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

[G004]

Bond-Based 3D-Chiral Linear Indices: Theory and QSAR Applications to Central Chirality Codification

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

[a]*Applied Chemistry Research Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.* ***e-mail:*** *jacgarit@yahoo.es, juancq@uclv.edu.cu or juancq.22@gmail.com*
[b]*Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Department of Pharmacy, Faculty of Chemistry-Pharmacy and Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.*
[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*

Abstract

The recently introduced non-stochastic and stochastic bond-based linear indices are been generalized to codify chemical structure information for chiral drugs, making use of a trigonometric 3D-chirality correction factor. These improved modified descriptors are applied to several well-known data sets in order to validate each one of them. 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 such as CoMFA, MQSM, CoMMA, E-state, and so on. For that reason, 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 that 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 linear indices appear to provide a very interesting alternative to other more common 3D-QSAR descriptors.

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

Introduction

Chirality is a fundamental aspect of molecular structure, which sometimes has profound influence on the properties of compounds. The non-superimposable mirror-image isomers are called enantiomers, but may also be referred to as enantiomorphs, optical isomers or optical antipodes.[1] Enantiomers quite often exhibit different chemical and physical properties as well as different biological activities.[2] Therefore, enantiomers of a given compound have identical chemical properties with regard to their reaction with non-chiral reagents, although they will give products with different configurations and they may show differences in behavior (both in reaction rates and in product stereochemistry) in their interactions with a chiral reagent. Furthermore, many biochemical processes and phenomena are stereospecific. For instance, L- and D-enantiomers of amino acids have different tastes,[3,4] enantiomers of some compounds have different odors,[5,6] and many medicinal preparations have physiological properties different from those of their enantiomers.[7-9] The case of thalidomide is an example of a problem that was, at least, complicated by the ignorance of stereochemical effects.[10] Thus, whenever a drug is to be 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 this connection, the regulation of the Food & Drug Administration (FDA) requires a detailed study of both enantiomers.[11]

Attempts to give quantitative meaning to molecular chirality can be dated almost as far back as van't Hoff and LeBel's proposition to extend the structural formulas of chemistry into three-dimencional (3D) space. In 1890 Guye introduced the first function designed to correlate a pseudoscalar property, *i.e.*, optical rotation, with the molecular structure of chiroids-the first example of a chirality function in chemistry.[12] Chirality, however, is an inherent molecular property that depends on only symmetry and that is independent of its physical and chemical manifestations. It should be possible, therefore, to quantify chirality, *i.e.,* to construct a chirality measure, without reference to any experimental data.

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.[1,2,13-21] Nevertheless, very 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.[22] Among the chiral topological indices (CTIs) published in the literature, Estrada and Uriarte mentioned some in a recent

review about topological indices (TIs).[22] Those derived by Pyka,[23-25] as well as Gutman and Pyka[26] have rationalized some of these indices from a mathematical point of view. The relationships between these indices and the Wiener index have been established. Moreover, Schultz et al.[20] 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.[27] 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'. These methods yield a single real value, usually an absolute quantity that is the same for both enantiomorphs. Recently, Benigni *et al.* proposed a chirality measure for molecules in a data set.[17] This measure is based on the comparison of the 3D structure for a molecule with all the others in a data set, in terms of electrostatic potential and shape indices. Moreau described a quantitative measure of the chirality of the environment around each atom.[28] However, applications of quantitative measures of chirality to the prediction of experimental observables have been quite limited.

A different idea was to incorporate R/S labels into conventional topological indices (TIs).[20] Derived chirality descriptors were correlated with biological activity by de Julián-Ortiz *et al.*,[16] Golbraikh *et al.*[18] and more recently by Díaz *et al*.[29] One of the first approaches to this field was introduced by de Julián-Ortiz *et al.*[16] in a study of the pharmacological activity of different pairs of enantiomers on the σ-receptor. Fortunately, the so-called CTIs are inexpensive in terms of computational time in comparison to grid dependent methods as CoMFA.[30] In any case, when chirality is considered many 3D-TIs become 'hard to interpret' in physical terms. For example, Golbraikh, Bonchev, and Tropsha's (GBT) work generated even complex numbers, which are incompatible with standard statistical software.[18]

In addition to CTIs, the characterization of symmetry, and specifically chiral structural features in computer-aided drug design (CADD), has become possible only since the development of 3D-QSAR (Quantitative Structure Activity Relationship) methods. Among these methods, special mention should be made of the use of CoMFA.[30] Evidently, chirality in CoMFA (Comparative Molecular Fields Analysis) is taken into account by default, since the 3D-field values of chiral isomers are different. Despite its wide popularity, CoMFA is not always applicable, especially in situations where compounds under investigation are highly flexible.

Even although most these difficulties are solved by Grid (a CoMFA-like last generation method), several drawbacks still remain when large data should be processed.[31]

In fact, recently one of the present authors, Y.M-P, has introduced a new set of atom-level molecular descriptors (MDs) of relevance to QSAR/QSPR (Quantitative Structure Activity/Property Relationship) studies and 'rational' drug design, non-stochastic and stochastic linear indices [$f_k(\bar{x})$ and ${}^s f_k(\bar{x})$, respectively]. These atom-level MDs have also been useful for the selection of novel subsystems of compounds having a desired property/activity.[32-35] In this sense, the method was successfully applied to the identification of new tyrosinase inhibitors[32] and virtual (computational) screening of novel anthelmintic compounds, which were then synthesized and in vivo evaluated on hepatic Fasciola.[35] Studies for the fast-track discovery of novel antibacterial,[34] antimalarial[36] and trichomonicidal,[33] lead-like chemicals were also conducted with this theoretical approach. In addition, the atom-based linear indices have been successfully employed in QSAR and in silico pharmacokinetic studies of Caco-2 permeability of drugs.[37] This approach has also been extended to consider three-dimensional features of small/medium-sized molecules based on the trigonometric-3D-chirality correction factor approach.[38]

More recently, a new set of non-stochastic and stochastic bond-based linear indices has been defined and applied to QSPR/QSAR studies.[39,40] The present report is written with two objectives in mind: 1) to extend the non-stochastic and stochastic bond-based 2D linear indices in order to codify chirality-related structural features and 2) to compare the results achieved with them with those obtained with other methods. The problems of classification of ACE (***A***ngiotesin-***C***onverting ***E***nzime) inhibitors, the prediction of σ-receptor antagonist activities and corticosteroid-binding globulin binding affinity of the Cramer's steroid data set are selected as an illustrative example of method applications. These examples will be used as a matter of comparison with other CTIs, 3D and quantum chemical descriptors as well.

Theoretical Framework

The basis of the extension of linear indices that will be given here is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature,[41,42] and rediscovered by Estrada as an important source of new molecular descriptors.[43-48] In this section, we first will define the nomenclature to be used in this report, then the atom-based molecular vector (**X**) will be redefined for bond characterization using the same approach as previously reported,

and finally some new definition of bond-based non-stochastic and stochastic linear indices will be given.

Background in Edge-Adjacency Matrix

Let $G = (V, E)$ be a simple graph, with $V = \{v_1, v_2, ..., v_n\}$ and $E = \{e_1, e_2, ...e_m\}$ being the vertex- and edge-sets of G, respectively. Then G represents a molecular graph having n vertices and m edges (bonds). The edge-adjacency matrix **E** of G (likewise called bond-adjacency matrix, **B**) is a square, symmetric matrix whose elements e_{ij} are "1" if and only if edge i is adjacent to edge j, otherwise "0".[45,48,49] Two edges are adjacent if they are incidental to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the i^{th} row (or column) of **E** is named the edge-degree of bond i, $\delta(e_i)$.[43,46,47,49]

Edge-Relations: Stochastic Edge-Adjacency Matrix

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

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

where, ${}^k e_{ij}$ are the elements of the k^{th} power of **E** and the SUM of the i^{th} row of $\mathbf{E}^k$ are named the k-order edge degree of bond i, ${}^k\delta(e)_i$. Note that the matrix $\mathbf{ES}^k$ in Eq. **1** has the property that *the sum of the elements in each row* is "1". An mxm matrix with nonnegative entries having this property is called a "*stochastic matrix*".[50-55]

Structural Representation Across of the Bond-Based Molecular Vector

The atom-based molecular vector (**X**) used to represent small-to-medium size organic chemicals has been explained elsewhere in some detail.[34,35,56,57] In a manner parallel to the development of **X**, we present the expansion of the bond-based molecular vector ($\overline{w}$). The components (w) of $\overline{w}$ are numerical values, which represent a certain standard bond property (bond label). That is to say, these weights correspond to different bond properties for organic molecules. Thus, a molecule having 5, 10, 15,..., m bonds can be represented by means of vectors, with 5, 10, 15,..., m components, belonging to the spaces $\Re^5$, $\Re^{10}$, $\Re^{15}$,...,$\Re^m$, respectively, where m is the dimension of the real sets ($\Re^m$). This approach allows us encoding

organic molecules such as 2-hydroxybut-2-enenitrile through the molecular vector $\overline{w}$ = [$w_{Csp3\text{-}Csp2}$, $w_{Csp2=Csp2}$, $w_{Csp2\text{-}Osp3}$, $w_{H\text{-}Osp3}$, $w_{Csp2\text{-}Csp}$, $w_{Csp\equiv Nsp}$]. 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) 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,[58-61] standard bond dipole[58-61] or even mathematical expressions involving atomic weights such as atomic Log P,[62] surface contributions of polar atoms,[63] atomic molar refractivity,[64] atomic hybrid polarizabilities,[65] and Gasteiger-Marsilli atomic charges,[66] atomic electronegativities in Pauling scale[67] and so on. Here, we characterized each bond with the following parameter:

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

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

Calculation of Linear Indices for Bonds, Bond-Type and the Whole Molecule

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

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

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

where m is the number of bonds in the molecule, and w_j 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.[50,68,69] For that reason, those coordinates can be considered as weights (bond labels) of the edge of the molecular graph. The coefficients ${}^{k}e_{ij}$ and ${}^{k}es_{ij}$ are the elements of the k^{th} power of the matrices **E**(G) and **ES**(G), correspondingly, of the molecular graph. The defining equations (**3**) and (**4**) for $\boldsymbol{f_k}(\overline{w})$ and ${}^{\mathbf{s}}\boldsymbol{f_k}(\overline{w})$, respectively, can also be written in matrix form, where [W] is a column

vector (an *m*x1 matrix) of the coordinates of $\overline{w}$ in the canonical basis set of $\Re^m$. Here, $\mathbf{E}^k$ and $\mathbf{ES}^k$ denote the matrices of linear maps with respect to the natural basis set.

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

$$f(\lambda_1 W_1 + \lambda_2 W_2) = \lambda_1 f(W_1) + \lambda_2 f(W_2) \quad (5)$$

for any scalars λ_1,λ_2 and any vectors W_1, W_2 in $\Re^m$.

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

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

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

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

Finally, in addition to total and bond linear indices computed for each bond in the molecule, a local-fragment (bond-type) formalism can be developed. The k^{th} bond-type linear index of the edge-adjacency matrix is calculated by adding the k^{th} bond linear indices of all bonds of the same bond type in the molecule. That is to say, this extension of the bond linear index is similar to the group additive schemes, in which an index appears for each bond type in the molecule together with its contribution based on the bond linear index. Consequently, if a molecule is partitioned into Z molecular fragments, the total non-stochastic [or stochastic] linear indices can be partitioned into Z local non-stochastic [or stochastic] linear indices $f_{kL}(\overline{w})$ [or ${}^s f_{kL}(\overline{w})$], L = 1, …, Z. Therefore, the total (both non-stochastic and stochastic) linear indices of order *k* can be expressed as the sum of the local linear indices of the Z fragments of the same order:

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

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

In the bond-type linear indices formalism, each bond in the molecule is classified into a bond-type (fragment). In this sense, bonds may be classified into bond types in terms of the characteristics of the two atoms that define the bond. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the k^{th} fragment (bond-type) linear indices provide much useful information. Thus, the development of the bond-type linear indices description provides the basis for application to a wider range of biological problems in which the local formalism is applicable without the need for superposition of a closely related set of structures. A detailed example of the calculus of the bond-based linear indices can be seen in previous publications.[39,40]

3D-Chiral non-stochastic and stochastic bond-based linear indices.

The total and local bon-based linear 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:

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

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:[38,70]

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.[71] This correction has principally a mathematical meaning and must not be the source of any misunderstanding.

A severe limitation of the GBT[18] approach is the existence of different chirality corrections, and we had great difficulty in selecting one of these. Therefore, the present *trigonometric 3D-chiral correction factor* is invariant with respect 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 this Table, it is clearly shown that the *trigonometric 3D-chirality factor* is invariant with respect to the selection of all possible real scales. Moreover, the factor gets ever the values 1, 0 and -1 for R, non-chiral and S atoms. As outlined above the demonstration of invariance for this factor with respect 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.[16,18,29]

Table 1. Values of *Trigonometric 3D-Chirality Correction Factor* {sin[(ω_A+4Δ)π/2]} within the Allowed Domain.

	Δ														
ω_A	***-7***	***-6***	***-5***	***-4***	***-3***	***-2***	***-1***	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***
$\omega_{R=}$ 1	1		1		1		1		1		1		1		1
$\omega_{non\text{-}chiral=}$ 0		0		0		0		0		0		0		0	
$\omega_{S=}$ -1	-1		-1		-1		-1		-1		-1		-1		-1

A very interesting point is that the present 3D-chiral descriptor reduces to simple (2D) bond-based linear indices for molecules without specific 3D characteristics because sin[(0+4Δ)π/2] = 0, being Δ zero or any even number. Therefore, when all the atoms in the molecule are not chiral, the bond-based linear indices or any GBT-like chiral TI do not change upon the introduction of this factor. This means that, for example $^{*}\overline{w} = \overline{w}$ and thus, $^{*}f_k(\overline{w}) = f_k(\overline{w})$.

Methods

TOMOCOMD-CARDD approach

Molecular fingerprints were generated by means of the interactive program for molecular design and bioinformatic research TOMOCOMD.[72] 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 linear 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:

Step 1. Draw the molecular structure for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol belonging to the different groups in the periodic table of the elements.

Step 2. Use appropriate weights in order to differentiate the atoms in the molecule. The weights used in this work are those previously proposed for the calculation of the DRAGON descriptors,[73-75] that is, atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (G).[67] The values of these atomic labels are shown in Table 2.[67,73-75]

Step 3. Compute the total and local (bond, group and bond-type) non-stochastic and stochastic linear 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) linear maps.

Step 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 linear fingerprints having, for instance, the following appearance,

$$\mathbf{A} = a_0 {}^{*}\boldsymbol{f}_0(\overline{w}) + a_1 {}^{*}\boldsymbol{f}_1(\overline{w}) + a_2 {}^{*}\boldsymbol{f}_2(\overline{w}) + \ldots + a_k {}^{*}\boldsymbol{f}_k(\overline{w}) + \mathrm{c} \qquad (\mathbf{12})$$

where **A** is the measured activity, ${}^{*}\boldsymbol{f}_k(\overline{w})$ are the k^{th} bond-based 3D-chiral linear indices, and the $\boldsymbol{a}_k$*'s* and c are the coefficients obtained by the MRA;

Step 5. Test the robustness and predictive power of the QSPR/QSAR equation by using internal (cross-validation) and external validation techniques.

Table 2. Values of the Atomic Weights Used for ***TOMOCOMD-CARDD*** MDs.[67,73-75]

ID	Atomic Mass	VdW[a] Volume (Å^3)	Polarizability (Å^3)	Mulliken Electronegativity	Pauling Electronegativity
H	1.01	6.709	0.667	2.592	2.20
B	10.81	17.875	3.030	2.275	2.04
C	12.01	22.449	1.760	2.746	2.55
N	14.01	15.599	1.100	3.194	3.04
O	16.00	11.494	0.802	3.654	3.44
F	19.00	9.203	0.557	4.000	3.98
S	32.07	24.429	2.900	2.957	2.58
Cl	35.45	23.228	2.180	3.475	3.16
Br	79.90	31.059	3.050	3.219	2.96
I	126.90	38.792	5.350	2.778	2.66

[a]VdW: van der Waals

The bond–based *TOMOCOMD-CARDD* descriptors computed in this study were the following:

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

Chemometric analysis

Statistical analysis was carried out with the STATISTICA software.[76] 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 (CBG) binding 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.[77,78] Therefore, the quality of the models was determined by examining the determination coefficients (also know as square correlation coefficient; 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*.[77,79]

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.[80] 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,[80] and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages.

Results and Discussion

With the objective of assessing the efficacy of bond-based 3D-chiral linear 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 several QSAR studies in recent years. Now we are going to discuss the use of the bond-based 3D-chiral linear indices descriptors in each one of these well-known series of compounds and a comparison with other previously reported approaches will be also developed.

Prediction of the Corticosteroid-Binding Globulin Binding Affinity of a Steroid Family.

The first molecular set used in our study is made up of 31 steroids for which the binding affinity to the corticosteroid-binding globulin was measured. We select this molecular data set because it is well known to QSAR researchers; the so-called Cramer's steroid database was introduced by Cramer *et al.*[30] in 1988 using Comparative Molecular Field Analysis (CoMFA) method and since then it has become a benchmark for the assessment of novel QSAR methods.[81,82] 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.[83] Even though this data set is not the ideal 3D benchmark data set[83] it was used for the shake of comparability.[84] We use this molecular set, because all compounds in this data set contain chiral atoms, and binding affinities of these compounds are available.[30] Due to the fact that the studied steroid molecular structures have been already depicted in several publications,[30,82] they will not be included here.

Different methods have been used to develop 3D-QSAR models for this data set, including CoMFA,[30] Comparative Molecular Similarity Indices Analysis (CoMSIA),[85] Molecular Quantum Similarity Measures (MQSM),[86] Topological Quantum Similarity Indices (TQSI),[87] Comparative Molecular Moment Analysis (CoMMA),[82] Mapping Property Distributions of Molecular Surfaces (MAP),[84] and so on.[88-91] Table 3 gathers the studied entire set with the experimental binding affinities, taken from Robert et al.[86] The obtained models (concentrations in mol^{-1}) are given below, together with their statistical parameters:

$$\mathbf{CBG} = -10.30(\pm2.35) + 18.22(\pm2.33)\,{}^{*G}\boldsymbol{f}_{0L}(\overline{w}_E) - 0.79(\pm0.16)\,{}^{*G}\boldsymbol{f}_{0L}{}^{H}(\overline{w}_E) - 17.88(\pm2.25)\,{}^{*K}\boldsymbol{f}_{0}(\overline{w}) + 0.07(\pm0.01)\,{}^{*M}\boldsymbol{f}_{1L}{}^{H}(\overline{w}_E) + 0.041(\pm0.006)\,{}^{*V}\boldsymbol{f}_{2}{}^{H}(\overline{w}) - 6.1\times10^{-3}(\pm0.98\times10^{-3})\,{}^{*V}\boldsymbol{f}_{3}{}^{H}(\overline{w}) \quad (\mathbf{13})$$

$N = 31 \quad R = 0.933 \quad R^2 = 0.870 \quad F = 26.76 \quad s = 0.436 \quad q^2 = 0.787 \quad s_{cv} = 0.499 \quad p < 0.0001$

$$\mathbf{CBG} = 8.20(\pm1.92) - 1.30(\pm0.24)\,{}^{*Vs}\boldsymbol{f}_{5}(\overline{w}) + 1.81(\pm0.50)\,{}^{*Vs}\boldsymbol{f}_{4}(\overline{w}) + 4.9\times10^{-2}(\pm1.4\times10^{-2})\,{}^{*Ms}\boldsymbol{f}_{7}{}^{H}(\overline{w}) + 9.3\times10^{-2}(\pm1.4\times10^{-2})\,{}^{*Vs}\boldsymbol{f}_{1L}(\overline{w}_E) - 0.59(\pm0.18)\,{}^{*Gs}\boldsymbol{f}_{5L}(\overline{w}_E) - 0.58(\pm0.28)\,{}^{*Vs}\boldsymbol{f}_{3}(\overline{w}) \quad (\mathbf{14})$$

$N = 31 \quad R = 0.954 \quad R^2 = 0.911 \quad F = 40.70 \quad s = 0.361 \quad q^2 = 0.869 \quad s_{cv} = 0.39 \quad p < 0.0001$

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. Both models have six variables, the non-stochastic model (Eq. **13**) explains

the 87% of the variance of the experimental CBG values, while the stochastic model (Eq. **14**) explains more than the 91% of the experimental variance. The predicted values for this data set using non-stochastic and stochastic bond-based 3D-chiral linear indices are also shown in Table 3. The values of the standard deviation for these models were small s = 0.436 and s = 0.361 for non-stochastic and stochastic models, respectively.

Table 3. Results of the Steroids Data Set Used for QSAR Study.

		Observed CBG affinity (pKa)[a]	Bond-based non-stochastic 3D-chiral linear indices			Bond-based stochastic 3D-chiral linear indices		
			Predicted value	%E[b]	$\%E_{cv}$	Predicted value	%E	$\%E_{cv}$
1	Aldosterone	-6.279	-5,874	6,450	9,722	-6,315	0,571	1,196
2	Androstanediol	-5.000	-5,644	12,876	21,566	-5,119	2,384	2,796
3	Androstenediol	-5.000	-4,811	3,786	4,861	-4,724	5,524	6,564
4	Androstenedione	-5.763	-6,322	9,696	11,559	-6,563	13,883	16,809
5	Androsterone	-5.613	-5,816	3,613	4,140	-5,332	4,999	6,352
6	Corticosterone	-7.881	-7,863	0,229	0,277	-7,383	6,314	7,950
7	Cortisol	-7.881	-7,729	1,922	2,206	-7,445	5,538	6,213
8	Cortisone	-6.892	-7,517	9,064	10,515	-7,482	8,560	9,682
9	Dehydroepiandrosterone	-5.000	-5,110	2,190	3,167	-4,852	2,953	3,386
10	Deoxycorticosterone	-7.653	-7,616	0,488	0,566	-7,900	3,233	3,936
11	Deoxycortisol	-7.881	-7,530	4,449	5,393	-7,956	0,952	1,328
12	Dihydrtestosterone	-5.919	-5,364	9,378	13,509	-5,986	1,127	1,198
13	Estradiol	-5.000	-5,168	3,368	4,393	-4,865	2,703	3,543
14	Estriol	-5.000	-4,717	5,658	9,409	-5,188	3,764	5,008
15	Estrone	-5.000	-5,125	2,510	3,694	-4,929	1,411	1,908
16	Ethiocholanolone	-5.255	-5,468	4,061	6,261	-5,193	1,173	1,410
17	Pregnenolone	-5.255	-5,328	1,397	2,298	-5,549	5,595	7,053
18	17-Hydroxyregnenolone	-5.000	-5,233	4,660	6,354	-5,402	8,043	9,927
19	Progesterone	-7.380	-6,875	6,838	7,868	-7,174	2,793	3,214
20	17-Hydroxyprogesterone	-7.740	-6,787	12,313	14,376	-7,014	9,384	10,491
21	Testosterone	-6.724	-6,365	5,343	6,552	-6,356	5,466	6,212
22	Prednisolone	-7.512	-7,208	4,052	4,552	-7,267	3,256	3,885
23	Cortisolacetate	-7.553	-7,456	1,283	1,969	-7,520	0,440	2,319
24	4-Pregnene-3,11,20-trione	-6.779	-6,913	1,975	2,306	-6,692	1,281	1,501
25	Epicorticosterone	-7.200	-7,850	9,031	10,139	-7,522	4,468	5,420
26	19-Nortestosterone	-6.144	-6,077	1,089	1,481	-6,105	0,640	0,885
27	16a,17a-Dihydroxyprogesterone	-6.247	-6,950	11,246	12,532	-6,562	5,046	5,803
28	16a-Methylprogesterone	-7.120	-6,951	2,370	3,054	-7,245	1,759	2,238
29	19-Norprogesterone	-6.817	-6,636	2,656	2,981	-6,958	2,061	3,386
30	2a-Methylcortisol	-7.688	-7,573	1,492	1,727	-7,394	3,823	4,775
31	2a-Methyl-9a-fluorocortisol	-5.797	-6,096	5,165	8,814	-5,980	3,149	6,727

[a]Observed CBG affinity values taken from Robert *et al*. concentration are expressed in nM.[86] [b]E: Error. [c]E_{cv}: Error of LOO validation.

An important aspect of QSAR modeling is the development of a way to perform 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 leave-

one-out (LOO) statistics have been used as a means of demonstrating predictive capability. These models showed cross-validation square correlation coefficients of 0.787 and 0.869, respectively. Both values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models.[77-79]

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 works are summarized in Table 4, where the results were arranged in decreasing order of q^2 and the comparison can be easily carried out. However, it should be remarked 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 statistical methods such as partial least squares, principal component 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 mechanical approaches.

Table 4. Comparison Between Prediction for the Steroid Data Set with Bond-Based 3D-Chiral Linear Indices and other 3D-QSAR Approaches.

QSAR Method	N	n	Statistical Method	q^2	Ref.
Similarity matrices-based molecular descriptors	31	6	genetic NN	0.940	90
Bond-based stochastic 3D-chiral linear indices	31	6	MLR	0.869	Eq 14
TQSAR	31	6	MLR after PCA	0.842	86
Atom-based 3D-chiral linear indices (*stochastic*)	31	7	MLR	0.788	38
Bond-based non-stochastic 3D-chiral linear indices	31	6	MLR	0.786	Eq 13
MEDV	31	5	GA and RLM	0.777	96
TQSI	31	3	MLR	0.775	87
Atom-based 3D-chiral linear indices (*non-stochastic*)	31	6	MLR	0.767	38
MEDV	31	6	GA and RLM	0.765	96
Similarity indices	31	1	PLS	0.734	89
E-State and kappa shape index	31	4	MLR*	0.730	97
MQSM	31	4	MLR after PLS	0.727	98
E-State and kappa shape index	31	4	MLR	0.720	97
MQMS	31	3	MLR and PCA	0.705	87
CoMMA	31	6	PCR	0.689	99
MEDV	31	4	GA and RLM	0.648	96
Wagener's	31	-	k-NN and FNN	0.630	88

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**

Modeling σ-receptor antagonist activities of 3-(3-hydroxyphenyl)piperidines

After the results obtained in the previous section, a short data set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl) piperidines that bind to σ-receptors are also selected as an

illustrative example of application of the bond-based 3D-chiral linear indices. This data set was introduced in QSAR studies by de Julian-Ortiz *et al.*[16] in 1998, and since then, it has been repeatedly used by some authors[29,38,92,93] in recent years, to validate new CTIs. The σ-receptors mediate severe side effects induced by various dopamine antagonists.[16]

Bond-based 3D-chiral linear indices are asymmetric and reduce to classical descriptors when symmetry is not codified. Besides, Gónzalez-Díaz *et al*. concluded that σ-receptor antagonist activity is not a pseudoscalar property[29] and we can expect at least a good correlation with bond-based 3D-chiral linear indices.

This experiment also permitted us to compare our method with other previously reported approaches. The multiple linear regression (MLR) analysis is used to develop QSAR models for the σ-receptor antagonist activities. The models obtained using non-stochastic and stochastic bond-based 3D-chiral linear indices for the σ-receptor antagonistic activities are given below:

$$\mathbf{logIC_{50}(\sigma)} = -5.84(\pm 0.45) + 2.8\text{x}10^{-4}(\pm 0.24\text{x}10^{-4})\,{}^{*G}\boldsymbol{f}_5^{H}(\overline{w}) + 1.28\text{x}10^{-8}(\pm 3.5\text{x}10^{-9})\,{}^{*P}\boldsymbol{f}_{15}(\overline{w}) \quad (\mathbf{15})$$

N = 14 $R^2 = 0.953$ $q^2_{LOO} = 0.892$ F(2, 11) = 111.18 $s = 0.239$ $s_{cv} = 0.333$ $p < 0.0001$

$$\mathbf{logIC_{50}(\sigma)} = -6.52(\pm 0.411) + 1.181(\pm 0.144)\,{}^{*\,Gs}\boldsymbol{f}_7(\overline{w}) - 0.912(\pm 0.129)\,{}^{*Ks}\boldsymbol{f}_{15}(\overline{w}) \quad (\mathbf{16})$$

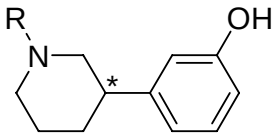
N = 14 $R^2 = 0.967$ $q^2_{LOO} = 0.956$ F(2, 11) = 161.80 $s = 0.200$ $s_{cv} = 0.211$ $p < 0.0001$

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 these models are appropriate for the description of the chemicals studied here. In Table 5, the structure and values of experimental[16,29] and calculated Log IC_{50} are show for this data set.

These QSAR models use two variables and explain more than the 95 % and 96% of the experimental values of log IC_{50} and show low values of standard deviation 0.239 and 0.200, for bond-based non-stochastic and stochastic 3D-chiral linear indices models, correspondingly.

The comparison with the results of other methods previously reported for the same activity is shown in Table 6. As can be seen, our models' results have statistical parameters better than the ones of models obtained with MARCH-INSIDE molecular descriptors,[29] other chiral TIs[16] and atom-based 3D-chiral linear indices.[38] Predictability and stability (robustness) of the

Table 5. Results of Multivariate Regression Analysis of the log IC_{50} of a Group of *n*-Alkylated 3-(3-Hidroxyphenyl)Piperidines for the σ-Receptor.



Compound (Alkyl group)[a]	Log IC_{50} (σ-receptor)								
	Obs.[b]	Cal.[c]	Res.[d]	Cal.[e]	Res.[d]	Cal.[f]	Res.[d]	Cal.[g]	Res.[d]
(R)-3-HPP									
H	-0.66	-0.54	-0.12	-0.48	-0.18	-0,38	-0,28	-0,43	-0,23
CH_3	0.43	0.13	0.30	0.28	0.15	0,06	0,37	0,28	0,15
C_2H_5	0.95	0.72	0.23	0.70	0.25	0,69	0,26	0,75	0,20
n-C_3H_7	1.52	1.32	0.20	1.45	0.07	1,38	0,14	1,34	0,18
i-C_3H_7	0.61	1.30	-0.69	0.84	-0.23	1,00	-0,39	1,08	-0,47
n-C_4H_9	2.05	1.93	0.12	1.89	0.16	2,09	-0,04	2,06	-0,01
2-Phenylethyl	2.10	2.22	-0.12	2.41	-0.31	2,23	-0,13	2,04	0,06
(S)-3-HPP									
H	-1.19	-1.09	-0.10	-0.80	-0.39	-0,98	-0,21	-1,23	0,04
CH_3	-0.28	-0.42	0.14	-0.56	0.28	-0,50	0,22	-0,38	0,10
C_2H_5	-0.01	0.17	-0.18	0.19	-0.20	0,14	-0,15	0,13	-0,14
n-C_3H_7	0.81	0.77	0.04	0.57	0.24	0,83	-0,02	0,78	0,03
i-C_3H_7	0.68	0.75	-0.07	0.62	0.06	0,51	0,17	0,52	0,16
n-C_4H_9	1.51	1.37	0.14	1.18	0.33	1,54	-0,03	1,60	-0,09
2-Phenylethyl	1.80	1.67	0.13	2.03	-0.23	1,70	0,10	1,80	0,00

[a]Alkylic (R) group at nitrogen ring. [b]Observed values of the Log IC_{50}, C in nM, for the σ-receptor taken from the literature.[16,29] [c]Values calculated using atom-based non-stochastic linear indices.[38] [d]Residual, defined as [Log IC_{50} (σ)Obs – Log IC_{50} (σ)Cal]. [e]Values calculated using atom-based stochastic linear indices.[38] [f]Values calculated using bond-based non-stochastic linear indices (Eq. 15). [g]Values calculated using bond-based stochastic linear indices (Eq. 16).
Abbreviations: HPP, *N*-alkylated 3-hydroxyphenyl piperidines.

Table 6. Statistical Parameters of the QSAR Models Obtained Using Bond-Based 3D-Chiral Linear Indices to Predict the σ-Receptor Antagonist Activity of 14 N-Alkylated 3-Hydroxyphenyl Piperidines.

index	*N*	n	R^2	*s*	q^2	s_{cv}	F
Bond-based stochastic linear indices (Eq. 16)	14	2	0.967	0.200	0.957	0.211	161.80
Bond-based non-stochastic linear indices (Eq. 15)	14	2	0.953	0.239	0.892	0.333	111.18
Atom-based Stochastic linear indices.[38]	14	2	0.941	0.267	0.900	0.319	87.93
Atom-based non-Stochastic linear indices.[38]	14	2	0.939	0.271	0.909	0.305	84.87
Chiral TIs[16]	14	3	0.931	0.301	*	*	45.70
MARCH-INSIDE molecular descriptors[29]	14	2	0.922	0.295	*	0.32	71.17

*Values are not reported in the literature.

obtained models with regard to data variation was carried out here by means of LOO cross-validation. The models showed values of cross-validation determination coefficient (q^2) of 0.892 and 0.957, when non-stochastic and stochastic bond-based linear 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.[77,78,94] Unfortunately, the authors of previous works, Diaz *et al*. [29] and de Julian de Ortiz *et al.* [16] 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 linear indices is the best QSAR model to describe the property studied in this section.

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

Finally, in order to validate even more the bond-based 3D-chiral linear indices in QSAR studies, a recently introduced data set of 32 perindoprilate stereoisomers, angiotensin-converting enzyme (ACE) inhibitors[95] 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).[95] 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:

$$\mathbf{ACEiactv} = 71.191 + 24.29\text{x}10^{-2*G}\boldsymbol{f}_3(\overline{w}) - 5.95\text{x}10^{-2*K}\boldsymbol{f}_{5L}(\overline{w}_E) \quad (\mathbf{17})$$

$N = 23 \quad \lambda = 0.401 \quad D^2 = 7.069 \quad F(2, 20) = 14.93 \quad p < 0.0001$

$$\mathbf{ACEiactv} = 2953.39 + 7.367^{*Ps}\boldsymbol{f}_3(\overline{w}) - 20.504^{*Vs}\boldsymbol{f}_{kL}(\overline{w}_E) \quad (\mathbf{18})$$

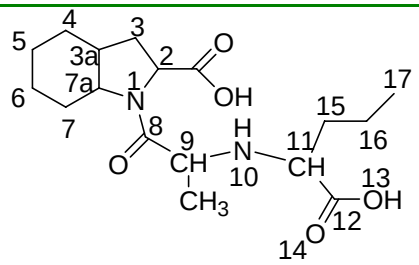
$N = 23 \quad \lambda = 0.392 \quad D^2 = 7.348 \quad F(2, 20) = 15.52 \quad p < 0.0001$

Model (**17**), which includes non-stochastic bond-based linear indices, classifies correctly 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. The most important criterion for the acceptance or not of a discriminant model is based on the statistics for external prediction sets. Therefore, model (**17**) correctly classifies 100.00% of the compounds in the test set; this model also showed a maximal Matthews' correlation coefficient (MCC) of 1.000 for the external validation set. In Table 7, we give the basic structure of perindoprilate stereoisomers and their classification in the training and prediction sets together with their posterior probabilities calculated from the Mahalanobis distance.

On the other hand, model (18), which includes stochastic bond-based linear indices, has an accuracy of 100% for the training set. This model showed a high Matthews' correlation coefficient (MCC) of 1. 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 a maximal accuracy of 100%, yielding a maximal MCC of 1.000 for the test set.

Table 7. 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]
				Eq. 17 Non-stochastic bond-based linear indices		Eq. 18 Stochastic bond-based linear indices	
				Active Compounds			
1	SSRSS*	+	1.1	+	0,643	+	0,816
2	SRSSS*	+	1.2	+	0,985	+	0,994
3	*SSSSS*	+	1.5	+	0,929	+	0,968
4	SRRSS*	+	3.3	+	0,916	+	0,964
5	*SSSSR*	+	12.2	+	0,931	+	0,945
6	*SSRSR*	+	29.4	+	0,649	+	0,701
7	*SRRSR*	+	39.8	+	0,918	+	0,938
8	*SRSSR*	+	54	+	0,986	+	0,990
9	*RRSSS*	+	108	-	-0,570	+	0,105
				Non-active Compounds			
10	*SSSRS*	-	$1.1x10^3$	-	-0,758	-	-0,804
11	*RSSSS*	-	$1.9x10^3$	-	-0,895	-	-0,635
12	SSRRR*	-	$2.6x10^3$	-	-0,954	-	-0,980
13	*RRSSR*	-	$5.5x10^3$	-	-0,562	-	-0,168
14	*SSRRS*	-	$7.1x10^3$	-	-0,955	-	-0,965
15	*RRSRS*	-	$7.8x10^3$	-	-0,997	-	-0,996
16	RSRRR*	-	$23x10^3$	-	-1,000	-	-1,000
17	*SRRRR*	-	$33x10^3$	-	-0,790	-	-0,894
18	*RSSSR*	-	$36x10^3$	-	-0,893	-	-0,771
19	*RSRSR*	-	$47x10^3$	-	-0,981	-	-0,959
20	RSRSS*	-	$60x10^3$	-	-0,981	-	-0,930
21	*RRRRR*	-	10^5	-	-1,000	-	-1,000
22	*SRRRS*	-	10^5	-	-0,794	-	-0,823
23	*RRRSS*	-	10^5	-	-0,912	-	-0,666
24	SRSRR*	-	10^5	-	-0,179	-	-0,486
25	*RRRRS*	-	10^5	-	-1,000	-	-0,999
26	*RRSRR*	-	10^5	-	-0,997	-	-0,997
27	*SSSRR*	-	10^5	-	-0,753	-	-0,882
28	RSSRS*	-	10^5	-	-0,999	-	-0,999
29	*RRRSR*	-	10^5	-	-0,910	-	-0,792
30	*RSSRR*	-	10^5	-	-0,999	-	-1,000
31	*RSRRS*	-	10^5	-	-1,000	-	-1,000
32	SRSRS*	-	10^5	-	-0,190	-	-0,251

*Compounds used in the Test set. [a]Notation of the chiral centres 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 concentration in nM taken from previous works.[16,29] [d] ΔP posterior probability predicted for each compound using Eqs. 17 and 18.

A comparison between the results obtained in our study and those achieved with other cheminformatic approaches is depicted in Table 8. It should be remarked that our models contain one variable less than the model obtained with MARCH-INSIDE molecular descriptors [29] and the same number of variables that were used by us with atom-based linear indices to develop the QSAR models.[70] 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 models.

Table 8. Classification of 32 Perindoprilate's 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 Stochastic Linear indices (Eq. 18)	2	0.392	7.348	100.00	100.00	15.52
Bond-based Non-Stochastic Linear indices (Eq. 17)	2	0.401	7.069	95.65	100.00	14.93
Atom-based Stochastic Linear indices [38]	2	0.399	7.789	95.65	100.00	15.02
Atom-based Non-Stochastic Linear indices [38]	2	0.398	7.82	100.00	88.88	15.08
MARCH-INSIDE molecular descriptors [29]	3	0.380	8.43	91.30	88.88	10.30

n: number of parameters in the obtained model.

Concluding Remarks

In this study, we showed that bond-based 3D-chiral linear indices 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.

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

References and Notes

1. de Julian-Ortiz, J. V.; Garcia-Domenech, R.; Galvez, J.; Soler , R.; García-March, F. J.; Antón-Fos, G. M. J Chromatogr A 1996, 719(1), 37-44.
2. Aires-de-Sousa, J.; Gasteiger, J. J Mol Graph Model 2002, 20(5), 373-388.
3. Solms, J.; Vuataz, L.; Egli, R. H. Experientia 1965, 21(12), 692-694.
4. Schiffman, S. S.; Clark, T. B., 3rd; Gagnon, J. Physiol Behav 1982, 28(3), 457-465.
5. Laska, M.; Teubner, P. Chem Sens 1999, 24(2), 161-170.
6. Polak, E. H.; Fombon, A. M.; Tilquin, C.; Punter, P. H. Behav Brain Res 1989, 31(3), 199-206.
7. DeCamp, W. H. Chirality 1989, 1, 2.
8. Hutt, A. J.; Tan, S. C. Drugs 1996, 52 Suppl 5, 1-12.
9. Wnendt, S.; Zwingenberger, K. Nature (London) 1997, 385(6614), 303-304.
10. Schumacher, H.; Blake, D. A.; Gurian, J. M.; Gillette, J. R. J Pharmacol Exp Ther 1968, 160(1), 189-200.
11. Stinson, S. C. Chem Eng News 2000, 78, 43.
12. Guye, P.-A. Compt Rend 1890(110), 714.
13. Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidovic, D. J Chem Inf Comput Sci 2004, 44(3), 831-836.
14. Aires-de-Sousa, J.; Gasteiger, J. J Chem Inf Comp Sci 2001, 41, 369-375.
15. Ruch, E. Acc Chem Res 1972, 5, 49-56.
16. de Julian-Ortiz, J. V.; de Gregorio Alapont, C.; Rios-Santamarina, I.; Garcia-Domenech, R.; Galvez, J. J Mol Graph Model 1998, 16(1), 14-18.
17. Benigni, R.; Cotta-Ramusino, M.; Gallo, G.; Giorgi, F.; Giuliani, A.; Vari, M. R. J Med Chem 2000, 43(20), 3699-3703.
18. Golbraikh, A.; Bonchev, D.; Tropsha, A. J Chem Inf Comput Sci 2001, 41(1), 147-158.
19. Wildman, S. A.; Crippen, G. M. J Chem Inf Comput Sci 2003, 43(2), 629-636.
20. Schultz, H. P.; Schultz, E. B.; Schultz, T. P. J Chem Inf Comput Sci 1995, 35, 864 - 870.
21. Aires-de-Sousa, J.; Gasteiger, J. J Comb Chem 2005, 7(2), 298-301.
22. Estrada, E.; Uriarte, E. Curr Med Chem 2001, 8(13), 1573-1588.
23. Pyka, A. J Serb Chem Soc 1997, 62(3), 251-269.
24. Pyka, A. J Planar Chromatogr Mod TLC 1993, 6, 282-288.
25. Pyka, A. J Liq Chromatogr Relat Technol 1999, 22, 41-50.

26. Gutman, I.; Pyka, A. J Serb Chem Soc 1997, 62(3), 261-265.
27. Buda , A. B.; Mislow, K. J Mol Struct (Theochem) 1991, 232(232), 1-12.
28. Moreau, G. J Chem Inf Comput Sci 1997, 37, 929 - 938.
29. Diaz, H. G.; Sanchez, I. H.; Uriarte, E.; Santana, L. Comput Biol Chem 2003, 27(3), 217-227.
30. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. J Am Chem Soc 1988, 110, 5959-5967.
31. Goodford, P. J. In QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications; Sanz, F.; Giraldo, J.; Manaut, F., Eds.; Prous Science: Barcelona, 1995, p 199-205.
32. Casanola-Martin, G. M.; Khan, M. T.; Marrero-Ponce, Y.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F. Bioorg Med Chem Lett 2006, 16(2), 324-330.
33. Marrero-Ponce, Y.; Machado-Tugores, Y.; Pereira, D. M.; Escario, J. A.; Barrio, A. G.; Nogal-Ruiz, J. J.; Ochoa, C.; Aran, V. J.; Martinez-Fernandez, A. R.; Sanchez, R. N.; Montero-Torres, A.; Torrens, F.; Meneses-Marcel, A. Curr Drug Discov Technol 2005, 2(4), 245-265.
34. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. J Mol Model 2006, 12(3), 255-271.
35. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sanchez, A. M.; Torrens, F.; Castro, E. A. Bioorg Med Chem 2005, 13(4), 1005-1020.
36. Marrero-Ponce, Y.; Montero-Torres, A.; Zaldivar, C. R.; Veitia, M. I.; Perez, M. M.; Sanchez, R. N. Bioorg Med Chem 2005, 13(4), 1293-1304.
37. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; García-Domenech, R. J Pharm Sci 2007, DOI: 10.1002/jps.21122.
38. Marrero-Ponce, Y.; Castillo-Garit, J. A. J Comput-Aided Mol Design 2005, 19(6), 369-383.
39. Marrero-Ponce, Y.; Torrens, F. (see ECSOC-9, Conference Hall G, G-015 http://wwwusces/congresos/ecsoc/9/ECSOC9HTM) 2006.
40. Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T. H.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. Bioorg Med Chem 2007, 15(3), 1483-1503.
41. Rouvray, D. H. In Chemical Applications of Graph Theory; Balaban, A. T., Ed.; Academic Press: London, 1976, p 180-181.

42. Trinajstić, N. Chemical Graph Theory; CRC Press: Boca Raton, FL, 1983.
43. Estrada, E. J Chem Inf Comput Sci 1995, 35, 31-33.
44. Estrada, E.; Ramírez, A. J Chem Inf Comput Sci 1996, 36, 837-843.
45. Estrada, E. J Chem Inf Comput Sci 1996, 36, 844-849.
46. Estrada, E.; Guevara, N.; Gutman, I. J Chem Inf Comput Sci 1998, 38, 428-431.
47. Estrada, E. J Chem Inf Comput Sci 1999, 39, 1042-1048.
48. Estrada, E.; Molina, E. J Mol Graphics Mod 2001, 20, 54-64.
49. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH: : Weinhein (Germany), 2000.
50. Edwards, C. H.; Penney, D. E. Elementary Linear Algebra; Prentice-Hall, Englewood Cliffs: New Jersey, USA, 1988.
51. 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(6), 395-407.
52. Gonzalez Diaz, H.; Olazabal, E.; Castanedo, N.; Sanchez, I. H.; Morales, A.; Serrano, H. S.; Gonzalez, J.; de Armas, R. R. J Mol Model (Online) 2002, 8(8), 237-245.
53. Gonzalez-Diaz, H.; Cruz-Monteagudo, M.; Vina, D.; Santana, L.; Uriarte, E.; De Clercq, E. Bioorg Med Chem Lett 2005, 15(6), 1651-1657.
54. Gonzalez-Diaz, H.; Uriarte, E.; Ramos de Armas, R. Bioorg Med Chem 2005, 13(2), 323-331.
55. González-Díaz, H.; Saíz-Urra, L.; Molina, R.; González-Díaz, Y.; Sánchez-González, A. J Comput Chem 2007, 28(6), 1042-1048.
56. Marrero-Ponce, Y. J Chem Inf Comput Sci 2004, 44(6), 2010-2026.
57. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. Molecules 2004, 9, 1100-1123.
58. Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. J Chem Inf Comput Sci 2002, 42(5), 1194-1203.
59. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. J Med Chem 2000, 43, 1975-1985.
60. Estrada, E.; Peña, A.; García-Domenech, R. J Comput-Aided Mol Des 1998, 12, 583-595.
61. Potapov, V. M. Stereochemistry; Khimia: Moscow, 1988.

62. Wang, R.; Gao, Y.; Lai, L. P. Drug Discov Des 2000, 19, 47-66.
63. Ertl, P.; Rohde, B.; Selzer, P. J Med Chem 2000, 43, 3714-3717.
64. Ghose, A. K.; Crippen, G. M. J Chem Inf Comput Sci 1987, 27, 21-35.
65. Millar, K. J. J Am Chem Soc 1990, 112, 8533-8542.
66. Gasteiger, J.; Marsilli, M. Tetrahedron Lett 1978, 34, 3181-3184.
67. Pauling, L. The Nature of Chemical Bond; Cornell University Press: Ithaca, NY, 1939.
68. Browder, A. Mathematical Analysis. An Introduction.; Springer-Verlag: New York, 1996.
69. Axler, S. Linear Algebra Done Right.; Springer-Verlag: New York, 1996.
70. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. Bioorg Med Chem 2006, 14(7), 2398-2408.
71. Eliel, E. L.; Wilen, S.; Mander, L. Stereochemistry of Organic Compounds;; John Wiley & Sons Inc: New York, 1994.
72. Marrero-Ponce, Y.; Romero, V.: Central University of Las Villas., 2002.
73. Kier L., B.; Hall, L., H. Molecular Connectivity in Structure–Activity Analysis; Research Studies Press: Letchworth, U. K, 1986.
74. Consonni, V.; Todeschini, R.; Pavan, M. J Chem Inf Comput Sci 2002, 42, 682-692.
75. Todeschini, R.; Gramatica, P. Persp Drug Disc Des 1998, 9-11, 355–380.
76. STATISTICA version. 6.0 Statsoft, I.
77. Wold, S.; Erikson, L. In Chemometric Methods in Molecular Design; van de Waterbeemd, H., Ed.; VCH Publishers: Weinheim, 1995, p 309-318.
78. Belsey, D. A.; Kuh, E.; Welsch, R. E. Regression Diagnostics; Wiley: New York, 1980.
79. Golbraikh, A.; Tropsha, A. J Mol Graph Model 2002, 20(4), 269-276.
80. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Bioinformatics 2000, 16, 412-424.
81. Coats, E. A. In 3D QSAR in Drug Design.; Kluwer/ESCOM: Dordrecht, 1998.
82. Silverman, B. D. Quant Struct-Act Relat 2000, 19, 237-246.
83. Coats, E. A. Persp Drug Disc Des 1998, 12-14, 199-213.
84. Stiefl, N.; Baumann, K. J Med Chem 2003, 46(8), 1390-1407.
85. Klebe, G.; Abraham, U.; Mietzner, T. J Med Chem 1994, 37, 4130-4146.
86. Robert, D.; Amat, L.; Carbo-Dorca, R. J Chem Inf Comput Sci 1999, 39(2), 333-344.

87. Lobato, M.; Amat, L.; Besalu, E.; Carbo-Dorca, R. Quant Struct-Act Relat 1997, 16, 465-472.
88. Wagener, M.; Sadowski, J.; Gasteiger, J. J Am Chem Soc 1995, 117, 7769-7775.
89. Parretti, M. F.; Kroemer, R. T.; Rothman, J. H.; Richards, W. G. J Comput Chem 1997, 18, 1334-1353.
90. So, S. S.; Karplus, M. J Med Chem 1997, 40(26), 4347-4359.
91. Chen, H.; Zhou, J.; Xie, G. J Chem Inf Comp Sci 1998, 38, 243-250.
92. Marrero-Ponce, Y.; Díaz, H. G.; Romero, V.; Torrens, F.; Castro, E. A. Bioorg Med Chem 2004, 12, 5331-5342.
93. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. J Mol Graph Model 2007, 26(1), 32-47.
94. Cronin, M. T.; Schultz, T. W. J Mol Struct (Theochem) 2003, 622, 39-51.
95. Vicent, M.; Marchand, B.; Rémond, G.; Jaquelin-Guinamant, S.; Damien, G.; Portevin, B.; Baumal, J.; Volland, J.; Bouchet, J.; Lambert, P.; Serkiz, B.; Luitjen, W.; Lauibie, M.; Schiavi, P. Drug Des Discov 1992, 9, 11.
96. Liu, S. S.; Yin, C. S.; Wang, L. S. J Chem Inf Comput Sci 2002, 42(3), 749-756.
97. Maw, H. H.; Hall, L. H. J Chem Inf Comput Sci 2001, 41(5), 1248-1254.
98. Besalu, E.; Girones, X.; Amat, L.; Carbo-Dorca, R. Acc Chem Res 2002, 35(5), 289-295.
99. Silverman, B. D.; Platt, D. E. J Med Chem 1996, 39(11), 2129-2140.