13rd International Electronic Conference on Synthetic Organic Chemistry (ECSOC-13), 1-30 November 2009.
http://www.mdpi.org/ecsoc-13/ & http://www.usc.es/congresos/ecsoc/13/

[g002]

Applications of Bond-Based 3D-Chiral Quadratic Indices in QSAR Studies Related to Central Chirality Codification.

Juan A. Castillo-Garit,[a,b,c,d,*] Yovani Marrero-Ponce,[b,d] Francisco Torrens,[d] Ramon García-Domenech[c] and J. Enrique Rodríguez-Borges.[e]

[a]*Applied Chemistry Research Center, Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.* **e-mail:** jacgarit@yahoo.es,or juancg@uclv.edu.cu
[b]*Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Department of Pharmacy, Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba*
[c]*Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, València, Spain.*
[d]*Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, 46071 Valencia, Spain*
[e]Centro de Investigação em Química (UP), Departamento de Química, Faculdade de Ciências, Universidade do Porto, R. Campo Alegre, 687, P-4169-007 Porto, Portugal

Abstract

The concept of bond-based quadratic indices is generalized to codify chemical structure information for chiral drugs, making use of a trigonometric 3D-chirality correction factor. In order to evaluate the effectiveness of this novel approach in drug design, we have modeled several well-known data sets. In particularly, Cramer's steroid data set has become a benchmark for the assessment of novel QSAR methods. This data set has been used by several researchers using 3D-QSAR approaches. Therefore, it is selected by us for the shake of comparability. In addition, to evaluate the effectiveness of this novel approach in drug design, we model the angiotensin-converting enzyme inhibitory activity of perindoprilate's σ-stereoisomers combinatorial library, as well as codify information related to a pharmacological property, highly dependent on the molecular symmetry, of a set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)-piperidines, which bind σ-receptors. The validation of this method is achieved by comparison with earlier publications applied to the same data sets. The non-stochastic and stochastic bond-based 3D-chiral quadratic indices appear to provide a rather interesting alternative to other more common 3D-QSAR descriptors.

Keywords: non-stochastic and stochastic bond-based 3D-chiral quadratic indices, 3D-QSAR, angiotesin-converting enzyme inhibitor, σ-receptor antagonist, binding affinity steroid.

1. Introduction

The asymmetry of atomic configurations is an important feature in determining the physical, chemical and biological properties of chemical substances [1]. In the literature, the asymmetric atoms are often referred to as chiral atoms, and molecules containing chiral atoms are referred to as chiral molecules. Two molecules with identical chemical formulas, but different states of symmetry of an only atom, are referred to as enantiomers, although may also be referred to as enantiomorphs, optical isomers or optical antipodes [2]. The molecules with identical 2D structural formulas containing more than one asymmetric atom are referred to as σ-diastereomers [3]. Usually enantiomers exhibit different chemical and physical properties, as well as different biological activities [4]. The case of thalidomide is an example of a problem that was, at least, complicated by the ignorance of stereochemical effects [5]. Thus, whenever a drug is obtained in a variety of chemically equivalent forms (such as a racemate), it is both good science and good sense to explore the potential for *in vivo* differences between these forms. In connection with this, the regulation of the US Food & Drug Administration (FDA) requires a detailed study of both enantiomers [6].

In view of the great importance of molecular chirality in chemistry, biochemistry, pharmacology, etc., much effort has been made to design theoretical methods by which enantiomeric species could be distinguished [1, 2, 4, 7-14]. Nevertheless, rather few of these descriptors have been reported in the literature to date, although the necessity of a more serious effort in this direction has been recognized by researchers in the area [15]. Among the chiral topological indices (CTIs) published in the literature, Estrada and Uriarte mentioned some in a recent review about topological indices (TIs) [15]. Pyka [16-18], as well as Gutman and Pyka [19] rationalized some of these indices from a mathematical point of view. The relationships between these indices and the Wiener index were established. Moreover, Schultz *et al.*[13] modified a series of TIs in order to introduce information regarding the chirality of stereocenters in the molecules.

Some years ago, Buda and Mislow distinguished between two classes of measures [20]. In the first class, ‘the degree of chirality expresses the extent to which a chiral object differs from an achiral reference object’. In the second one, ‘it expresses the extent to which two enantiomorphs differ from one another’. Either method yields a single real value, usually an absolute quantity which is the same for both enantiomorphs. A different idea was to incorporate R/S labels into conventional topological indices [13]. Derived chirality descriptors were correlated with biological activity by de Julián-Ortiz *et al.* [10], Golbraikh *et al.* [1] and more recently by Díaz *et al*. [21]. One of the first approaches in this field was introduced by de Julián-Ortiz *et al.* [10], in a study of the pharmacological activity of different pairs of

enantiomers on dopamine D_2 and the σ-receptor. Fortunately, the so-called chiral topological indices (CTIs) are inexpensive in terms of computational time, in comparison to grid dependent methods like CoMFA [22]. Anyways, when chirality is considered, many 3D-TIs become 'hard to interpret' in physical terms. For example, Golbraikh, Bonchev, and Tropsha's work generated even complex numbers that are incompatible with standard statistical software [1].

In two recent works of Aires-de-Sousa and Gasteiger [4, 8], two different kinds of chirality codes were designed, named "conformation-independent chirality code" (CICC) and "conformation-dependent chirality code" (CDCC). The chirality code is a molecular transformation that *represents* the chirality of a molecule by using a spectrum-like, fixed-length code and includes information about the geometry of chiral centers, properties of the atoms in the neighborhoods, and bonds lengths. The code distinguishes between enantiomers and yields descriptors with symmetrical values for opposite enantiomers [14].

Recently, a novel scheme to the rational *–in silico-* molecular design and to QSAR/QSPR has been introduced by our research group: ***TOMOCOMD*** **(**acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign). It calculates several new families of 2D, 3D-Chiral (2.5) and 3D (geometric and topographical) non-stochastic and stochastic atom- and bond-based molecular descriptors, based on algebraic theory and discrete mathematics. They are denoted quadratic, linear and bilinear indices, and have been defined in analogy to the quadratic, linear and bilinear mathematical maps [23-27]. These approaches describe changes in the electron distribution with time throughout the molecular backbone, and they have been successfully employed in the prediction of several physical, physicochemical, chemical, biological, pharmacokinetic and toxicological properties of organic compounds [28-38]. Besides, these indices have been extended to consider three-dimensional features of small/medium-sized molecules based on the *trigonometric 3D-chirality correction factor approach* [39-43]. In earlier publications, we have obtained rather promising results when stochastic and non-stochastic atom-based 3D-chiral quadratic, linear and bilinear indices were applied to three of the most commonly used chiral data sets [39-43].

The present report is written with two objectives in mind. First, to extend the non-stochastic and stochastic bond-based 2D quadratic indices in order to codify chirality-related structural features and second, to compare the achieved results with those obtained by other methods. The problems of the prediction of corticosteroid-binding globulin (CBG) affinity of the Cramer's steroid data set, σ-receptor antagonist activities and the classification of ACE (angiotesin-converting enzime) inhibitors are selected as illustrative examples of applications of the method. These examples will be used as a matter of comparison with other CTIs, 3D and quantum chemical descriptors as well.

2. Theoretical Scaffold

The basis of the extension of quadratic 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 Molecular Descriptors (MDs) [46-51]. In this section, first we will define the nomenclature to be used in this work, then, the atom-based molecular vector ($\bar{x}$) will be redefined for bond characterization, using the same approach as previously reported and, finally, some new definition of bond-based non-stochastic and stochastic quadratic indices, with its peculiar mathematical properties, will be given.

2.1 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 **E** of G (likewise called bond-adjacency matrix, **B**) is a symmetric square matrix, whose elements e_{ij} are 1 if and only if edge i is adjacent to edge j [46, 49, 52]. Two edges are adjacent if they are incidental to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the ith row (or column) of **E** is named the edge degree of bond i, $\delta(e_i)$ [46-52].

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

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

where ${}^{k}e_{ij}$ are the elements of the kth power of **E,** and the SUM of the ith row of $\mathbf{E}^{k}$ is named the k-order edge degree of bond i, ${}^{k}\delta(e)_{i}$. Note that the matrix $\mathbf{ES}^{k}$ in Eq. **1** has the property that *the sum of the elements in each row* is 1. An mxm matrix with nonnegative entries having this property is called a "**stochastic matrix**" [53]. Recently, some authors have introduced the stochastic approach to atomic relationships to derive new MDs [54-59].

2.2 Chemical Information and Bond-Based Molecular Vector

The atom-based molecular vector ($\bar{x}$), used to represent small-to-medium sized organic chemicals has been explained elsewhere in some detail [23, 60]. In a manner parallel to the development of $\bar{x}$, we present the expansion of the bond-based molecular vector ($\bar{w}$). The components ($\bar{w}$) of w are numeric

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 us to encode organic molecules, such as 2-hydroxybut-2-enenitrile, through the molecular vector $\overline{w}$ = [$w_{\text{Csp3-Csp2}}$, $w_{\text{Csp2=Csp2}}$, $w_{\text{Csp2-Osp3}}$, $w_{\text{H-Osp3}}$, $w_{\text{Csp2-Csp}}$, $w_{\text{Csp}\equiv\text{Nsp}}$]. This vector belongs to the product space $\Re^6$. These properties characterize each kind of bond (and bond-type) in the molecule. Diverse kinds of bond weights (*w*) can be used, in order to codify information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distance [61, 62], standard bond dipole [61, 62] or even mathematical expressions involving atomic weights such as atomic log P [63], surface contributions of polar atoms [64], atomic molar refractivity [65], atomic hybrid polarizabilities [66], Gasteiger-Marsilli atomic charge [67], atomic electronegativity in Pauling scale [68] and so on. Here, we characterized each bond with the following parameter:

$$w_i = x_i/\delta_i + x_j/\delta_j \quad \textbf{(2)}$$

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

2.3 Theory of Non-Stochastic and Stochastic Total (Whole) and Local (Bond, Group and Bond-Type) Quadratic Indices

If a molecule consists of *m* bonds (*vector of* $\Re^m$), then the k^{th} total quadratic indices are calculated as quadratic maps (quadratic form) in $\Re^m$, in canonical basis set. Specifically, the k^{th} total non-stochastic and stochastic quadratic indices, $q_k(\overline{w})$ and ${}^{s}q_k(\overline{w})$, are computed from these k^{th} non-stochastic and stochastic edge adjacency matrices, $\mathbf{E}^k$ and $\mathbf{ES}^k$, as shown in Eqs. 3 and 4, correspondingly:

$$q_k(\overline{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^{k}e_{ij}\, w^i w^j = [\mathrm{W}]^{\mathrm{t}} \mathbf{E}^k [\mathrm{W}] \quad \textbf{(3)}$$

$${}^{s}q_k(\overline{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^{k}es_{ij}\, w^i w^j = [\mathrm{W}]^{\mathrm{t}} \mathbf{ES}^k [\mathrm{W}] \quad \textbf{(4)}$$

where *m* is the number of bonds of the molecule, and $w^1,...,w^m$ are the coordinates of the bond-based molecular vector ($\overline{w}$) in the so-called canonical ('natural') basis set. In this basis set, the coordinates of any vector $\overline{w}$ coincide with the components of this vector [69, 70]. Therefore, those coordinates can be considered as weights (bond labels) of the edge of the molecular graph. The coefficients ${}^{k}e_{ij}$ and ${}^{k}es_{ij}$ are the elements of the k^{th} power of the matrices **E**(G) and **ES**(G), correspondingly, of the molecular graph.

The defining equations (**3**) and (**4**) for $q_k(\overline{w})$ and ${}^{s}q_k(\overline{w})$, respectively, may be also written in matrix form (see Eqs. **3** and **4**), where [W] is a column vector (an mx1 matrix) of the coordinates of $\overline{w}$ in the canonical basis set of $\Re^m$, and $[W]^t$ (an 1xm matrix) is the transpose of [W]. Here, $\mathbf{E}^k$ and $\mathbf{ES}^k$ denote the matrices of quadratic maps with regard to the natural basis set.

In addition to total bond-based quadratic indices computed for the whole molecule, a local-fragment (bond, group and bond-type) formalism can be developed. These MDs are termed local non-stochastic and stochastic quadratic indices, $q_{kL}(\overline{w})$ and ${}^{s}q_{kL}(\overline{w})$, respectively. The definition of these descriptors is as follows:

$$q_{kL}(\overline{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^{k}e_{ijL}\, w^{i} w^{j} = [W]^{t}\, \mathbf{E}^{k}_{L}\, [W] \qquad (\mathbf{5})$$

$${}^{s}q_{kL}(\overline{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^{k}es_{ijL}\, w^{i} w^{j} = [W]^{t}\, \mathbf{ES}^{k}_{L}\, [W] \qquad (\mathbf{6})$$

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

${}^{k}e_{ijL}$ [${}^{k}es_{ijL}$] = ${}^{k}e_{ij}$ [${}^{k}es_{ij}$], *if* both e_i or e_j are edges (bonds) contained in the molecular fragment

= ½ ${}^{k}e_{ij}$ [${}^{k}es_{ij}$] *if* either e_i and e_j are edges (bonds) contained in the molecular fragment

= 0, otherwise (**7**)

Notice that the scheme above follows the spirit of a Mulliken population analysis.[71] Note also that for every partitioning of a molecule into Z molecular fragments, there will be Z local molecular-fragment matrices. In particular, if a molecule is partitioned into Z molecular fragments, the matrices $\mathbf{E}^k$ [$\mathbf{ES}^k$] can be 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 non-stochastic and stochastic bond-based quadratic indices are the sum of the non-stochastic and stochastic bond-based quadratic indices, respectively, of the Z molecular fragments:

$$q_k(\overline{w}) = \sum_{L=1}^{Z} q_{kL}(\overline{w}) \qquad (\mathbf{8})$$

$${}^{s}q_k(\overline{w}) = \sum_{L=1}^{Z} {}^{s}q_{kL}(\overline{w}) \qquad (\mathbf{9})$$

Bond, group and bond-type quadratic fingerprints are specific cases of local bond-based quadratic indices. Therefore, the *k*th bond-type quadratic indices are calculated by adding the k^{th}-bond quadratic indices for all bonds of the same type in the molecule. Likewise, this extension of the bond quadratic 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 quadratic index.

In the bond-type quadratic indices formalism, each bond in the molecule is classified into a bond-type (fragment). Therefore, bonds may be classified into bond types in terms of the characteristics of the two atoms that define the bond. For all data sets, including those with a common molecular scaffold as well as those with rather diverse structure, the k^{th} fragment (bond-type) quadratic indices provide much useful information. Thus, the development of the bond-type quadratic indices description provides the basis for an 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. The bond-type descriptors combine three important aspects of structure information: 1) electron accessibility for the bonds of the same type, 2) presence/absence of the bond type, and 3) count of the bonds in the bond type.

Finally, these local MDs can be calculated by a chemical (or functional) group in the molecule, such as heteroatoms (O, N and S in all valence states and including the number of attached H-atoms), hydrogen bonding (H-bonding) to heteroatoms (O, N and S in all valence states), halogen atoms (F, Cl, Br and I), all aliphatic carbon chains (several bond-types), all aromatic bonds (aromatic rings), and so on. The group-level quadratic indices are the sum of the individual bond-level quadratic indices for a particular group of bonds. For all data set structures, the k^{th} group-based quadratic indices provide also important information for QSAR/QSPR studies. A detailed example of the calculation of the bond-based quadratic indices can be seen in an earlier publication [27].

2.4 3D-Chiral Non-stochastic and Stochastic Bond-Based Quadratic Indices.

The total and local bond-based quadratic indices, as defined above, cannot codify any information about the 3D molecular structure. In order to solve this problem, we introduced a trigonometric 3D-chirality correction factor in the components (w) of $\overline{w}$**.** Therefore, a chirality molecular vector is obtained $^*\overline{w}$ and each bond will be characterized by the following parameter:

$$w_i = {}^*x_i/\delta_i + {}^*x_j/\delta_j \qquad \textbf{(10)}$$

Notice that this equation is quite similar to Eq. 2, but the atomic weights of the atoms that characterize the bond, x_i and x_j, where replaced by the terms $^*x_i = \{x_i + \sin[(\omega_A + 4\Delta)\pi/2]\}$ and $^*x_j = \{x_j + \sin[(\omega_A + 4\Delta)\pi/2]\}$ to take into account the 3D environment.

The *trigonometric 3D-chirality correction factor* uses a dummy variable, ω_A, and an integer parameter, Δ [39, 41, 43]:

ω_A = 1, and Δ is an odd number when *A* has R (rectus), E (entgegen), or *a* (axial) notation according to Cahn-Ingold-Prelog (CIP) IUPAC rules (**11**)

= 0, and Δ is an even number, if *A* does not have 3D specific enviroment

= -1, and Δ is an odd number when *A* has S (sinister), Z (zusammen), or *e* (equatorial) notation according to CIP rules

Thus, this 3D-chirality factor, $\sin[(\omega_A+4\Delta)\pi/2]$, takes different values in order to codify specific stereochemical information such as chirality, Z/E isomerism, and so on. This factor, therefore, takes values in the following order 1 > 0 > -1 for atoms that have specific 3D environments. The chemical idea here is not that the attraction of electrons by an atom depends on its chirality, because experience shows that chirality does not change the electronegativities of atoms in the molecule, in an isotropic environment in an observable manner [72]. This correction has mainly a mathematical meaning and must not be the source of any misunderstanding.

Table 7. Classification of 32 perindoprilate stereoisomers and the statistical parameters of the QSAR models obtained using different MDs.

Index	n	λ	D^2	% Accuracy (Training)	%Accuracy (Test)	F
Bond-based non-stochastic quadratic indices (Eq. 16)	2	0.447	5.86	100.00	88.88	12.37
Bond-based stochastic quadratic indices (Eq. 17)	2	0.448	5.84	95.65	88.88	12.32
Atom-based non-stochastic quadratic indices [40]	2	0.420	7.12	95.65	100.00	13.73
MARCH-INSIDE molecular descriptors [21]	3	0.380	8.43	91.30	88.88	10.30

n: number of parameters in the obtained model.

A severe limitation of the GBT [1] approach is the existence of different chirality corrections, and we have great difficulty in selecting one of these. Therefore, the present *trigonometric 3D-chiral correction factor* is invariant with regard to the selection of other chirality scales for all kinds of such chiral TIs (GBT-like ones). Table 1 depicts the values of the *trigonometric 3D-chirality correction factor* for all allowed values of ω_A and Δ (GBT-like chirality scale and other alternative chirality scales). In Table 1, it is clearly shown that the *trigonometric 3D-chirality factor* is invariant with regard to the selection of all possible real scales. Moreover, the factor ever gets the values 1, 0 and -1 for R, non-chiral and S atoms. As outlined above, the demonstration of invariance for this factor with regard to other 3D features such as *a*/*e* substitutions and Z/E or π-isomers is straightforward to realize by homology. Henceforth, we do

not need to answer the question regarding the best value for chirality correction, at least for linear scales [1, 10, 21].

A rather interesting point is that, for molecules without specific 3D characteristics, the present 3D-chiral descriptor is reduced to simple (2D) bond-based quadratic indices, because $\sin[(0+4\Delta)\pi/2] = 0$, being Δ zero or any even number. Therefore, when all the atoms in the molecule are achiral, the bond-based quadratic indices or any GBT-like chiral TIs do not change upon the introduction of this factor. Therefore, for example $^{*}\overline{w} = \overline{w}$ and thus, $^{*}\boldsymbol{q}_{k}(\overline{w}) = \boldsymbol{q}_{k}(\overline{w})$.

3. Experimental Section

3.1 Computational Strategies

All the computations were carried out on a PC Pentium-4 3.2 GHz. The TOMOCOMD package for Windows, developed in our laboratory, was used to compute the molecular descriptors for the dataset of compounds. This software is an interactive program for molecular design and bioinformatics research [73]. It consists of four subprograms; each one of them allows both drawing the structures (drawing mode) and calculating molecular 2D/3D descriptors (calculation mode). The modules are named CARDD (Computed-Aided 'Rational' Drug Design), CAMPS (Computed-Aided Modeling in Protein Science), CANAR (Computed-Aided Nucleic Acid Research) and CABPD (Computed-Aided Bio-Polymers Docking). In the present report, we outline salient features concerned with only one of these subprograms, CARDD, and with the calculation of non-stochastic and stochastic bond-based 3D chiral quadratic indices.

The main steps for the application of the present method in QSAR/QSPR and drug design can be summarized briefly in the following algorithm: 1) Draw the molecular structure for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol, which belong to the different groups in the periodic table of the elements; 2) Use appropriate weights in order to differentiate the atoms in the molecule. The weight used in this work is the atomic electronegativity in Pauling scale (E) [68]; 3) Compute the total and local (bond, group and bond-type) non-stochastic and stochastic quadratic indices. It can be carried out in the software calculation mode, where one can select the atomic property and the descriptor family before calculating the molecular indices. This software generates a table in which the rows correspond to the compounds, and columns correspond to the bond-based (both total and local) quadratic maps; 4) Find a QSPR/QSAR equation by using several multivariate analytical techniques, such as multilinear regression analysis (MRA), neural networks, linear discrimination analysis, and so on. Therefore, one can find a

quantitative relation between an activity **A** and the quadratic fingerprints having, for instance, the following appearance,

$$\mathbf{A} = a_0 {}^{*}\boldsymbol{q}_0(\overline{w}) + a_1 {}^{*}\boldsymbol{q}_1(\overline{w}) + a_2 {}^{*}\boldsymbol{q}_2(\overline{w}) + \ldots + a_k {}^{*}\boldsymbol{q}_k(\overline{w}) + \mathrm{c}$$

where **A** is the measured activity, ${}^{*}\boldsymbol{q}_k(\overline{w})$ are the k^{th} bond-based 3D-chiral quadratic indices, and the $\boldsymbol{a}_k$*'s* and c are the coefficients obtained by the MRA; 5) Test the robustness and predictive power of the QSPR/QSAR equation by using internal (cross-validation) and external validation techniques.

The bond–based *quadratic indices* descriptors computed in this study were the following:

1) k^{th} (k = 15) total non-stochastic bond-based 3D-chiral quadratic indices, not considering and considering H-atoms in the molecular graph (G) [${}^{*}\boldsymbol{q}_k(\overline{w})$ and ${}^{*}\boldsymbol{q}_k{}^{\mathrm{H}}(\overline{w})$, respectively].
2) k^{th} (k = 15) total stochastic bond-based 3D-chiral quadratic indices, not considering and considering H-atoms in the molecular graph (G) [${}^{*\mathbf{s}}\boldsymbol{q}_k(\overline{w})$ and ${}^{*\mathbf{s}}\boldsymbol{q}_k{}^{\mathrm{H}}(\overline{w})$, respectively].
3) k^{th} (k = 15) bond-type local (group = heteroatoms: S, N, O) non-stochastic 3D-chiral quadratic indices, not considering and considering H-atoms in the molecular graph (G) [${}^{*}\boldsymbol{q}_{k\mathrm{L}}(\overline{w}_{\mathrm{E}})$ and ${}^{*}\boldsymbol{q}_{k\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}})$, correspondingly]. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.
4) k^{th} (k = 15) bond-type local (group = heteroatoms: S, N, O) stochastic 3D-chiral quadratic indices, not considering and considering H-atoms in the molecular graph (G) [${}^{*\mathrm{s}}\boldsymbol{q}_{k\mathrm{L}}(\overline{w}_{\mathrm{E}})$ and ${}^{*\mathbf{s}}\boldsymbol{q}_{k\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}})$, correspondingly]. These local descriptors are also putative molecular charge, dipole moment, and H-bonding acceptors.

3.2. Chemometric analysis

Statistical analysis was carried out with the STATISTICA software [74]. The considered tolerance parameter (proportion of variance that is unique to the respective variable) was the default value for minimum acceptable tolerance, which is 0.01. Forward stepwise procedure was fixed as the strategy for variable selection. The principle of maximal parsimony (Occam's razor) was taken into account as the strategy for model selection. Therefore, we selected the model with the highest statistical signification, but having as few parameters (a_k) as possible.

Multiple Linear Regression (MLR) analysis was carried out to predict corticosteroid-binding globulin affinity of a steroid data set and the σ-receptor antagonist activities of 3-(3-hydroxyphenyl)piperidines. The quality of the models was determined by examining the regression's statistical parameters and those of the cross-validation analysis [75, 76]. Therefore, the quality of the models was determined by examining the determination coefficients (also know as square correlation coefficients, R^2), Fisher-

ratio's *p*-level [*p*(F)], standard deviation of the regression (*s*), and the leave-*one*-out (LOO) press statistics (q^2, s_{cv}) analogues to R^2 and *s* [75, 77].

On the other hand, *linear discriminant analysis* (LDA) was performed to classify the 32 perindoprilate stereoisomers as angiotensin-converting enzyme (ACE) inhibitors or not. The quality of the models were determined by examining Wilks' λ parameter (*U*-statistic), square Mahalanobis distance (D^2), Fisher ratio (F) and the corresponding *p*-level [p(F)] as well as the percentage of good classification in the training and test sets. The statistical robustness and predictive power of the obtained model was assessed by using an external prediction (test) set. In developing classification models, the values of 1 and -1 were assigned to active and inactive compounds, respectively. By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(Active) - P(Inactive)]x100$ or as inactive, otherwise. P(Active) and P(Inactive) are the probabilities with which the equations classify a compound as active and inactive, correspondingly.

Finally, the calculation of percentages of global good classification (accuracy) and Matthews' correlation coefficient (MCC), in the training and test sets, permitted the assessment of the model [78]. The MCC always takes values between -1 and +1. A value of -1 indicates total disagreement (all-false predictions), and +1, total agreement (perfect predictions). The MCC is 0 for completely random predictions and, therefore, it yields easy comparison with regard to a random baseline. Therefore, MCC quantifies the strength of the linear relation between the molecular descriptors and the classifications [78], and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages.

4. Result and Discussion

With the objective of assessing the efficacy of bond-based 3D-chiral quadratic indices, we have tested their ability to predict pharmacological properties, in several groups of compounds with a known stereochemical influence. We select these data sets because they have been repeatedly used in a number of QSAR studies in recent years. Now, we are going to discuss the use of the bond-based 3D-chiral quadratic indices descriptors in each one of these well-known series of compounds, and a comparison with other previously reported approaches will be also developed.

4.1 Prediction of the Corticosteroid-Binding Globulin (CBG) Binding Affinity of a Steroid Family.

The first molecular set used in our study is made up of 31 steroids, for which the affinity to the corticosteroid-binding globulin was measured. The so-called Cramer's steroid data set is a well-known benchmark to QSAR researchers [22, 79-89]. Various groups used this data set to compare the quality of their 3D-QSAR methods. Hence, this data set has become one of the most often discussed ones and can

be seen as a point of reference data set for novel MDs [90]. It was used here for the shake of comparability [88]. We use this molecular set because all the compounds in this data set contain chiral atoms, and binding affinities of these compounds are available [22]. Due to the fact that the studied steroid molecular structures have been already depicted in several publications [22, 80], they will not be included here. Table 2 gathers the entire studied set with the experimental binding affinities, taken from Robert *et al.* [87]. The obtained models are given below together with their statistical parameters:

$$\mathbf{CBG} = -6.17(\pm 1.54) - 0.044(\pm 0.02)\,{}^{*}\boldsymbol{q}_0(\overline{w}) + 6.8\text{x}10^{-2}(\pm 0.8\text{x}10^{-2})\,{}^{*}\boldsymbol{q}_2(\overline{w}) - 3.6\text{x}10^{-2}(\pm 0.5\text{x}10^{-2})\,{}^{*}\boldsymbol{q}_3(\overline{w}) + 0.5\text{x}10^{-2}(\pm 0.1\text{x}10^{-2})\,{}^{*}\boldsymbol{q}_4(\overline{w}) + 5.1\text{x}10^{-2}(\pm 1.2\text{x}10^{-2})\,{}^{*}\boldsymbol{q}_1{}^{\mathrm{H}}(\overline{w}) - 7.4\text{x}10^{-2}(\pm 1.8\text{x}10^{-2})\,{}^{*}\boldsymbol{q}_{2\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) + 3.9\text{x}10^{-3}(\pm 1.1\text{x}10^{-3})\,{}^{*}\boldsymbol{q}_{4\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) \quad (\mathbf{12})$$

$N = 31 \quad R = 0.923 \quad R^2 = 0.852 \quad F(7,23) = 18.93 \quad s = 0.47 \quad q^2 = 0.74 \quad s_{\mathrm{cv}} = 0.55 \quad p < 0.0001$

$$\mathbf{CBG} = -8.51(\pm 0.69) + 0.756(\pm 0.174)\,{}^{*\mathbf{s}}\boldsymbol{q}_{1\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) + 0.470 \pm 0.082\,{}^{*\mathbf{s}}\boldsymbol{q}_2{}^{\mathrm{H}}(\overline{w}) - 2.069(\pm 0.894)\,{}^{*\mathbf{s}}\boldsymbol{q}_{3\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) - 0.435(\pm 0.093)\,{}^{*\mathbf{s}}\boldsymbol{q}_1{}^{\mathrm{H}}(\overline{w}) + 19.992(\pm 4.975)\,{}^{*\mathbf{s}}\boldsymbol{q}_{8\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) - 29.764(\pm 8.794)\,{}^{*\mathbf{s}}\boldsymbol{q}_{7\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) + 11.280(\pm 4.382)\,{}^{*\mathbf{s}}\boldsymbol{q}_{6\mathrm{L}}{}^{\mathrm{H}}(\overline{w}_{\mathrm{E}}) \quad (\mathbf{13})$$

$N = 31 \quad R = 0.931 \quad R^2 = 0.867 \quad F(7,23) = 21.44 \quad s = 0.45 \quad q^2 = 0.78 \quad s_{\mathrm{cv}} = 0.50 \quad p < 0.0001$

where N is the size of the data set, R^2 is the square regression coefficient (determination coefficient), s is the standard deviation of the regression, F is the Fischer ratio and q^2 (s_{cv}) are the square correlation coefficient (standard deviation) of the cross-validation performed by the *leave-one-out* (LOO) procedure. As can be seen, the non-stochastic model (Eq. **12**) explains more than 85% of the variance of the experimental CBG values, using seven variables to describe the 31 steroids, while the stochastic model (Eq. **13**) explains more than 86% of this experimental value also using seven variables. The predicted values for this data set, using non-stochastic and stochastic bond-based 3D-chiral quadratic indices, are also shown in Table 2. The model's obtained deviation, with bond-based 3D-chiral quadratic indices, shows a smaller value of standard deviation (s = 0.45) than the one obtained with non-stochastic indices (s = 0.47).

An important aspect of QSAR modeling is the development of a way of performing the statistical validation of the models. Good direct statistical criteria to fit the data set are not a guarantee that the model can make accurate predictions for compounds outside the data set. The LOO statistic has been used as means of demonstrating predictive capability. These models showed cross-validation square correlation coefficients of 0.740 and 0.784, respectively. These values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models [75-77].

Table 2. Results of the steroids data set used for QSAR study.

		Observed CBG affinity (pKa)[a]	Non-stochastic bond-based 3D-chiral quadratic indices			Stochastic bond-based 3D-chiral quadratic indices		
			Predicted value	%E[b]	$\%E_{cv}$[c]	Predicted value	%E[b]	$\%E_{cv}$[c]
1	Aldosterone	-6.279	-6.152	2.016	5.262	-6.346	1.061	1.378
2	Androstanediol	-5.000	-5.425	8.504	13.763	-5.388	7.763	12.064
3	Androstenediol	-5.000	-5.082	1.633	1.995	-4.957	0.869	1.136
4	Androstenedione	-5.763	-6.658	15.524	19.881	-6.451	11.945	13.865
5	Androsterone	-5.613	-5.680	1.193	1.298	-5.862	4.432	5.739
6	Corticosterone	-7.881	-7.437	5.638	7.181	-7.375	6.423	7.575
7	Cortisol	-7.881	-7.512	4.677	5.281	-7.463	5.305	6.465
8	Cortisone	-6.892	-7.461	8.259	9.716	-7.542	9.437	11.412
9	Dehydroepiandrosterone	-5.000	-4.745	5.093	7.469	-4.923	1.548	2.681
10	Deoxycorticosterone	-7.653	-7.629	0.320	0.415	-8.008	4.641	6.417
11	Deoxycortisol	-7.881	-7.907	0.326	0.391	-7.772	1.380	1.807
12	Dihydrotestosterone	-5.919	-5.191	12.301	14.859	-5.037	14.896	18.513
13	Estradiol	-5.000	-4.821	3.571	4.457	-4.841	3.185	4.750
14	Estriol	-5.000	-5.103	2.055	3.664	-5.035	0.696	1.585
15	Estrone	-5.000	-5.254	5.081	7.260	-4.975	0.506	0.752
16	Ethiocholanolone	-5.255	-4.977	5.294	6.612	-5.181	1.407	1.816
17	Pregnenolone	-5.255	-5.936	12.968	14.338	-5.782	10.033	11.251
18	17-Hydroxyregnenolone	-5.000	-5.735	14.700	20.475	-5.584	11.680	15.140
19	Progesterone	-7.380	-7.105	3.731	4.294	-7.138	3.276	3.852
20	17-Hydroxyprogesterone	-7.740	-6.878	11.143	15.096	-6.975	9.890	13.210
21	Testosterone	-6.724	-6.226	7.412	8.731	-6.332	5.832	6.708
22	Prednisolone	-7.512	-7.460	0.696	0.865	-7.728	2.873	3.463
23	Cortisolacetate	-7.553	-7.430	1.623	3.788	-7.486	0.882	8.721
24	4-Pregnene-3,11,20-trione	-6.779	-6.835	0.825	0.975	-6.861	1.205	1.670
25	Epicorticosterone	-7.200	-7.779	8.042	11.291	-7.120	1.105	1.239
26	19-Nortestosterone	-6.144	-5.954	3.092	4.822	-6.277	2.165	2.664
27	16α,17α-Dihydroxyprogesterone	-6.247	-6.291	0.697	0.857	-6.839	9.471	10.610
28	16α -Methylprogesterone	-7.120	-7.203	1.172	1.495	-7.052	0.955	1.114
29	19-Norprogesterone	-6.817	-6.840	0.344	0.467	-6.692	1.829	2.076
30	2α -Methylcortisol	-7.688	-7.237	5.860	7.836	-7.157	6.913	8.113
31	2α -Methyl-9α-fluorocortisol	-5.797	-6.030	4.020	10.274	-5.795	0.034	0.115

[a]Observed CBG affinity values taken from Robert *et al.*; concentration are expressed in nM [87]. [b]E: Error. [c]E_{cv}: Error of LOO validation.

As we previously pointed out, one of the objectives of the present report is to compare with other methods used for this data set. The results of these publications are summarized in Table 3, where the results were arranged in decreasing value of q^2, and the comparison can more easily be carried out. It is remarkable that the present QSAR method, non-stochastic and stochastic bond-based 3D-chiral linear indices, obtains results that favorably compare to other highly predictive QSAR models, even when they use more sophisticated statistic methods such as partial least squared, principal components analysis, non-linear neural network techniques and so on. Many of the models, object of comparison, were

obtained from different procedures based on quantum mechanics and/or geometric principles, as well as molecular mechanic approaches.

Table 3. Comparison between prediction for the steroid data set with bond-based 3D-chiral quadratic indices and other 3D QSAR approaches.

QSAR Method	N	n	Statistical method	q^2	Ref.
TQSAR	31	6	MLR after PCA	0.842	[87]
Bond-based stochastic 3D-chiral quadratic indices	31	7	MLR	0.784	Eq 13
Atom-based 3D-chiral quadratic indices (*non-stochastic*)	31	6	MLR	0.781	[41]
MEDV	31	5	GA and RLM	0.777	[93]
TQSI	31	3	MLR	0.775	[83]
MEDV	31	6	GA and RLM	0.765	[93]
Bond-based non-stochastic 3D-chiral quadratic indices	31	7	MLR	0.740	Eq 12
Atom-based 3D-chiral quadratic indices (*stochastic*)	31	7	MLR	0.735	[41]
Similarity indices	31	1	PLS	0.734	[85]
E-State and kappa shape index	31	4	MLR*	0.730	[94]
MQSM	31	4	MLR after PLS	0.727	[84]
E-State and kappa shape index	31	4	MLR	0.720	[94]
MQMS	31	3	MLR and PCA	0.705	[83]
CoMMA	31	6	PCR	0.689	[95]
MEDV	31	4	GA and RLM	0.648	[93]
Wagener's	31	-	k-NN and FNN	0.630	[81]

N: number of steroids. **n:** number of variables. q^2**:** leave-one-out cross-validated coefficient of determination.
***One variable has a non-linear relationship

4.2 Modeling σ-Receptor Antagonist Activities of 3-(3-hydroxyphenyl)piperidines

In a second application, we investigate the ability of bond-based 3D-chiral quadratic indices to predict σ receptor antagonistic activities. A short data set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)piperidines which bind to σ-receptors, are also selected as illustrative example of the 3D-chiral bilinear indices application. This data set was introduced in QSAR studies by de Julian-Ortiz *et al.* [10] in 1998 and, after that, it has been repeatedly used by some authors [21, 39, 40, 43] in recent years to validate new CTIs. The σ-receptors mediate severe side effects induced by various dopamine antagonists [10].

Bond-based 3D-chiral quadratic indices are non-symmetric and reduce to classical descriptors when symmetry is not codified. Besides, Gónzalez-Díaz *et al*. conclude that σ-receptor antagonist activity is not a pseudoscalar property [21], and we can expect, at least, a good correlation with bond-based 3D-chiral quadratic indices. The multiple linear regression (MLR) analysis was used to develop QSAR models for the σ-receptor antagonistic activities. The models obtained using non-stochastic and

stochastic bond based 3D-chiral quadratic indices for the σ-receptor antagonistic activities are given below:

$$\mathbf{logIC_{50}(\sigma)} = -9.214(\pm0.961) + 0.146(\pm0.022)\,{}^{*}\boldsymbol{q}_{0}(\overline{w}) + 5.86\times10^{-11}(\pm2.22\times10^{-11})\,{}^{*}\boldsymbol{q}_{15}{}^{H}(\overline{w}) \quad (\mathbf{14})$$

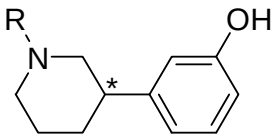
$N = 14 \quad R^2 = 0.939 \quad q^2_{LOO} = 0.882 \quad F(2, 11) = 84.53 \quad s = 0.272 \quad s_{cv} = 0.348 \quad p < 0.0001$

$$\mathbf{logIC_{50}(\sigma)} = -7.372(\pm0.442) + 1.042(\pm0.102)\,{}^{*s}\boldsymbol{q}_{0}{}^{H}(\overline{w}) - 1.278(\pm0.148)\,{}^{*s}\boldsymbol{q}_{2}{}^{H}(\overline{w}) \quad (\mathbf{15})$$

$N = 14 \quad R^2 = 0.969 \quad q^2_{LOO} = 0.956 \quad F(2, 11) = 172.57 \quad s = 0.194 \quad s_{cv} = 0.213 \quad p < 0.0001$

where N is the size of the data set, R^2 is the square correlation coefficient (determination coefficient), s is the standard deviation of the regression, F is the Fischer ratio and q^2 (s_{cv}) is the square correlation coefficient (standard deviation) of the cross-validation performed by the LOO procedure. These statistics indicate that both models are appropriate for the description of the chemicals studied here. In Table 4, the structure and values of the experimental and predicted values of Log IC_{50} (50 percent inhibitory concentration in nM) are shown for this data set.

Table 4. Results of multivariate regression analysis of the log IC_{50} of a group of *n*-alkylated 3-(3-hydroxyphenyl)piperidines for the σ-receptor.



Compound (Alkyl group)[a]	Log IC_{50} (σ-receptor)				
	Obs.[b]	Cal.[c]	Res.[d]	Cal.[e]	Res.[d]
(R)-3-HPP					
H	-0.66	-0.58	-0.08	-0.74	0.08
CH_3	0.43	0.00	0.43	0.44	-0.01
C_2H_5	0.95	0.71	0.24	0.70	0.25
n-C_3H_7	1.52	1.37	0.15	1.32	0.20
i-C_3H_7	0.61	1.11	-0.50	0.91	-0.30
n-C_4H_9	2.05	2.06	-0.01	1.97	0.08
2-Phenylethyl	2.10	1.93	0.17	2.22	-0.12
(S)-3-HPP					
H	-1.19	-1.09	-0.10	-1.26	0.07
CH_3	-0.28	-0.44	0.16	-0.02	-0.26
C_2H_5	-0.01	0.30	-0.31	0.23	-0.24
n-C_3H_7	0.81	0.98	-0.17	0.85	-0.04
i-C_3H_7	0.68	0.75	-0.07	0.43	0.25
n-C_4H_9	1.51	1.68	-0.17	1.51	0.00
2-Phenylethyl	1.80	1.55	0.25	1.76	0.04

[a]Alkyl (R) group in nitrogen ring. [b]Observed values of the Log IC_{50}, C in nM, for the σ-receptor taken from the literature[10, 21, 40]. [c]Values calculated using non-stochastic bond-based 3D-chiral quadratic indices (Eq. 14). [d]Residual defined as [Log IC_{50} (σ)Obs – Log IC_{50} (σ)Cal]. [e]Values calculated using stochastic bond-based 3D-chiral quadratic indices (Eq. 15).
Abbreviations: HPP, *N*-alkylated 3-Hydroxyphenyl piperidines.

These QSAR models use two variables, explain about the 94 % and 97% of the experimental values of log IC_{50} and show low values of standard deviation, 0.272 and 0.194, for bond-based non-stochastic and stochastic 3D-chiral linear indices models, correspondingly. The comparison of the results with the ones of other methods previously reported for the same activity is shown in Table 5, where an analyis between all these approaches can be easily carried out. As can be seen, the non-stochastic model results have statistical parameters which are similar to those obtained by Marrero-Ponce *et al*., using atom-based 3D-chiral linear indices [39], and better than the ones of models obtained with MARCH-INSIDE molecular descriptors [21] and other chiral TIs [10]. Moreover, the model obtained with stochastic bond-based quadratic indices showed better statistical parameters than the other QSAR models.

Predictability and stability (robustness) of the obtained models, with regard to data variation were carried out here by means of LOO cross-validation. The models showed values of cross-validation determination coefficient (q^2) of 0.882 and 0.956, when non-stochastic and stochastic bond-based quadratic indices were used, respectively. The values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models [75, 76, 91]. Unfortunately, the authors of previous works, Diaz *et al*. [21] and de Julian de Ortiz *et al.* [10], did not report the result of the cross-validation. Considering all these statistical criteria, we can conclude that the model obtained with stochastic bond-based 3D-chiral quadratic indices is the best QSAR model for describing the property studied in this section.

Table 5. Statistical parameters of the QSAR models obtained using bond-based 3D-chiral quadratic indices to predict the σ-Receptor antagonist activity of 14 N-alkylated 3-Hydroxyphenyl piperidines.

index	***N***	**n**	$\mathbf{R^2}$	***s***	q^2	s_{cv}	**F**
Bond-based stochastic quadratic indices	14	2	0.969	0.194	0.956	0.213	172.57
Bond-based non-stochastic quadratic indices	14	2	0.939	0.272	0.882	0.348	84.53
Atom-based non-Stochastic Quadratic indices [40]	14	2	0.940	0.270	0.912	0.289	85.82
Chiral TIs [10]	14	3	0.931	0.301	*	*	45.70
MARCH-INSIDE molecular descriptors [21]	14	2	0.922	0.295	*	0.32	71.17

*Values are not reported in the literature.

4.3 Classification of the ACE Inhibitory Activity of 32 Perindoprilate's σ-Stereoisomers

Finally, in order to validate even more the bond-based 3D-chiral quadratic indices in QSAR studies, a recently introduced data set of 32 perindoprilate stereoisomers, angiotensin-converting enzyme (ACE) inhibitors [92] was used. Enzyme ACE acts in plasma and blood vessels, removing the C-terminal dipeptide of decapeptide Angiotesin I to produce the potent blood vessel-constricting octapeptide Angiotesin II. In addition, ACE inactivates the hypotensive nonapeptide bradykinin. Therefore, ACE is

the biological target of many important antihypertensive drugs called ACE inhibitors (ACEIs) [92]. In this study, 'active' is taken to mean a compound that has an IC_{50} value not greater than 110 nM. The obtained classification models are given below, together with the LDA statistical parameters:

$$\mathbf{ACEiactv} = 26.765 + 4.297\text{x}10^{-10*}\boldsymbol{q}_{15}^{\mathbf{H}}(\overline{w}) - 2.821\text{x}10^{-5*}\boldsymbol{q}_{9}(\overline{w}) \quad (\mathbf{16})$$

N = 23 λ = 0.447 $D^2 = 5.857$ F(2, 20) = 12.37 $p < 0.0003$

$$\mathbf{ACEiactv} = 30.845 + 1.390^{*\mathbf{s}}\boldsymbol{q}_{3}(\overline{w}) - 5.144^{*\mathbf{s}}\boldsymbol{q}_{1}^{\mathbf{H}}(\overline{w}) \quad (\mathbf{17})$$

N = 23 λ = 0.448 $D^2 = 5.835$ F(2, 20) = 12.32 $p < 0.0003$

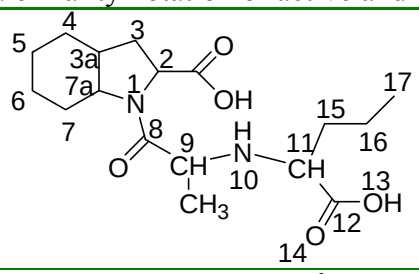
where N is the number of compounds, λ is the Wilks' statistic, D^2 is the square Mahalanobis distance, F is the Fisher ratio and *p*-value is the significance level.

The model (**16**), which includes non-stochastic bond-based quadratic indices, has an accuracy of 100% for the training set. This model showed a high Matthews' correlation coefficient (MCC) of 1. The most important criterion for the acceptance or not of a discriminant model is based on the statistics for the external prediction set. Model (**16**) correctly classifies 100.00% of active (isomers 1, 2 and 4) and 83.33% of inactive (isomers 12, 16, 20, 24 and 28) compounds in the test set, for an accuracy of 88.88% (MCC = 0.79). In Table 6, we give the basic structure of perindoprilate stereoisomers and their classification into the training and prediction sets, together with their posterior probabilities calculated from the Mahalanobis distance.

A similar behaviour was obtained with stochastic linear indices (Eq. **17**). In this case, the model correctly classifies 83.33% of active (isomers 3, 5, 6, 7 and 8) and 100% of inactive ones (compounds 10, 11, 13-15, 17-19, 21-23, 25-27, 29-31), for accuracy of 95.65% and a high MCC of 0.887 for the training set. As we previously pointed out, the analisys of the statistics for external prediction sets is the main criterion for the acceptance or not of a discriminant model. In this sense, the stochastic model shows the same behaviour as the non-stochastic model with an accuracy of 88.88% and MCC = 0.79.

A comparison between the results obtained in our study and those achieved with other cheminformatic approaches is depicted in Table 7. It should be remarked that our models contain one variable less than the model obtained with MARCH-INSIDE molecular descriptors,[21] as well as the same number of variables that were used by us with atom-based linear indices to develop the QSAR models [41]. The obtained results with non-stochastic bond-based linear indices are quite similar to the ones obtained with atom-based linear indices, but the statistical parameters of the model developed with stochastic bond-based linear indices are the best of all the models.

Table 6. Basic structure and chirality notation of active and inactive perindoprilate stereoisomers.



No	Comp.[a]	Class[b]	IC_{50}[c]	Class	ΔP[d]	Class	ΔP[d]
				Non-stochastic bond-based quadratic indices (Eq. 16)		Stochastic bond-based quadratic indices (Eq. 17)	
Active Compounds							
1	SSRSS*	+	1.1	+	0.610	+	0.864
2	SRSSS*	+	1.2	+	0.990	+	0.986
3	*SSSSS*	+	1.5	+	0.936	+	0.931
4	SRRSS*	+	3.3	+	0.950	+	0.952
5	*SSSSR*	+	12.2	+	0.799	+	0.749
6	*SSRSR*	+	29.4	+	0.082	+	0.544
7	*SRRSR*	+	39.8	+	0.834	+	0.818
8	*SRSSR*	+	54	+	0.968	+	0.945
9	*RRSSS*	+	108	+	0.103	-	-0.619
Non-active Compounds							
10	*SSSRS*	-	$1.1x10^3$	-	-0.277	-	-0.298
11	*RSSSS*	-	$1.9x10^3$	-	-0.705	-	-0.926
12	SSRRR*	-	$2.6x10^3$	-	-0.969	-	-0.842
13	*RRSSR*	-	$5.5x10^3$	-	-0.485	-	-0.890
14	*SSRRS*	-	$7.1x10^3$	-	-0.890	-	-0.514
15	*RRSRS*	-	$7.8x10^3$	-	-0.971	-	-0.989
16	RSRRR*	-	$23x10^3$	-	-1.000	-	-1.000
17	*SRRRR*	-	$33x10^3$	-	-0.774	-	-0.587
18	*RSSSR*	-	$36x10^3$	-	-0.906	-	-0.981
19	*RSRSR*	-	$47x10^3$	-	-0.991	-	-0.989
20	RSRSS*	-	$60x10^3$	-	-0.966	-	-0.957
21	*RRRRR*	-	10^5	-	-0.999	-	-0.999
22	*SRRRS*	-	10^5	-	-0.353	-	-0.013
23	*RRRSS*	-	10^5	-	-0.711	-	-0.853
24	SRSRR*	-	10^5	-	-0.029	-	-0.143
25	*RRRRS*	-	10^5	-	-0.997	-	-0.995
26	*RRSRR*	-	10^5	-	-0.992	-	-0.997
27	*SSSRR*	-	10^5	-	-0.727	-	-0.748
28	RSSRS*	-	10^5	-	-0.995	-	-0.998
29	*RRRSR*	-	10^5	-	-0.912	-	-0.961
30	*RSSRR*	-	10^5	-	-0.999	-	-1.000
31	*RSRRS*	-	10^5	-	-1.000	-	-0.999
32	SRSRS*	-	10^5	+	0.547	+	0.475

*Compounds used in the test set. [a]Notation of the chiral centers in each perindoprilate derivative in the following order C_2, C_{3a}, C_{7a}, C_9, C_{11}. [b]Classification according to the value of the IC_{50}. [c]Values of the IC_{50}, of the compound, for ACE in nM taken from previous works [10, 21].

Table 7. Classification of 32 perindoprilate stereoisomers and the statistical parameters of the QSAR models obtained using different MDs.

Index	n	λ	D^2	% Accuracy (Training)	%Accuracy (Test)	F
Bond-based non-stochastic quadratic indices (Eq. 16)	2	0.447	5.86	100.00	88.88	12.37
Bond-based stochastic quadratic indices (Eq. 17)	2	0.448	5.84	95.65	88.88	12.32
Atom-based non-stochastic quadratic indices [40]	2	0.420	7.12	95.65	100.00	13.73
MARCH-INSIDE molecular descriptors [21]	3	0.380	8.43	91.30	88.88	10.30

n: number of parameters in the obtained model.

5. Concluding Remarks

The non-stochastic and stochastic bond-based 3D-chiral quadratic indices are a novel set of MDs. They can be successfully applied in QSAR studies that include chiral molecules. Therefore, we suggest that 2D-QSAR methods, improved by chirality descriptors, could be a powerful alternative to popular 3D-QSAR approaches.

Our studies demonstrated that bond-based 3D-chiral quadratic indices are able not only to discriminate between active and inactive perindoprilate stereoisomers, but also to codify information related to the pharmacological property, highly dependent on molecular symmetry, of a set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)-piperidines that bind σ-receptors, as well as to predict the corticosteroid-binding globulin affinity of the Cramer's steroid data set. Moreover, we show that for the three data sets the chiral-QSAR models obtained with bond-based 3D-chiral quadratic indices had better or similar predictive ability, as compared to other chiral and/or 3D-QSARs previously reported.

Acknowledgement: Castillo-Garit, J.A. and M-P, Y. acknowledges to the program 'Estades Temporals per a Investigadors Convidats' for a fellowship to work at Valencia University in 2008, C-G. J.A. also thanks to the program 'Becas para la formación especializada de jóvenes investigadores de países en vías de desarrollo' for a fellowship to work at Valencia University (2007). RGD acknowledges financial support of the Fondo de Investigación Sanitaria, Ministerio de Sanidad, Spain (project SAF2005–PI052128). FT acknowledges financial support from the Spanish MEC (Project Nos. CTQ2004-07768-C02-01/BQU) and CCT-005-07-00365) and EU (Program FEDER).

References and Notes

1. A. Golbraikh, D. Bonchev,A. Tropsha, *J Chem Inf Comput Sci* **2001,** *41*, 147-58.

2. J. V. de Julian-Ortiz, R. Garcia-Domenech, J. Galvez, R. Soler , F. J. García-March,G. M. Antón-Fos, *J Chromatogr A* **1996,** *719*, 37-44.

3. V. M. Potapov, *Stereochemistry*. Khimia: Moscow, 1988;

4. J. Aires-de-Sousa,J. Gasteiger, *J Mol Graph Model* **2002,** *20*, 373-88.

5. H. Schumacher, D. A. Blake, J. M. Gurian,J. R. Gillette, *J Pharmacol Exp Ther* **1968,** *160*, 189-200.

6. S. C. Stinson, *Chem. Eng. News.* **2000,** *78*, 43.

7. J. Aires-de-Sousa, J. Gasteiger, I. Gutman,D. Vidovic, *J Chem Inf Comput Sci* **2004,** *44*, 831-6.

8. J. Aires-de-Sousa,J. Gasteiger, *J Chem Inf Comp Sci* **2001,** *41*, 369-375.

9. E. Ruch, *Acc Chem Res* **1972,** *5*, 49-56.

10. J. V. de Julian-Ortiz, C. de Gregorio Alapont, I. Rios-Santamarina, R. Garcia-Domenech,J. Galvez, *J Mol Graph Model* **1998,** *16*, 14-18.

11. R. Benigni, M. Cotta-Ramusino, G. Gallo, F. Giorgi, A. Giuliani,M. R. Vari, *J Med Chem* **2000,** *43*, 3699-703.

12. S. A. Wildman,G. M. Crippen, *J Chem Inf Comput Sci* **2003,** *43*, 629-36.

13. H. P. Schultz, E. B. Schultz,T. P. Schultz, *J. Chem. Inf. Comput. Sci.* **1995,** *35*, 864 - 870.

14. J. Aires-de-Sousa,J. Gasteiger, *J Comb Chem* **2005,** *7*, 298-301.

15. E. Estrada,E. Uriarte, *Curr Med Chem* **2001,** *8*, 1573-88.

16. A. Pyka, *J. Serb. Chem. Soc.* **1997,** *62*, 251-269.

17. A. Pyka, *J. Planar Chromatogr. Mod. TLC* **1993,** *6*, 282-288.

18. A. Pyka, *J. Liq. Chromatogr. Relat. Technol.* **1999,** *22*, 41-50.

19. I. Gutman,A. Pyka, *J. Serb. Chem. Soc.* **1997,** *62*, 261-265.

20. A. B. Buda ,K. Mislow, *J Mol. Struct. (Theochem)* **1991,** *232*, 1-12.

21. H. G. Diaz, I. H. Sanchez, E. Uriarte,L. Santana, *Comput Biol Chem* **2003,** *27*, 217-27.

22. R. D. Cramer, D. E. Patterson,J. D. Bunce, *J. Am. Chem. Soc.* **1988,** *110*, 5959-5967.

23. Y. Marrero-Ponce, *Molecules* **2003,** *8*, 687-726.

24. Y. Marrero-Ponce, *J Chem Inf Comput Sci* **2004,** *44*, 2010-2026.

25. Y. Marrero-Ponce, F. Torrens, Y. J. Alvarado,R. Rotondo, *J Comput-Aided Mol Design* **2006,** *20*, 685-701.

26. G. M. Casanola-Martin, Y. Marrero-Ponce, M. T. Khan, A. Ather, S. Sultan, F. Torrens,R. Rotondo, *Bioorg Med Chem* **2007,** *15*, 1483-503.

27. Y. Marrero-Ponce, M. T. Khan, G. Casañola-Martín, A. Ather, M. N. Sultankhodzhaev, R. García-Domenech, F. Torrens,R. Rotondo, *J. Comput.-Aided Mol. Design* **2007,** *21*, 167-188.

28. Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar,E. A. Castro, *Bioorg Med Chem* **2005,** *13*, 2881-2899.

29. Y. Marrero-Ponce, R. M. Marrero, F. Torrens, Y. Martinez, M. G. Bernal, V. R. Zaldivar, E. A. Castro,R. G. Abalo, *J Mol Model* **2006,** *12*, 255-71.

30. Y. Marrero-Ponce, Y. Machado-Tugores, D. M. Pereira, J. A. Escario, A. G. Barrio, J. J. Nogal-Ruiz, C. Ochoa, V. J. Aran, A. R. Martinez-Fernandez, R. N. Sanchez, A. Montero-Torres, F. Torrens,A. Meneses-Marcel, *Curr Drug Discov Technol* **2005,** *2*, 245-65.

31. Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal, H. S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A. M. Sanchez, F. Torrens,E. A. Castro, *Bioorg Med Chem* **2005,** *13*, 1005-1020.

32. Y. Marrero-Ponce, M. A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio,F. Torrens, *Internet Electron J Mol Des* **2005,** *4* 124-150.

33. A. Montero-Torres, M. C. Vega, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J. A. Escario, V. J. Aran, A. R. Martinez-Fernandez,A. Meneses-Marcel, *Bioorg Med Chem* **2005,** *13*, 6264-75.

34. A. Montero-Torres, R. N. Garcia-Sanchez, Y. Marrero-Ponce, Y. Machado-Tugores, J. J. Nogal-Ruiz, A. R. Martinez-Fernandez, V. J. Aran, C. Ochoa, A. Meneses-Marcel,F. Torrens, *Eur J Med Chem* **2006**.

35. G. M. Casanola-Martin, M. T. Khan, Y. Marrero-Ponce, A. Ather, M. N. Sultankhodzhaev,F. Torrens, *Bioorg Med Chem Lett* **2006,** *16*, 324-30.

36. Y. Marrero Ponce, A. Meneses-Marcel, J. A. Castillo Garit, Y. Machado-Tugores, J. A. Escario, A. G. Barrio, D. M. Pereira, J. J. Nogal-Ruiz, V. J. Arán, A. R. Martínez-Fernández, F. Torrens, R. Rotondo, F. Ibarra-Velarde,Y. J. Alvarado, *Bioorg Med Chem* **2006,** *14*, 6502-6524.

37. J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens,R. García-Domenech, *J. Pharm. Sci.* **2008,** *97*, 1946-1976.

38. J. A. Castillo-Garit, Y. Marrero-Ponce, J. Escobar, F. Torrens,R. Rotondo, *Chemosphere* **2008,** *73*, 415 - 427

39. Y. Marrero-Ponce,J. A. Castillo-Garit, *J. Comput.-Aided Mol. Design* **2005,** *19*, 369-83.

40. Y. Marrero-Ponce, H. G. Díaz, V. Romero, F. Torrens,E. A. Castro, *Bioorg. Med. Chem.* **2004,** *12*, 5331-5342.

41. J. A. Castillo-Garit, Y. Marrero-Ponce,F. Torrens, *Bioorg Med Chem* **2006,** *14*, 2398-2408.

42. J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, R. García-Domenech,V. Romero-Zaldivar, *J. Comput. Chem.* **2008,** *29*, 2500 - 2512.

43. J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens,R. Rotondo, *J. Mol. Graphics Model.* **2007,** *26*, 32-47.

44. D. H. Rouvray, Academic Press: London, 1976;

45. N. Trinajstić, *Chemical Graph Theory*. CRC Press: Boca Raton, FL., 1983;

46. E. Estrada,E. Molina, *J Mol Graph Model* **2001,** *20*, 54-64.

47. E. Estrada, *J Chem Inf Comput Sci* **1995,** *35*, 31-33.

48. E. Estrada,A. Ramirez, *J Chem Inf Comput Sci* **1996,** *36*, 837-43.

49. E. Estrada, *J Chem Inf Comput Sci* **1996,** *36*, 844-49.

50. E. Estrada, N. Guevara,I. Gutman, *J Chem Inf Comput Sci* **1998,** *38*, 428-31.

51. E. Estrada, *J Chem Inf Comput Sci* **1999,** *39*, 1042-48.

52. R. Todeschini,V. Consonni, *Handbook of Molecular Descriptors*. Wiley-VCH: Germany, 2000;

53. C. H. Edwards,D. E. Penney, *Elementary Linear Algebra*. Prentice-Hall, Englewood Cliffs: New Jersey, USA 1988;

54. H. Gonzalez-Diaz, E. Tenorio, N. Castanedo, L. Santana,E. Uriarte, *Bioorg Med Chem* **2005,** *13*, 1523-30.

55. H. G. Diaz, I. Bastida, N. Castanedo, O. Nasco, E. Olazabal, A. Morales, H. S. Serrano,R. R. de Armas, *Bull Math Biol* **2004,** *66*, 1285-311.

56. H. Gonzales-Diaz, O. Gia, E. Uriarte, I. Hernadez, R. Ramos, M. Chaviano, S. Seijo, J. A. Castillo, L. Morales, L. Santana, D. Akpaloo, E. Molina, M. Cruz, L. A. Torres,M. A. Cabrera, *J Mol Model (Online)* **2003,** *9*, 395-407.

57. Y. Marrero-Ponce, M. Iyarreta-Veitia, A. Montero-Torres, C. Romero-Zaldivar, C. A. Brandt, P. E. Avila, K. Kirchgatter,Y. Machado, *J Chem Inf Model* **2005,** *45*, 1082-100.

58. A. Montero-Torres, R. N. García-Sánchez, Y. Marrero-Ponce, Y. Machado-Tugores, J. J. Nogal-Ruiz, A. R. Martínez-Fernández, V. J. Arán, C. Ochoa, A. Meneses-Marcel,F. Torrens, *Eur. J. Med. Chem.* **2006,** *41*, 483–493.

59. Y. Marrero-Ponce, R. Medina-Marrero, Y. Martinez, F. Torrens, V. Romero-Zaldivar,E. A. Castro, *J. Mol. Mod.* **2006,** *12*, 255-271.

60. Y. Marrero-Ponce, *Bioorg. Med. Chem.* **2004,** *12*, 6351-6369.

61. S. Vilar, E. Estrada, E. Uriarte, L. Santana,Y. Gutierrez, *J Chem Inf Model* **2005,** *45*, 502-14.

62. V. M. Potapov, *Stereochemistry*. Mir: Moscow, 1978;

63. R. Wang, Y. Gao,L. Lai, *Perspect. Drug Discov. Des.* **2000,** *19*, 47-66.

64. P. Ertl, B. Rohde,P. Selzer, *J Med Chem* **2000,** *43*, 3714-3717.

65. A. K. Ghose,G. M. Crippen, *J Chem Inf Comput Sci* **1987,** *27*, 21-35.

66. K. J. Millar, *J. Am. Chem. Soc.* **1990,** *112*, 8533-8542.

67. J. Gasteiger,M. A. Marsilli, *Tetrahedron Lett.* **1978,** *34*, 3181-3184.

68. L. Pauling, *The Nature of Chemical Bond*. Cornell University Press: Ithaca, NY, 1939;

69. A. Browder, *Mathematical Analysis. An Introduction.*, Springer-Verlag: New York, Inc. , 1996;

70. S. Axler, *Linear Algebra Done Right*. Springer-Verlag: New York, 1996;

71. P. D. Walker,P. G. Mezey, *J. Am. Chem. Soc.* **1993,** *115*, 12423-12430.

72. E. L. Eliel, S. Wilen,L. Mander, *Stereochemistry of Organic Compounds;*. John Wiley & Sons Inc: New York, 1994;

73. Y. Marrero-Ponce,V. Romero *TOMOCOMD software. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es*, Central University of Las Villas., 2002.

74. I. STATISTICA version. 6.0 Statsoft.

75. S. Wold,L. Erikson, in: *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, (Ed.) VCH Publishers: Weinheim, 1995; pp 309-318.

76. D. A. Belsey, E. Kuh,R. E. Welsch, *Regression Diagnostics*. Wiley: New York, 1980;

77. A. Golbraikh,A. Tropsha, *J Mol Graph Model* **2002,** *20*, 269-76.

78. P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen,H. Nielsen, *Bioinformatics* **2000,** *16*, 412-424.

79. E. A. Coats, *In 3D QSAR in Drug Design.*, Kluwer/ESCOM: Dordrecht, 1998;

80. B. D. Silverman, *Quant. Struct.-Act. Relat* **2000,** *19*, 237-246.

81. M. Wagener, J. Sadowski,J. Gasteiger, *J. Am. Chem. Soc.* **1995,** *117*, 7769-7775.

82. S. S. So,M. Karplus, *J Med Chem* **1997,** *40*, 4347-4359.

83. M. Lobato, L. Amat, E. Besalu,R. Carbo-Dorca, *Quant. Struct.-Act. Relat* **1997,** *16*, 465-472.

84. E. Besalu, X. Girones, L. Amat,R. Carbo-Dorca, *Acc Chem Res* **2002,** *35*, 289-95.

85. M. F. Parretti, R. T. Kroemer, J. H. Rothman,W. G. Richards, *J Comput Chem* **1997,** *18*, 1334-1353.

86. G. Klebe, U. Abraham,T. Mietzner, *J. Med. Chem.* **1994,** *37*, 4130-4146.

87. D. Robert, L. Amat,R. Carbo-Dorca, *J Chem Inf Comput Sci* **1999,** *39*, 333-44.

88. N. Stiefl,K. Baumann, *J Med Chem* **2003,** *46*, 1390-407.

89. H. Chen, J. Zhou,G. Xie, *J. Chem. Inf. Comp. Sci.* **1998,** *38*, 243-250.

90. E. A. Coats, *Persp. Drug Disc. Des* **1998,** *12-14*, 199-213.

91. M. T. Cronin,T. W. Schultz, *J. Mol. Struct. (Theochem).* **2003,** *622*, 39-51.

92. M. Vicent, B. Marchand, G. Rémond, S. Jaquelin-Guinamant, G. Damien, B. Portevin, J. Baumal, J. Volland, J. Bouchet, P. Lambert, B. Serkiz, W. Luitjen, M. Lauibie,P. Schiavi, *Drug Des. Discov* **1992,** *9*, 11.

93. S. S. Liu, C. S. Yin,L. S. Wang, *J Chem Inf Comput Sci* **2002,** *42*, 749-56.

94. H. H. Maw,L. H. Hall, *J Chem Inf Comput Sci* **2001,** *41*, 1248-54.

95. B. D. Silverman,D. E. Platt, *J Med Chem* **1996,** *39*, 2129-40.