0 | -1.32 | 188.1 | -0.07 |
| $CH_3(CH_2)_7$- | 195,1 | 195.0 | 0.17 | 195.3 | -0.33 |
| $CH_3(CH_2)_5C(CH_3)_2$- | 178,0 | 177.7 | 0.40 | 177.7 | 0.39 |
| $(CH_3CH_2CH_2)_2(CH_3CH_2)C$- | 182,0 | 181.6 | 0.64 | 181.8 | 0.21 |

[a]Experimental values of boiling point (°C). [b,d]Predicted values using non-stochastic (Eq. **36**) and stochastic (Eq. **37**) bond-based bilinear indices, respectively. [c,e]Residual values of LOO cross-validation process using non-stochastic and stochastic bond-based bilinear indices, correspondingly [$Res_{CV-LOO}$ = Bp(Obsd.) - Bp($Pred._{CV-LOO}$)].

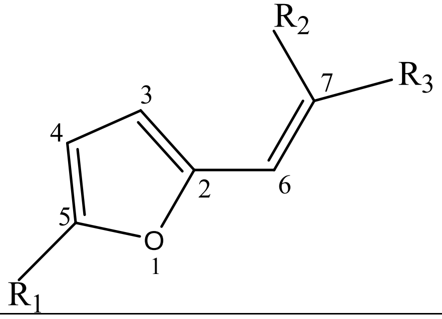
**Table 5.** Statistical Information for Best Multiple Regression (2, 3 and 4 parameters) Models Obtained by Using Non-Stochastic and Stochastic Bond-Level Bilinear Indices to describe the Bp of Alkyl-alcohols.

| zise | Molecular Descriptors | $R^2$ | s | $q^2$ | $s_{cv}$ | F |
|---|---|---|---|---|---|---|
| 2 | ${}^{\mathrm{MK}}\boldsymbol{b}_2{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MK}}\boldsymbol{b}_0\,(w,v)$ | 0.994 | 2.3 | 0.999 | 2.5 | 2024.0 |
| 3 | ${}^{\mathrm{MK}}\boldsymbol{b}_2{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MK}}\boldsymbol{b}_0\,(w,v)$, ${}^{\mathrm{MK}}\boldsymbol{b}_{13}\,(w,v)$ | 0.998 | 1.4 | 0.997 | 1.6 | 3399.0 |
| 4 | ${}^{\mathrm{MV}}\boldsymbol{b}_1(w,v)$, ${}^{\mathrm{MP}}\boldsymbol{b}_{2\mathrm{L}}{}^{\mathrm{H}}\,(w,v)_E$, ${}^{\mathrm{PK}}\boldsymbol{b}_0{}^{\mathrm{H}}\,(w,v)$, ${}^{\mathrm{PK}}\boldsymbol{b}_{12\mathrm{L}}(w,v)_E$ | 0.999 | 1.1 | 0.998 | 1.3 | 4174.7 |
| 2 | ${}^{\mathrm{MKe}}\boldsymbol{b}_4{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MKe}}\boldsymbol{b}_0\,(w,v)$ | 0.995 | 2.1 | 0.993 | 2.3 | 2308.0 |
| 3 | ${}^{\mathrm{MKe}}\boldsymbol{b}_3{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MKe}}\boldsymbol{b}_0\,(w,v)$, ${}^{\mathrm{MKe}}\boldsymbol{b}_1\,(w,v)$ | 0.997 | 1.6 | 0.996 | 1.7 | 2766.9 |
| 4 | ${}^{\mathrm{MVe}}\boldsymbol{b}_4{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MKe}}\boldsymbol{b}_3{}^{\mathrm{H}}(w,v)$, ${}^{\mathrm{MKe}}\boldsymbol{b}_1\,(w,v)$, ${}^{\mathrm{PVe}}\boldsymbol{b}_0\,(w,v)$ | 0.999 | 1.1 | 0.998 | 1.2 | 4404.8 |

**Table 6.** Statistical Parameters for the Models Describing the Bp of Alkyl-alcohols by Using Different MDs.

| **Molecular Descriptors** | **n** | $R^2$ | $q^2$ | **s** | $s_{CV}$ | **F** |
|---|---|---|---|---|---|---|
| Best Models From Literature | | | | | | |
| Local spectral moments[57] | 5 | 0.982 | * | 4.2 | * | 23.8 |
| E-State[42] | 3 | 0.926 | * | 5.8 | * | 204 |
| 2D Bond- and Atom-based ***TOMOCOMD-CARDD*** MDs | | | | | | |
| Bond-based non-Stocastic Bilinear Indices (Eq. **36**) | 4 | 0.999 | 0.998 | 1.1 | 1.3 | 4174.7 |
| Bond-based Stocastic Bilinear Indices (Eq. **37**) | 4 | 0.999 | 0.998 | 1.1 | 1.2 | 4404.8 |
| Atom-based Linear Indices[76] | 4 | 0.993 | 0.990 | 2.48 | 2.79 | 871.96 |
| Bond-based Quadratic Indices[34] | 3 | 0.998 | 0.997 | 1.46 | 1.60 | 3394.6 |

**Table 7.** Chemical structures and numbering of atoms in the 2-furylethylene compounds used in this study.



| no. | $R_1$ | $R_2$ | $R_3$ | no. | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|---|---|---|---|
| 1 | H | $NO_2$ | $COOCH_3$ | 18 | $NO_2$ | H | $CONHCH(CH_3)C_2H_5$ |
| 2 | $CH_3$ | $NO_2$ | $COOCH_3$ | 19 | $NO_2$ | H | $CONHC(CH_3)_3$ |
| 3 | Br | $NO_2$ | $COOCH_3$ | 20 | $NO_2$ | H | $CONHCH_2C(CH_3)_3$ |
| 4 | I | $NO_2$ | $COOCH_3$ | 21 | $NO_2$ | H | $COOCH_3$ |
| 5 | $COOCH_3$ | $NO_2$ | $COOCH_3$ | 22 | $NO_2$ | H | $COOC_2H_5$ |
| 6 | $NO_2$ | $NO_2$ | $COOCH_3$ | 23 | $NO_2$ | H | $COO(CH_2)_2CH_3$ |
| 7 | $NO_2$ | $COOC_2H_5$ | $COOC_2H_5$ | 24 | $NO_2$ | H | $COOCH(CH_3)_2$ |
| 8 | $NO_2$ | H | $NO_2$ | 25 | $NO_2$ | H | $COO(CH_2)_3CH_3$ |
| 9 | H | H | $NO_2$ | 26 | $NO_2$ | H | $COOCH_2CH(CH_3)_2$ |
| 10 | $NO_2$ | H | $CONH_2$ | 27 | $NO_2$ | H | $COOCH(CH_3)C_2H_5$ |
| 11 | $NO_2$ | H | $CONHCH_3$ | 28 | $NO_2$ | H | $COOC(CH_3)_3$ |
| 12 | $NO_2$ | H | $CON(CH_3)_2$ | 29 | $NO_2$ | H | $COO(CH_2)_4CH_3$ |
| 13 | $NO_2$ | H | $CONHC_2H_5$ | 30 | $NO_2$ | H | Br |
| 14 | $NO_2$ | H | $CONH(CH_2)_2CH_3$ | 31 | $NO_2$ | H | CN |
| 15 | $NO_2$ | H | $CONHCH(CH_3)_2$ | 32 | $NO_2$ | H | $OCH_3$ |
| 16 | $NO_2$ | H | $CONH(CH_2)_3CH_3$ | 33 | $NO_2$ | H | H |
| 17 | $NO_2$ | H | $CONHCH_2CH(CH_3)_2$ | 34 | $NO_2$ | CN | $COOCH_3$ |

Novel $R_1$,$R_2$-Substituted 2-Furylethylenes ($R_3 = NO_2$) used as ***external test set*** to assess the predictive power of the classification model for antibacterial activity

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | Br | Br | $NO_2$ | 6 | H | I | $NO_2$ |
| 2 | I | I | $NO_2$ | 7 | H | $CH_3$ | $NO_2$ |
| 3 | Br | H | $NO_2$ | 8 | Br | $CH_3$ | $NO_2$ |
| 4 | H | Br | $NO_2$ | 9 | I | $CH_3$ | $NO_2$ |
| 5 | I | H | $NO_2$ | | | | |

**Table 8**. Experimental and calculated values of the specific rate constant for the reaction of nucleophilic addition of thiols (log k) to the exocyclic double bond of the studied 2-furylethylenes.

| **no.** | **Obsd.**[a] | **Pred.**[b] | **$Res_{VC\text{-}LOO}$**[c] | **Pred.**[d] | **$Res_{VC\text{-}LOO}$**[e] |
|---|---|---|---|---|---|
| 1 | 6.591 | 6.853 | -0.262 | 6.474 | 0.117 |
| 2 | 6.518 | 6.085 | 0.433 | 6.612 | -0.094 |
| 3 | 6.914 | 7.110 | -0.196 | 6.829 | 0.085 |
| 4 | 6.982 | 6.817 | 0.165 | 6.966 | 0.016 |
| 5 | 7.176 | 7.087 | 0.089 | 7.135 | 0.041 |
| 6 | 7.602 | 7.874 | -0.272 | 7.807 | -0.205 |
| 7 | 5.255 | 5.352 | -0.097 | 5.311 | -0.056 |
| 8 | 6.763 | 6.545 | 0.218 | 6.821 | -0.058 |
| 9 | 5.623 | 5.535 | 0.088 | 5.483 | 0.140 |
| 10 | 3.813 | 3.933 | -0.120 | 4.137 | -0.324 |
| 11 | 3.840 | 3.947 | -0.107 | 3.771 | 0.069 |
| 12 | 3.874 | 3.489 | 0.385 | 3.731 | 0.143 |
| 13 | 3.825 | 3.686 | 0.139 | 3.512 | 0.313 |
| 14 | 3.623 | 3.677 | -0.054 | 3.623 | 0.000 |
| 15 | 3.751 | 3.718 | 0.033 | 3.693 | 0.058 |
| 16 | 3.784 | 3.725 | 0.059 | 3.628 | 0.156 |
| 17 | 3.697 | 3.668 | 0.029 | 3.711 | -0.014 |
| 18 | 3.705 | 3.709 | -0.004 | 3.763 | -0.058 |
| 19 | 3.697 | 3.845 | -0.148 | 3.895 | -0.198 |
| 20 | 3.650 | 3.660 | -0.010 | 3.777 | -0.127 |
| 21 | 4.000 | 4.153 | -0.153 | 4.191 | -0.191 |
| 22 | 3.920 | 3.828 | 0.092 | 3.592 | 0.328 |
| 23 | 3.790 | 3.659 | 0.131 | 3.612 | 0.178 |
| 24 | 3.763 | 3.748 | 0.015 | 3.686 | 0.077 |
| 25 | 3.623 | 3.653 | -0.030 | 3.587 | 0.036 |
| 26 | 3.650 | 3.485 | 0.165 | 3.648 | 0.002 |
| 27 | 3.592 | 3.576 | 0.016 | 3.726 | -0.134 |
| 28 | 3.584 | 3.713 | -0.129 | 3.862 | -0.278 |
| 29 | 3.590 | 3.648 | -0.058 | 3.582 | 0.008 |
| 30 | 2.987 | 2.909 | 0.078 | 3.046 | -0.059 |
| 31 | 3.273 | 3.350 | -0.077 | 3.355 | -0.082 |
| 32 | 2.140 | 2.599 | -0.459 | 2.371 | -0.231 |
| 33 | 3.553 | 3.485 | 0.068 | 3.496 | 0.057 |
| 34 | 5.557 | 5.581 | -0.024 | 5.272 | 0.285 |

[a]Experimental values of log k. [b,d]Predicted values using non-stochastic (Eq. **38**) and stochastic (Eq. **39**) bond-based bilinear indices, respectively. [c,e]Residual values of LOO cross-validation process using non-stochastic and stochastic bond-based bilinear indices, correspondingly [$Res_{CV\text{-}LOO}$ = log k (Obsd.) – log k ($Pred._{CV\text{-}LOO}$)].

**Table 9.** Experimental and calculated values of the partition coefficient *n*-octanol/water (log *P*) for the furylethylenes studied.

| no. | Obsd.[a] | Pred.[b] | Res$_{\mathrm{VC\text{-}LOO}}$[c] | Pred.[d] | Res$_{\mathrm{VC\text{-}LOO}}$[e] |
|---|---|---|---|---|---|
| 1 | 1.879 | 1.811 | 0.068 | 1.697 | 0.182 |
| 2 | 2.439 | 2.246 | 0.193 | 2.513 | -0.074 |
| 3 | 2.739 | 2.629 | 0.110 | 2.723 | 0.016 |
| 4 | 2.999 | 3.042 | -0.043 | 3.008 | -0.009 |
| 5 | 1.869 | 1.899 | -0.030 | 1.865 | 0.004 |
| 6 | 1.599 | 1.539 | 0.060 | 1.570 | 0.029 |
| 7 | 2.504 | 2.720 | -0.216 | 2.522 | -0.018 |
| 8 | 1.303 | 1.294 | 0.009 | 1.467 | -0.164 |
| 9 | 1.583 | 1.572 | 0.011 | 1.603 | -0.020 |
| 10 | 0.649 | 0.816 | -0.167 | 0.687 | -0.038 |
| 11 | 0.984 | 1.023 | -0.039 | 1.128 | -0.144 |
| 12 | 0.819 | 0.702 | 0.117 | 0.755 | 0.064 |
| 13 | 1.386 | 1.402 | -0.016 | 1.354 | 0.032 |
| 14 | 1.860 | 2.013 | -0.153 | 1.888 | -0.028 |
| 15 | 1.803 | 1.811 | -0.008 | 1.822 | -0.019 |
| 16 | 2.356 | 2.359 | -0.003 | 2.262 | 0.094 |
| 17 | 2.225 | 2.354 | -0.129 | 2.284 | -0.059 |
| 18 | 2.284 | 2.429 | -0.145 | 2.440 | -0.156 |
| 19 | 2.333 | 2.151 | 0.182 | 2.261 | 0.072 |
| 20 | 2.605 | 2.506 | 0.099 | 2.674 | -0.069 |
| 21 | 1.652 | 1.684 | -0.032 | 1.821 | -0.169 |
| 22 | 2.098 | 2.093 | 0.005 | 2.123 | -0.025 |
| 23 | 2.673 | 2.678 | -0.005 | 2.624 | 0.049 |
| 24 | 2.641 | 2.567 | 0.074 | 2.522 | 0.119 |
| 25 | 2.827 | 3.036 | -0.209 | 2.928 | -0.101 |
| 26 | 3.135 | 3.011 | 0.124 | 2.994 | 0.141 |
| 27 | 3.091 | 3.166 | -0.075 | 3.123 | -0.032 |
| 28 | 3.060 | 3.005 | 0.055 | 2.910 | 0.150 |
| 29 | 3.404 | 3.296 | 0.108 | 3.495 | -0.091 |
| 30 | 2.447 | 2.397 | 0.050 | 2.437 | 0.010 |
| 31 | 1.050 | 1.330 | -0.280 | 1.167 | -0.117 |
| 32 | 1.591 | 1.613 | -0.022 | 1.459 | 0.132 |
| 33 | 1.611 | 1.311 | 0.300 | 1.456 | 0.155 |
| 34 | 1.488 | 1.480 | 0.008 | 1.402 | 0.086 |

[a]Experimental values of log P. [b,d]Predicted values using non-stochastic (Eq. **40**) and stochastic (Eq. **41**) bond-based bilinear indices, respectively. [c,e]Residual values of LOO cross-validation process using non-stochastic and stochastic bond-based bilinear indices, correspondingly [Res$_{\mathrm{CV\text{-}LOO}}$ = log P (Obsd.) – log P (Pred.$_{\mathrm{CV\text{-}LOO}}$)].

**Table 10.** Statistical parameters of the QSPR/QSAR models obtained by using different MDs in order to model the properties of molecules included in the **CASE 3**.

| **Index**[a] | **n** | **$R^2$** | ***s*** | ***$q^2$*** | ***$s_{CV\text{-}LOO}$*** | **F** |
|---|---|---|---|---|---|---|
| **Reactivity (log *k*) of 34 2-Furylethylenes** | | | | | | |
| 2D Bond- and Atom-based ***TOMOCOMD-CARDD*** MDs | | | | | | |
| Bond-based NS BI(Eq.38) | 7 | 0.985 | 0.198 | 0.964 | 0.270 | 242.04 |
| Bond-based SS BI(Eq.39) | 7 | 0.987 | 0.178 | 0.971 | 0.240 | 293.74 |
| Atom-based NS QI[33] | 7 | 0.968 | 0.285 | 0.922 | 0.298 | 115.14 |
| Bond-based NS QI[34] | 7 | 0.967 | 0.292 | 0.940 | 0.345 | 108.79 |
| Bond-based SS QI[34] | 7 | 0.975 | 0.257 | 0.958 | 0.288 | 142.07 |
| Atom-based NS LI[76] | 6 | 0.973 | 0.26 | 0.948 | 0.33 | 161.22 |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | |
| Conn. Indices[57] | 7 | 0.821 | 0.681 | * | * | 17.1 |
| Global spectral moments[57] | 7 | 0.843 | 0.655 | * | * | 18.8 |
| Local spectral moments[57] | 7 | 0.964 | 0.320 | * | * | 70.4 |
| Quantum chemical descriptors[57] | 7 | 0.968 | 0.288 | * | * | 112.2 |
| **Partition Coefficient *n*-Octanol/Water (log *P*) of 34 2-Furylethylenes** | | | | | | |
| 2D Bond- and Atom-based ***TOMOCOMD-CARDD*** MDs | | | | | | |
| Bond-based NS BI(Eq.40) | 7 | 0.970 | 0.140 | 0.910 | 0.216 | 119.14 |
| Bond-based SS BI(Eq.41) | 7 | 0.981 | 0.110 | 0.949 | 0.162 | 197.61 |
| Atom-based NS QI[33] | 7 | 0.969 | 0.142 | 0.951 | 0.156 | 116.76 |
| Atom-based NS LI[76] | 7 | 0.968 | 0.143 | 0.938 | 0.176 | 113.38 |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | |
| Vertex and Edge Conn. Indices[58] | 7 | 0.939 | 0.199 | * | 0.247 | 56.9 |
| Topographic descriptors[58] | 7 | 0.964 | 0.155 | * | 0.176 | 84.6 |
| Quantum chemical descriptors[58] | used the Rogers and Cammarata approach | 0.875 | 0.319 | * | 0.370 | 45.5 |
| **Classification of 34 2-Furylethylene Derivatives as Antibacterial** | | | | | | |
| **index** | **n** | **λ** | **$D^2$** | **Accuracy (Training)** | **Accuracy (Test)** | **F** |
| 2D Bond- and Atom-based ***TOMOCOMD-CARDD*** MDs | | | | | | |
| Bond-based NS BI(Eq.42) | 3 | 0.26 | 11.05 | 97.1 | 100% | 28.42 |
| Bond-based SS BI(Eq.43) | 3 | 0.27 | 10.59 | 94.1 | 100% | 27.25 |
| Atom-based NS LI[76] | 3 | 0.30 | 9.44 | 94.12% | 100% | 22.9 |
| Atom-based NS QI[33] | 5 | 0.26 | 11.78 | 97.1% | 100% | 15.98 |
| Best Models From Literature by Using 2D and 3D MDs | | | | | | |
| Vertex and Edge Conn. Indices[58] | 5 | 0.43 | 5.7 | 91.2% | 100% | 7.7 |
| Topographic descriptors[58] | 5 | 0.38 | 6.7 | 94.1% | 100% | 9.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Quantum chemical descriptors[58] | 5 | 0.44 | 5.2 | 88.2% | 100% | 7.1 |

**Table 11.** Classification of 2-furylethylene derivatives as antibacterial according to the two best(3-fold) obtained models with our MDs.

| Compd. | Obsd. | non-stochastic (Eq.) bond-based bilinear indice | | Stochastic (Eq.) bond-based bilinear indice | |
|---|---|---|---|---|---|
| | | Class. | Prob. | Class. | Prob. |
| | | *Learning Group* | | | |
| 1 | + | + | 99.98 | + | 99.99 |
| 2 | + | + | 99.98 | + | 99.55 |
| 3 | + | + | 99.99 | + | 99.95 |
| 4 | + | + | 99.99 | + | 99.64 |
| 5 | + | + | 99.87 | + | 99.99 |
| 6 | + | + | 99.97 | + | 99.95 |
| 7 | + | + | 99.59 | + | 99.57 |
| 8 | + | + | 89.31 | + | 88.58 |
| 9 | + | + | 91.91 | + | 98.05 |
| 10 | + | + | 93.07 | + | 66.53 |
| 11 | + | + | 86.35 | + | 85.23 |
| 12 | + | + | 96.73 | + | 99.93 |
| 13 | + | + | 59.50 | + | 59.53 |
| 14 | - | - | 3.06 | - | 0.64 |
| 15 | - | - | 3.41 | - | 1.16 |
| 16 | - | - | 0.99 | - | 0.22 |
| 17 | - | - | 0.12 | - | 0.05 |
| 18 | - | - | 0.08 | - | 0.09 |
| 19 | - | - | 0.02 | - | 0.05 |
| 20 | - | - | 0.00 | - | 0.01 |
| 21 | - | + | **87.20** | + | **84.53** |
| 22 | - | - | 44.36 | - | 32.10 |
| 23 | - | - | 1.00 | - | 0.20 |
| 24 | - | - | 0.64 | - | 0.29 |
| 25 | - | - | 0.23 | - | 0.06 |
| 26 | - | - | 0.02 | - | 0.01 |
| 27 | - | - | 0.01 | - | 0.02 |
| 28 | - | - | 0.00 | - | 0.01 |
| 29 | - | - | 0.14 | - | 0.02 |
| 30 | - | - | 0.04 | - | 9.91 |
| 31 | - | - | 0.26 | - | 0.83 |
| 32 | - | - | 17.93 | + | **82.82** |
| 33 | - | - | 0.03 | - | 0.10 |
| 34 | + | + | 96.85 | + | 92.09 |
| | | *Prediction group* | | | |
| 1 | + | + | 80.08 | + | 98.33 |
| 2 | + | + | 77.39 | + | 86.38 |
| 3 | + | + | 97.69 | + | 96.62 |
| 4 | + | + | 64.12 | + | 99.96 |
| 5 | + | + | 98.12 | + | 89.52 |
| 6 | + | + | 56.14 | + | 99.99 |
| 7 | + | + | 100.00 | + | 98.41 |
| 8 | + | + | 100.00 | + | 96.38 |
| 9 | + | + | 100.00 | + | 83.51 |

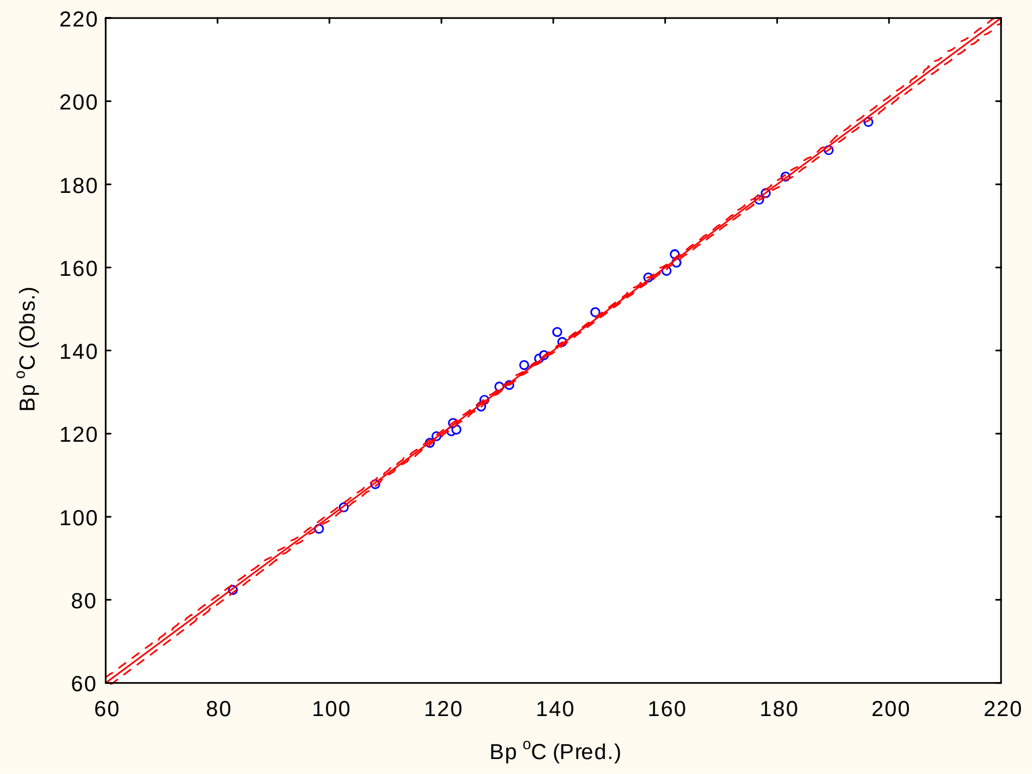


**Figure 1.** Linear correlations of observed versus calculated boiling point according to the model obtained from non-stochastic bond-level bilinear indices (Eq. **36**).

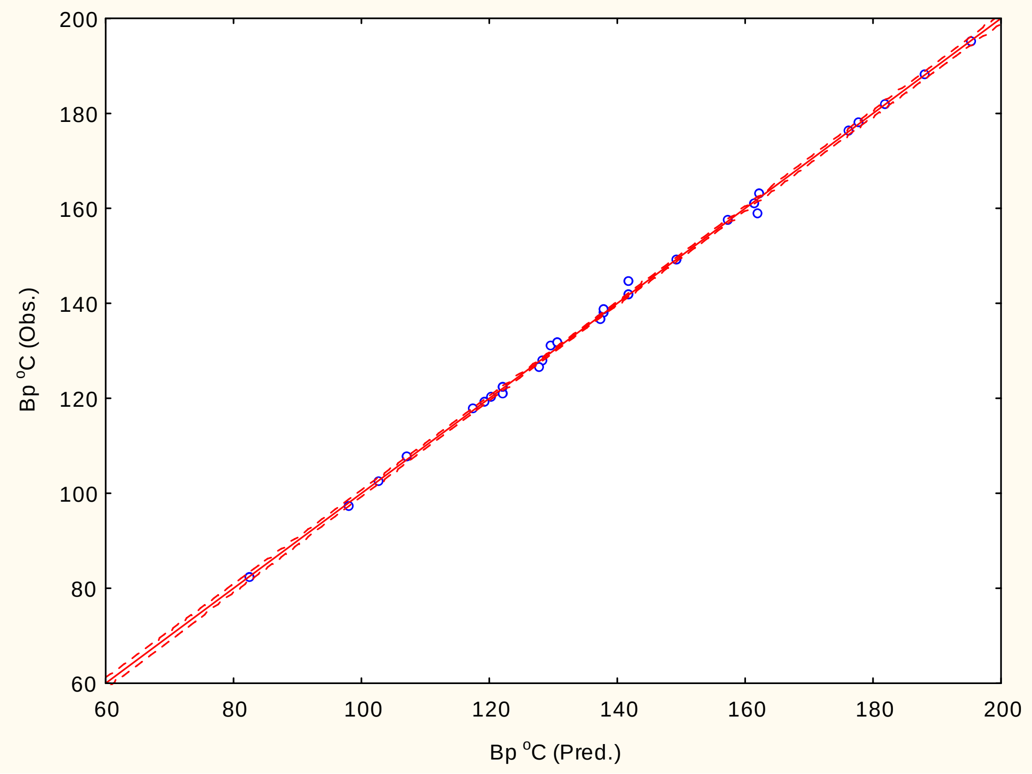


**Figure 2.** Linear correlations of observed versus calculated boiling point according to the model obtained from stochastic bond-level bilinear indices (Eq. **37**).