

# 3D-Chiral quadratic indices of the ‘molecular pseudograph’s atom adjacency matrix’ and their application to central chirality codification: classification of ACE inhibitors and prediction of $\sigma$ -receptor antagonist activities

Yovani Marrero Ponce,<sup>a,b,\*</sup> Humberto González Díaz,<sup>b,c</sup> Vicente Romero Zaldivar,<sup>d</sup> Francisco Torrens<sup>e</sup> and Eduardo A. Castro<sup>f</sup>

<sup>a</sup>Department of Pharmacy, Faculty of Chemical-Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

<sup>b</sup>Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

<sup>c</sup>Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Spain

<sup>d</sup>Faculty of Informatics, University of Cienfuegos, Cienfuegos 59430, Cuba

<sup>e</sup>Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (València), Spain

<sup>f</sup>CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de la Plata, CC 962, 1900 La Plata, Buenos Aires, Argentina

Received 10 April 2004; revised 22 July 2004; accepted 24 July 2004

Available online 11 September 2004

**Abstract**—Quadratic indices of the ‘molecular pseudograph’s atom adjacency matrix’ have been generalized to codify chemical structure information for chiral drugs. These 3D-chiral quadratic indices make use of a trigonometric 3D-chirality correction factor. These indices are nonsymmetric and reduced to classical (2D) descriptors when symmetry is not codified. By this reason, it is expected that they will be useful to predict symmetry-dependent properties. 3D-Chirality quadratic indices are real numbers and thus, can be easily calculated in TOMOCOMD-CARDD software. These descriptors circumvent the inability of conventional 2D quadratic indices (*Molecules* **2003**, 8, 687–726. <http://www.mdpi.org>) and other (chirality insensitive) topological indices to distinguish  $\sigma$ -stereoisomers. In this paper, we extend our earlier work by applying 3D-chirality quadratic indices to two data sets containing chiral compounds. Consequently, in order to test the potential of this novel approach in drug design we have modelled the angiotensin-converting enzyme inhibitory activity of perindoprilate’s  $\sigma$ -stereoisomers combinatorial library. Two linear discriminant analysis (LDA) models were obtained. The first one model was performed considering all data set as training series and classifies correctly 88.89% of active compounds and 100.00% of nonactive one for a global good classification of 96.87%. The second one LDA-QSAR model classified correctly 83.33% of the active and 100.00% of the inactive compounds in a training set, result that represent a total of 95.65% accuracy in classification. On the other hand, the model classifies 100.00% of these compounds in the test set. Similar predictive behaviour was observed in a leave-one-out cross-validation procedure for both equations. Canonical regression analysis corroborated the statistical quality of these models ( $R_{\text{can}}$  of 0.82 and of 0.76, respectively) and was also used to compute biology activity canonical scores for each compound. Finally, prediction of the biological activities of chiral 3-(3-hydroxyphenyl)piperidines, which are  $\sigma$ -receptor antagonists, by linear multiple regression analysis was carried out. Two statistically significant QSAR models were obtained ( $R^2 = 0.940$ ,  $s = 0.270$  and  $R^2 = 0.977$ ,  $s = 0.175$ ). These models showed high stability to data variation in the leave-one-out cross-validation procedure ( $q^2 = 0.912$ ,  $s_{\text{cv}} = 0.289$  and  $q^2 = 0.957$ ,  $s_{\text{cv}} = 0.211$ ). The results of this study compare favourably with those obtained with other chirality descriptors applied to the same data set. The 3D-chiral TOMOCOMD-CARDD approach provides a powerful alternative to 3D-QSAR.

© 2004 Elsevier Ltd. All rights reserved.

## 1. Introduction

Molecular isomerism is a subject of fundamental interest to organic chemists. Two molecules with identical chemical formulas but different states of symmetry of

**Keywords:** Angiotensin-converting enzyme inhibitors;  $\sigma$ -Receptor antagonists; TOMOCOMD-CARDD Approach; 3D-Chiral quadratic indices.

\*Corresponding author. Tel.: +53-42-281192/281473; fax: +53-42-281130/281455; e-mail addresses: [yovanimp@qf.uclv.edu.cu](mailto:yovanimp@qf.uclv.edu.cu); [ymarrero77@yahoo.es](mailto:ymarrero77@yahoo.es)

only one atom are referred to as enantiomers. Enantiomers are nonsuperimposable mirror image isomers and they may also be referred to as enantiomorphs, optical isomer or optical antipodes (because, they rotate the polarization plane in opposite direction). The molecules with identical 2D structural formulas containing more than one asymmetric atom as referred to as  $\sigma$ -diastereomers.<sup>1</sup> In the literature, the asymmetric atoms are often referred to as chiral atoms and molecules containing chiral atoms are referred to as chiral molecules.

However, if a molecule contains chiral atoms, it can be an achiral molecule, because it may present a symmetric element such as symmetric centre or plane (i.e., meso-compounds: tartaric acid). In addition, chirality can be caused by a spatial isomerism resulting from the lack of free rotation around single or double bonds such as in derivatives of biphenyl or in allenes, rather than due to chiral atoms.<sup>1,2</sup> Thus, a necessary and sufficient condition to consider a compound as chiral is the absence of an element of symmetry, which avoids chirality. That is to say, if there is not an inverse axis ( $S_n$ ) of symmetry (including the plane and centre as self-containing cases of  $S_n$ ) we can affirm that we are in the presence of a chiral molecule (having or not chiral atoms).<sup>1,3</sup>

In the original definition, Kelvin formulated the concept of chirality as an abstract property of geometric objects: 'I call any geometrical figure, or group of points, *chiral*... if its image in a plane mirror, ideally realized cannot be brought to coincide with itself'.<sup>4,5</sup> Although enantiomorphs cannot be placed upon each other (superimposed or overlapped) so that at least parts of them coincide.

Hence, chirality cannot be equated with asymmetry (i.e., the total absence of symmetry) because all molecules present at least one simple axis of symmetry. Anyhow, this element does not preclude chirality. For this reason, in the past the word 'dissymmetry' was often used as a synonym for what we now call chirality. Pasteur was well aware of the difference between 'dissymmetry' and asymmetry, as evidenced by the French title of his lecture 'Recherches sur la Dissymétrie moléculaire des produits Organiques naturels'.<sup>6</sup> Unfortunately, this was translated into English as 'Researches on the Molecular Asymmetry of Natural Organic Products'.<sup>7</sup> The word dissymmetry, in the sense of what we now call chirality seems to have been lost to the English language over time.<sup>3</sup>

Enantiomers of a given compounds have identical chemical properties with regard to their reaction with nonchiral reagents, although they will give products with different configurations. In addition, they may show differences in behaviour (both in reaction rates and in product stereochemistry) in their interactions with a chiral reagent. In this sense, many biochemical process and phenomena are stereospecific. For instance, L- and D-enantiomers of amino acids have different tastes,<sup>8,9</sup> enantiomers of some compounds have different odours<sup>10,11</sup> and many medicinal preparation have phys-

iological properties different from those of their enantiomers.<sup>12–14</sup> The case of thalidomide is an example of a problem that was, at least, complicated by the ignorance of stereochemical effects.<sup>15</sup> 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 Food & Drug Administration (FDA) requires a detailed study of both enantiomers.<sup>16</sup>

Based on the experience of chemists, we can recognize at least three kinds of chemical properties or biological activities that depend on the symmetry properties of the molecule, the environment and the apparatus used to measure the property.<sup>3,17</sup>

1. Symmetry-independent properties (they have the same absolute value and sign for both enantiomers and are invariant to both proper and improper operations of symmetry, they are known as scalar properties). These properties are generally measured in an isotropic environment with symmetric apparatus, for example, boiling points, density. These properties do not need the use of chiral molecular descriptors to be predicted.
2. Molecular symmetry dependent properties measured in isotropic environments with specific symmetric apparatus (they have the same absolute value but opposite sign and are in general pseudoscalar properties). Pseudoscalar properties are those, which remains invariant to proper operations of symmetry (rotations) but changes sign under an improper operation (reflections). These properties depend in their absolute magnitude on molecular symmetry and need the use chiral molecular symmetric indices to be predicted, for example, optical activity.
3. Molecular symmetry dependent properties measured in nonisotropic environments (they have different absolute value and could have or not opposite mathematical sign depending on the scale). They could be scalar or pseudoscalar with respect to the system as a whole (the molecule and the molecular environment). This specific group of properties need the use of nonsymmetric chiral descriptors, for example, retention time of enantiomers in chiral chromatographic or  $\sigma$ -receptors antagonists activities (see below).

Attempts to give quantitative meaning to molecular chirality can be dated almost as far back as van't Hoff's and LeBel's proposition to extend the structural formulas of chemistry into three-dimensional space (3D). In 1890 Guye introduced the first function designed to correlated a pseudoscalar property, that is optical rotation, with the molecular structure of chiroids—the first example of a chirality function in chemistry.<sup>18</sup> Chirality, however, is an inherent molecular property that depends only on symmetry and that is independent of its physical and chemical manifestations. It should therefore be possible to quantify chirality, that is to construct a chirality measure, without reference to any experimental data.

Several quantitative measures of chirality have been developed in the past and were extensively reviewed.<sup>19–21</sup> Buda and Mislow distinguished between two classes of measures.<sup>19</sup> 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.<sup>22</sup> 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 of each atom.<sup>23</sup> Application 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).<sup>24</sup> Derived chirality descriptors were correlated with biological activity by Julián-Ortiz et al.,<sup>25</sup> Golbraikh et al.<sup>26</sup> and more recently by González-Díaz et al.<sup>17</sup> These indices are referred as chirality TIs (CTIs). The main purpose on developing these descriptors is to be able to account for chiral molecules, which are well known to play an import role in medicinal chemistry. 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.<sup>27</sup>

In addition to CTIs, the characterization of symmetry, and specifically chiral structural features in Computer-Aided Drug Discovery (CADD), has become possible only since the development of 3D-QSAR methods. Among these methods, special mention must be made of the use of CoMFA.<sup>28</sup> Evidently; the chirality in CoMFA is taken into account by default, since 3D field values of chiral isomers are different. Despite its wide popularity, CoMFA is not always applicable, especially in situation where compounds under investigation are highly flexible. Even almost these difficulties are solved by Grid (a CoMFA like last generation method) several drawbacks still remain when large data must be processed.<sup>29</sup>

Recently, the present author has introduced the novel computer-aided molecular design scheme TOMOCOMD (acronym of TOPological MOlecular COMputer Design). It calculates several new families of topologic molecular descriptors. One of these families has been defined as molecular quadratic indices by analogy with the quadratic mathematical forms.<sup>30</sup> This point of view was successfully applied to the prediction of physical properties and Caco-2 permeability of organic compounds and drugs, respectively.<sup>30,31</sup> The method is very flexible and makes possible the study of small molecules as well as macromolecules such as nucleic acid.<sup>32</sup>

In this context, the main aim of the present manuscript is to extend quadratic indices of the ‘molecular pseudograph’s atom adjacency matrix’ in order to codify chirality related structural features. The problem of classification of ACE (angiotensin-converting enzyme) inhibitors and prediction of  $\sigma$ -receptor antagonist activities are selected as illustrative example of method application. These examples will be used as matter of comparison with other CTIs as well.

## 2. Theoretical approach

The general principles of the quadratic indices of the ‘molecular pseudograph’s atom adjacent matrix’ for small-to-medium sized organic compounds have been explained in some detail elsewhere.<sup>30,31</sup> However, an overview of this approach will be given.

The molecular vector ( $X$ ) is constructed in order to calculate the molecular quadratic indices for a molecule where the components of this vector are numeric values, which represent a certain atomic property. These properties characterize each kind atom within the molecule. Such properties can be the electronegativity, density (standard state), atomic radii and so on.

If a molecule consists of  $n$  atoms (*vector of  $\mathfrak{R}^n$* ), then the  $k$ th total quadratic indices,  $q_k(x)$  are calculated as a quadratic form ( $q: \mathfrak{R}^n \rightarrow \mathfrak{R}$ ) in canonical basis as shown in Eq. 1,

$$q_k(x) = \sum_{i=1}^n \sum_{j=1}^n {}^k a_{ij} X_i X_j \quad (1)$$

where  ${}^k a_{ij} = {}^k a_{ji}$  (symmetric square matrix),  $n$  is the number of atoms of the molecule and  $X_1, \dots, X_n$  are the coordinates of the molecular vector ( $X$ ) in a system of basis vectors of the  $\mathfrak{R}^n$ . One can choose the basis vectors the coordinates of the same vector will be different.<sup>33–35</sup> The values of the coordinates depend thus in an essential way on the choice of the basis. The so-called canonical (‘natural’) bases,  $e_j$  denote the  $n$ -tuple having 1 in the  $j$ th position and 0’s elsewhere.<sup>33–35</sup> In the canonical bases, the coordinates of any vector  $X$  coincide with the components of this vector.<sup>30–35</sup> For that reason, those coordinates can be considered as weights (atom-labels) of the vertices of the pseudograph.

The coefficients  ${}^k a_{ij}$  are the elements of the  $k$ th power of the matrix  $\mathbf{M}(G)$  of the molecular pseudograph ( $G$ ). Here,  $\mathbf{M}(G) = \mathbf{M} = [a_{ij}]$ , denote the matrix of  $q_k(x)$  with respect to the natural basis. In this matrix  $n$  is the number of vertices (atoms) of  $G$  and the elements  $a_{ij}$  are defined as follows:<sup>30,31</sup>

$$\begin{aligned} a_{ij} &= P_{ij} & \text{if } i \neq j & \text{ and } \exists e_k \in E(G) \\ &= L_{ii} & \text{if } i = j \\ &= 0 & \text{otherwise} \end{aligned} \quad (2)$$

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

Given that  $a_{ij} = P_{ij}$ , the elements  $a_{ij}$  of this matrix represent the number of bonds between an atom  $i$  and other  $j$ . The matrix  $\mathbf{M}^k$  provides the number of walks of length  $k$  that links the vertices  $v_i$  and  $v_j$ . For this reason, each edge in  $\mathbf{M}^1$  represents two electrons belonging to the covalent bond between atoms (vertices)  $v_i$  and  $v_j$ ; for example, the inputs of  $\mathbf{M}^1$  are equal to 1, 2 or 3 when appears simple, double or triple bonds between vertices  $v_i$  and  $v_j$ , respectively. On the other hand, molecules containing aromatic rings with more than one canonical structure are represented like a pseudograph. It happens for substituted aromatic compounds such as pyridine, naphthalene, quinoline and so on, where the presence  $\text{PI}(\pi)$  electrons is accounted by means of loops in each atom of the aromatic ring. Conversely, aromatic rings having only one canonical structure, such as furan, thiophene and pyrrole are represented like a multigraph.

Note that the mathematical quadratic form's matrices,  $\mathbf{M}^k$ , are graph-theoretic electronic-structure models, like an 'extended Hückel' model. The  $\mathbf{M}^1$  matrix considers all valence-bond electrons ( $\sigma$ - and  $\pi$ -networks) in one step and their power ( $k = 0, 1, 2, 3, \dots$ ) can be considering as an interacting-electron chemical-network model in  $k$  step. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical-bonding ideas.

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

On the other hand, the defining equation 1 for  $q_k(x)$  may be written as the single matrix equation:

$$q_k(x) = [X_1 \quad \dots \quad X_n] \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}^k \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad (3)$$

or in the more compact form,

$$q_k(x) = [X]^t \mathbf{M}^k [X] \quad (4)$$

where  $[X]$  is a column vector (a  $n \times 1$  matrix) of the coordinates of  $X$  in the canonical base of  $\mathcal{R}^n$ ,  $[X]^t$  the transpose of  $[X]$  (a  $1 \times n$  matrix) and  $\mathbf{M}^k$  the  $k$ th power of the matrix  $\mathbf{M}$  of the molecular pseudograph  $G$  (mathematical quadratic form's matrix). In Table 1 depict the calculation of the quadratic indices of the molecular pseudograph's atom adjacency matrix for amino-pyridin-4-yl-acetaldehyde.

In addition to an total quadratic indices computed for the whole molecule an local-fragment (atom and atom-type) formalism can be developed. These descriptors

are termed local quadratic indices of the 'molecular pseudograph's atom adjacency matrix',  $q_{kL}(x)$ . The  $q_{kL}(x)$  are graph-theoretical invariant for a given fragment  $F_R$  (connected subgraph) within a specific pseudograph  $G$ . The definition of these descriptors is as follows:

$$q_{kL}(x) = \sum_{i=1}^m \sum_{j=1}^m {}^k a_{ijL} X_i X_j \quad (5)$$

where  $m$  is the number of atoms of the fragment of interest and  ${}^k a_{ijL}$  is the element of the file 'i' and column 'j' of the matrix  $\mathbf{M}_L^k = \mathbf{M}^k(G, F_R)$  [ $q_{kL}(x) = q_k(x, F_R)$ ]. This matrix is extracted from the  $\mathbf{M}^k$  matrix and contains the information referred to the vertices of the specific molecular fragments ( $F_R$ ) and also of the molecular environment.

The matrix  $\mathbf{M}_L^k = [{}^k a_{ijL}]$  with elements  ${}^k a_{ijL}$  is defined as follows:

$$\begin{aligned} {}^k a_{ijL} &= {}^k a_{ij} && \text{if both } v_i \text{ and } v_j \text{ are vertices contained} \\ &&& \text{within } F_R \\ &= \frac{1}{2} {}^k a_{ij} && \text{if } v_i \text{ or } v_j \text{ is contained within } F_R \\ &= 0 && \text{otherwise} \end{aligned} \quad (6)$$

with the  ${}^k a_{ij}$  being the elements of the  $k$ th power of  $\mathbf{M}$ . These local analogues can also be expressed in matrix form by the expression:

$$q_{kL}(x) = [X]^t \mathbf{M}_L^k [X] \quad (7)$$

Note that for every partitioning of a molecule into  $Z$  molecular fragment there will be  $Z$  local molecular fragment matrices. That is to say, if a molecule is partitioned into  $Z$  molecular fragments, the matrix  $\mathbf{M}^k$  can be partitioned into  $Z$  local matrices  $\mathbf{M}_L^k$ ,  $L = 1, \dots, Z$  and the  $k$ th power of matrix  $\mathbf{M}$  is exactly the sum of the  $k$ th power of the local ('molecular fragment')  $Z$  matrices,

$$\mathbf{M}^k = \sum_{L=1}^Z \mathbf{M}_L^k \quad (8)$$

or in the same way as  $\mathbf{M}^k = [{}^k a_{ij}]$  where,

$${}^k a_{ij} = \sum_{L=1}^Z {}^k a_{ijL} \quad (9)$$

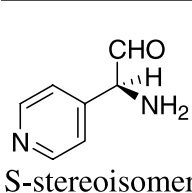
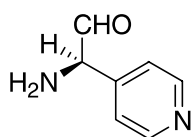
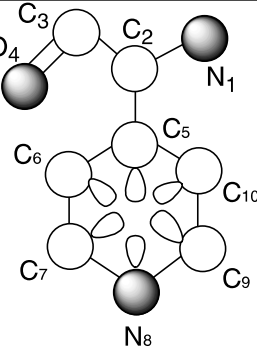
and the total quadratic indices is the sum in the quadratic indices of the  $Z$  molecular fragments (see Table 1),

$$q_k(x) = \sum_{L=1}^Z q_{kL}(x) \quad (10)$$

Any local quadratic index has a particular meaning, especially for the first values of  $k$ , where the information about the structure of the fragment  $F_R$  is contained. High values of  $k$  are in relation with the environment information of the fragment  $F_R$  considered inside the molecular pseudograph ( $G$ ).



**Table 1.** Definition and calculation of total (whole-molecule) and local (atom) 3D-chiral and simple 2D-quadratic indices of the molecular pseudograph's atom adjacency matrix of the amino-pyridin-4-yl-acetaldehyde molecule

<div><p>S-stereoisomer</p></div> <div><p>R-stereoisomer</p></div> <div><p><b>Molecular Structure</b></p></div>	<div><p><b>Molecular Pseudograph</b> (Hydrogen Suppressed-pseudograph)</p></div>	<p><math>^*\mathbf{X} = [\text{N}_1, \text{C}_2, \text{C}_3, \text{O}_4, \text{C}_5, \text{C}_6, \text{C}_7, \text{N}_8, \text{C}_9, \text{C}_{10}]</math> Chiral Molecular Vector: <math>^*\mathbf{X} \in \mathfrak{R}^{10}</math> In the definition of the <math>^*\mathbf{X}</math>, as chiral molecular vector, the chemical symbol of the element is used to indicate the corresponding electronegativity value + 3D-chirality factor. That is: if we write O it means <math>\chi(\text{O})</math> (oxygen Mulliken electronegativity) + <math>\sin((\omega_A+4\Delta)\pi/2)</math>. Therefore, if we use the canonical basis of <math>\mathfrak{R}^{10}</math>, the coordinates of any vector <math>^*\mathbf{X}</math> coincide with the components of that chiral molecular vector. <math>\sin((\omega_A+4\Delta)\pi/2)</math> is the trigonometric chirality correction factor and take different values in order to codify specific stereochemical information such as chirality. 3D-chiral descriptor reduces to simples (2D) quadratic indices ones for molecules without specific 3D characteristics.</p>																																																																																																														
<p><math>[\mathbf{X}]</math>: Column vector of coordinates of <math>^*\mathbf{X}</math> in the Canonical basis of <math>\mathfrak{R}^{10}</math> (a <math>n \times 1</math> matrix) <math>[\mathbf{X}] = [2.33, \mathbf{2.63}, 2.63, 3.17, 2.63, 2.63, 2.63, 2.33, 2.63, 2.63]</math> for chirality insensitive quadratic indices <math>[\mathbf{X}] = [2.33, \mathbf{3.63}, 2.63, 3.17, 2.63, 2.63, 2.63, 2.33, 2.63, 2.63]</math> for R-stereoisomer <math>[\mathbf{X}] = [2.33, \mathbf{1.63}, 2.63, 3.17, 2.63, 2.63, 2.63, 2.33, 2.63, 2.63]</math> for S-stereoisomer <math>[\mathbf{X}]^t</math>: Transposed of <math>[\mathbf{X}]</math> (a <math>1 \times n</math> matrix)</p>																																																																																																																
<div><div><math display="block">^*q_1(x) = \sum_{i=1}^{10} \sum_{j=1}^{10} {}^1a_{ij} {}^*X_i {}^*X_j = [\mathbf{X}]^t \mathbf{M}^1 [\mathbf{X}] = [\text{N C C O C C C N C C}]</math></div><div><table><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>N</td></tr><tr><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>C</td></tr><tr><td>0</td><td>1</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>C</td></tr><tr><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>O</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>C</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>C</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>C</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>N</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>C</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>C</td></tr></table></div></div>			0	1	0	0	0	0	0	0	0	0	N	1	0	1	0	1	0	0	0	0	0	C	0	1	0	2	0	0	0	0	0	0	C	0	0	2	0	0	0	0	0	0	0	O	0	1	0	0	1	1	0	0	0	1	C	0	0	0	0	1	1	1	0	0	0	C	0	0	0	0	0	1	1	1	0	0	C	0	0	0	0	0	0	1	1	1	0	N	0	0	0	0	0	0	0	1	1	1	C	0	0	0	0	1	0	0	0	1	1	C
0	1	0	0	0	0	0	0	0	0	N																																																																																																						
1	0	1	0	1	0	0	0	0	0	C																																																																																																						
0	1	0	2	0	0	0	0	0	0	C																																																																																																						
0	0	2	0	0	0	0	0	0	0	O																																																																																																						
0	1	0	0	1	1	0	0	0	1	C																																																																																																						
0	0	0	0	1	1	1	0	0	0	C																																																																																																						
0	0	0	0	0	1	1	1	0	0	C																																																																																																						
0	0	0	0	0	0	1	1	1	0	N																																																																																																						
0	0	0	0	0	0	0	1	1	1	C																																																																																																						
0	0	0	0	1	0	0	0	1	1	C																																																																																																						

3D-chiral quadratic indices of first order is a quadratic form;  $^*q_1(x): \mathfrak{R}^n \rightarrow \mathfrak{R}$  such that,  
 $^*q_1(\text{N}_1, \text{C}_2, \text{C}_3, \text{O}_4, \text{C}_5, \text{C}_6, \text{C}_7, \text{N}_8, \text{C}_9, \text{C}_{10}) = (1(\text{C}_5)^2 + 1(\text{C}_6)^2 + 1(\text{C}_7)^2 + 1(\text{N}_8)^2 + 1(\text{C}_9)^2 + 1(\text{C}_{10})^2 + 2\text{N}_1\text{C}_2 + 2\text{C}_2\text{C}_3 + 2\text{C}_2\text{C}_5 + 4\text{C}_3\text{O}_4 + 2\text{C}_5\text{C}_6 + 2\text{C}_5\text{C}_{10} + 2\text{C}_6\text{C}_7 + 2\text{C}_7\text{N}_8 + 2\text{N}_8\text{C}_9 + 2\text{C}_9\text{C}_{10})$

The  $k^{\text{th}}$  total (whole-molecule) 3D-chiral quadratic indices are calculated by summing the local (atom) 3D-chiral quadratic indices of all atoms in the molecule.

Local and total quadratic indices of order 0-2 ( $k = 0-2$ )									
Atom (f)	3D (R)-stereoisomer			'Classical' 2D-indices			3D (S)-stereoisomer		
	$^*q_0(x,f)$	$^*q_1(x,f)$	$^*q_2(x,f)$	$q_0(x,f)$	$q_1(x,f)$	$q_2(x,f)$	$^*q_0(x,f)$	$^*q_1(x,f)$	$^*q_2(x,f)$
N <sub>1</sub>	5.4289	8.4579	17.6847	5.4289	6.1279	17.6847	5.4289	3.7979	17.6847
*C <sub>2</sub>	13.1769	27.5517	91.1856	6.9169	19.9617	58.1756	2.6569	12.3717	31.1656
C <sub>3</sub>	6.9169	26.2211	47.6293	6.9169	23.5911	47.6293	6.9169	20.9611	47.6293
O <sub>4</sub>	10.0489	16.6742	63.2098	10.0489	16.6742	56.8698	10.0489	16.6742	50.5298
C <sub>5</sub>	6.9169	30.2976	91.7607	6.9169	27.6676	89.1307	6.9169	25.0376	86.5007
C <sub>6</sub>	6.9169	20.7507	71.01	6.9169	20.7507	68.38	6.9169	20.7507	65.75
C <sub>7</sub>	6.9169	19.9617	60.6741	6.9169	19.9617	60.6741	6.9169	19.9617	60.6741
N <sub>8</sub>	5.4289	17.6847	53.0541	5.4289	17.6847	53.0541	5.4289	17.6847	53.0541
C <sub>9</sub>	6.9169	19.9617	60.6741	6.9169	19.9617	60.6741	6.9169	19.9617	60.6741
C <sub>10</sub>	6.9169	20.7507	71.01	6.9169	20.7507	68.38	6.9169	20.7507	65.75
Total	75.585	208.312	627.8924	69.325	193.132	580.6524	65.065	177.952	539.4124

Atom and atom-type quadratic indices are specific case of local quadratic indices (for  $F_R$  = atom or atom-type). In this sense, the  $k^{\text{th}}$  atom-type quadratic indices are calculated by summing the  $k^{\text{th}}$  atom quadratic indices of all atoms of the same atom type in the molecule.

In the atom-type quadratic indices formalism, each atom in the molecule is classified into an atom-type (fragment), such as heteroatoms, heteroatoms H-bonding acceptor (O, N and S), halogens, aliphatic carbon chain, aromatic atoms (aromatic rings) and so on.<sup>30,31</sup>

**Table 2.** Values of trigonometric 3D-chirality correction factor  $[\sin((\omega_A + 4\Delta)\pi/2)]$  within the allowed domain

$\omega_A$	$\Delta$														
	−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
$\omega_{\text{nonchiral}} = 0$		0		0		0		0		0		0		0	
$\omega_S = -1$	−1		−1		−1		−1		−1		−1		−1		−1

For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the  $k$ th atom-type quadratic indices provide much useful information.

In any case, whether a complete series of indices is considered, a specific characterization of the chemical structure is obtained (whole structure or fragment), which is not repeated in any other molecule. The generalization of the matrices and descriptors to ‘superior analogues’ is necessary for the evaluation of situations where only one descriptor is unable to bring a good structural characterization.<sup>36</sup> These local indices can also be used together with total indices as variables of QSAR and QSPR models for properties or activities that depend more on a region or a fragment than on the whole molecule.

The total and local quadratic indices, as defined above, cannot codify any information about 3D molecular structure. In order to solve this problem we introduced a *trigonometric 3D-chirality correction factor* in molecular vector  $X$ . In these sense, a chirality molecular vector is obtained ( $*X$ ), where the components of  $X$  (for instance, Mulliken electronegativity ( $X_A$ )<sup>37</sup> of the atom  $A$ ) are substituted by the following term  $[\chi_A + \sin((\omega_A + 4\Delta)\pi/2)]$ .

The trigonometric 3D-chirality correction factor use a dummy variable,  $\omega_A$ <sup>17</sup> and an integer parameter,  $\Delta$ :

- $\omega_A = 1$  and  $\Delta$  is an odd number when  $A$  has  $R$  (rectus),  $E$  (entgegen), or  $a$  (axial) notation according to Cahn–Ingold–Prelog rules  
 $= 0$  and  $\Delta$  is an even number, if  $A$  does not have 3D specific environment  
 $= -1$  and  $\Delta$  is an odd number when  $A$  has  $S$  (sinister),  $Z$  (zusammen), or  $e$  (equatorial) notation according to Cahn–Ingold–Prelog rules (11)

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 their chirality, due to experience shows that chirality does not change the electronegativities of atoms in the molecule in an isotropic environment in an observable way.<sup>3</sup> This correction has principally a mathematical means and must not be source of any misunderstanding.

A severe limitation of the GBT approach is the existence of different chirality corrections and we had great difficulty in selecting one of these. In this sense, González-Dí et al.<sup>17</sup> introduced an exponential chirality factor ( $\exp(\omega_A \Delta)$ ), which eliminated indetermination in the selection of chirality and 3D scales for stochastic topologic indices. Unfortunately, this exponential factor does not solve the problem in GBT-like approaches. In this connection, the present trigonometric 3D-chiral correction factor is invariant with respect to the selection of other chirality scales for all kinds of such chiral topologic indices (GBT-like ones). Table 2 depicts the values of the trigonometric 3D-chirality correction factor for all allowed values of  $\omega_A$  and  $\Delta$  (GBT-like chirality scale and other alternative chirality scales). This table clearly shows that the trigonometric 3D-chirality factor is invariant with respect to the selection of all possible real scales. That is to say, the factor gets ever the values 1, 0 and  $-1$  for  $R$ , nonchiral and  $S$  atoms. As outlined above the demonstration of invariance for this factor with respect to other 3D features such as *ale* substitutions and  $Z/E$  or  $\pi$ -isomer 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.<sup>17,25,26</sup>

A very interesting point is that the present 3D-chiral descriptor reduces to simples (2D) quadratic indices ones for molecules without specific 3D characteristics because  $\sin(0 + 4\Delta)\pi/2 = 0$ , being  $\Delta$  zero or any even number. That is, when all the atoms in the molecule are not chiral, the TOMOCOMD-CARDD molecular descriptors or any GBT-like chiral topologic index do not change upon the introduction of this factor. This means that  $*X = X$  and thus,  $*q_k(x) = q_k(x)$ .

### 3. Experimental methods

#### 3.1. Computational methods. TOMOCOMD-CARDD approach

TOMOCOMD is an interactive program for molecular design and bioinformatics research.<sup>38</sup> The program is composed by four subprograms, each one of them dealing with drawing structures (drawing mode) and calculating 2D and 3D molecular 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 this paper we outline salient features concerned with only one of these subprograms: CARDD. This subprogram was developed upon the base of a user-friendly philosophy. The calculation of 3D-chi-

ral total and local quadratic indices for any organic molecule (or any drug-like compounds) was implemented in the TOMOCOMD-CARDD software.<sup>38</sup>

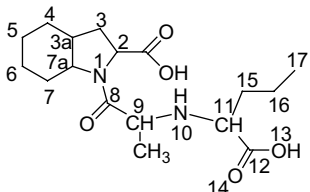
### 3.2. Data sets for 3D-chiral QSAR study

To evaluate the effectiveness of 3D-chiral quadratic indices, we have tested their ability to predict pharmacologi-

cal properties in groups with a known stereochemical influence. First, a data set of 32 perindoprilate stereoisomers, an ACE Inhibitors was used to test the applicability of the method.<sup>17,39</sup> The basic structure of perindoprilate stereoisomer and their chirality notation are shown in Table 3.

ACE acts in plasma and blood vessels, removing the C-terminal dipeptide of undecapeptide Angiotensin I to

**Table 3.** Basic structure and chirality notation of active and nonactive perindoprilate stereoisomers with their posterior probabilities in data (or data split in training and test sets) and LOO cross-validation procedure, as well as the canonical scores



Compd <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	$P_{\text{Data}}\%$ <sup>c</sup>	$P_{\text{cv}}\%$ <sup>d</sup>	Score <sup>e</sup>	$P_{\text{Train}}\%$ <sup>f</sup>	$P_{\text{Test}}\%$ <sup>g</sup>	$P_{\text{cv}}\%$ <sup>d</sup>	Score <sup>e</sup>	$P_{\text{Train}}\%$ <sup>f</sup>	$P_{\text{Test}}\%$ <sup>g</sup>
Eqs. 13 and 14 (two variables)					Eqs. 15 and 16 (two variables)			Eq. 17 (three variables)		
<i>Active compounds</i>										
SSRSS	1.1	97.49	96.63	−2.17		92.10		−1.98		99.93
SRSSS	1.2	98.60	98.55	−2.36		94.83		−2.16		99.82
SSSSS	1.5	99.31	98.89	−2.60	97.04		94.45	−2.39	99.97	
SRRSS	3.3	99.68	99.54	−2.85		98.07		−2.56		65.54
SSSSR	12.2	99.22	98.92	−2.56	96.60		94.9	−2.33	99.83	
SSRSR	29.4	97.92	97.19	−2.23	92.72		89.63	−2.02	99.52	
SRRSR	39.8	99.62	99.36	−2.80	97.72		94.83	−2.49	26.37*	
SRSSR	54	97.76	97.06	−2.21	92.46		89.57	−2.00	99.14	
RRSSS	108	5.72*	1.75*	−0.06	8.44*		2.20*	−0.08	70.94	
<i>Nonactive compounds</i>										
SSSRS	$1.1 \times 10^3$	94.60	90.73	−0.04	91.30		83.71	−0.10	94.62	
RSSSS	$1.9 \times 10^3$	86.97	82.04	−0.36	84.03		76.77	−0.37	92.46	
SSRRR	$2.6 \times 10^3$	98.09	97.82	0.31		96.34		0.26		21.29*
RRSSR	$5.5 \times 10^3$	95.08	94.47	−0.01	92.78		91.78	−0.02	71.19	
SSRRS	$7.1 \times 10^3$	86.36	84.94	−0.37	84.21		82.02	−0.36	15.07*	
RRSRS	$7.8 \times 10^3$	99.99	99.99	2.13	99.94		99.96	1.89	98.74	
RSRRR	$23 \times 10^3$	100.00	99.99	2.37		99.97		2.13		99.09
SRRRR	$33 \times 10^3$	97.12	96.02	0.17	95.35		93.75	0.17	99.1	
RSSSR	$36 \times 10^3$	84.63	83.04	−0.42	82.80		80.46	−0.40	98.72	
RSRSR	$47 \times 10^3$	99.25	99.10	0.62	98.17		97.77	0.54	96.3	
RSRSS	$60 \times 10^3$	99.55	99.43	0.78		98.67		0.67		82.6
RRRRR	$10^5$	100.00	99.99	2.30	99.96		99.98	2.09	98.4	
SRRRS	$10^5$	74.51	67.85	−0.63	76.22		70.56	−0.56	97.91	
RRRSS	$10^5$	97.07	96.48	0.16	95.14		94.17	0.15	50.91	
SRRRR	$10^5$	99.95	99.95	1.52		99.75		1.33		98.83
RRRRS	$10^5$	99.96	99.95	1.60	99.82		99.79	1.45	97.54	
RRSRR	$10^5$	100.00	99.99	2.93	99.99		99.97	2.62	99.09	
SSSRR	$10^5$	99.49	99.32	0.74	98.49		97.79	0.62	94.91	
RSSRS	$10^5$	99.93	99.92	1.41		99.68		1.23		99.91
RRRSR	$10^5$	96.59	94.34	0.11	94.85		91.97	0.12	82.87	
RSSRR	$10^5$	99.99	99.99	2.10	99.93		99.94	1.86	99.95	
RSRRS	$10^5$	99.98	99.97	1.79	99.87		99.93	1.60	99.91	
SRSRS	$10^5$	99.25	99.01	0.62		98.01		0.51		98.58

\* Misclassified compounds.

<sup>a</sup> Notation of the chiral centres in each perindoprilate derivative in the following order C<sub>2</sub>, C<sub>3a</sub>, C<sub>7a</sub>, C<sub>9</sub>, C<sub>11</sub>.

<sup>b</sup> Values of the IC<sub>50</sub>, of the compound, for ACE in nM taken from Refs. 17 and 39.

<sup>c</sup> Posterior probability predicted for each compound in data set (32 compounds).

<sup>d</sup> Posterior probability predicted for each compound in LOO cross-validation procedure.

<sup>e</sup> Canonical scores predicted using canonical analysis.

<sup>f</sup> Posterior probability predicted for each compound in training set.

<sup>g</sup> Posterior probability predicted for each compound in test set.

**Table 4.** Results of multivariate regression analysis of the Log IC<sub>50</sub> of a group of *N*-alkylated 3-(3-hydroxyphenyl)piperidines for the  $\sigma$ -receptor

Compound (alkyl group) <sup>a</sup>	Log IC <sub>50</sub> (σ-receptor)								
	Obs. <sup>b</sup>	Cal. <sup>c</sup>	Res. <sup>d</sup>	Cal. <sup>e</sup>	Res. <sup>d</sup>	Cal. <sup>f</sup>	Res. <sup>d</sup>	Cal. <sup>g</sup>	Res. <sup>d</sup>
<i>(R)</i> -3-HPP									
H	−0.66	−0.43	−0.23	−0.43	−0.23	−0.61	−0.05	−0.77	0.11
CH <sub>3</sub>	0.43	0.12	0.31	0.14	0.29	0.36	0.07	0.11	0.32
C <sub>2</sub> H <sub>5</sub>	0.95	0.72	0.23	0.78	0.17	0.87	0.08	0.77	0.18
<i>n</i> -C <sub>3</sub> H <sub>7</sub>	1.52	1.36	0.16	1.46	0.06	1.30	0.22	1.53	−0.01
<i>i</i> -C <sub>3</sub> H <sub>7</sub>	0.61	1.27	−0.66	<i>Outlier</i>	−	1.30	−0.69	1.12	−0.51
<i>n</i> -C <sub>4</sub> H <sub>9</sub>	2.05	2.00	0.05	2.13	−0.08	1.66	0.39	2.07	−0.02
2-Phenylethyl	2.10	2.22	−0.12	2.27	−0.17	2.09	0.01	2.02	0.08
<i>(S)</i> -3-HPP									
H	−1.19	−1.06	−0.13	−1.07	−0.12	−1.41	0.22	−0.77	−0.42
CH <sub>3</sub>	−0.28	−0.48	0.20	−0.47	0.19	−0.27	−0.01	−0.63	0.04
C <sub>2</sub> H <sub>5</sub>	−0.01	0.13	−0.14	0.17	−0.18	0.34	−0.35	0.21	−0.22
<i>n</i> -C <sub>3</sub> H <sub>7</sub>	0.81	0.77	0.04	0.84	−0.03	0.84	−0.03	0.96	−0.15
<i>i</i> -C <sub>3</sub> H <sub>7</sub>	0.68	0.68	0.00	0.74	−0.06	0.84	−0.16	0.73	−0.05
<i>n</i> -C <sub>4</sub> H <sub>9</sub>	1.51	1.40	0.11	1.51	0.00	1.26	0.25	1.51	0.00
2-Phenylethyl	1.80	1.62	0.18	1.65	0.15	1.75	0.05	1.46	0.34

Abbreviations: HPP, *N*-alkylated 3-hydroxyphenyl piperidines.

<sup>a</sup> Alkyl group at nitrogen ring.

<sup>b</sup> Observed values of the Log IC<sub>50</sub> for the  $\sigma$ -receptor taken from Ref. 17.

<sup>c</sup> Values calculated from Eq. 19.

<sup>d</sup> Residual, defined as [Log IC<sub>50</sub> ( $\sigma$ )Obs − Log IC<sub>50</sub> ( $\sigma$ )Cal].

<sup>e</sup> Values calculated from Eq. 20.

<sup>f</sup> Values calculated from Eq. 21 (see Ref. 17).

<sup>g</sup> Values calculated from Eq. 22 (see Ref. 25).

produce the potent blood vessel constricting octapeptide Angiotensin II. In addition, ACE inactivates the hypotensive nonapeptide Bradykinin. For these reasons, ACE is the biological target of many important antihypertensive drugs called ACEis.<sup>39</sup> Furthermore, a short data set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)piperidines that bind  $\sigma$ -receptors, are also selected as illustrative example of method application. The  $\sigma$ -receptors mediate effects induced by various opioids.<sup>25</sup> In Table 4, we depict the chemical structure and the Log IC<sub>50</sub> (50% inhibitory concentration) for this group of compounds. These examples will be used also with the aim of comparing 3D-chiral quadratic indices with other CTIs.

### 3.3. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR equation using the TOMOCOMD-CARDD descriptors and statistical techniques, such as multilinear regression analysis, linear discrimination analysis (LDA) and so on. That is to say, we can find a quantitative relation between an activity *A* and the quadratic indices having, for instance, the following appearance,

$$A = a_0^* q_0(x) + a_1^* q_1(x) + a_2^* q_2(x) + \dots + a_k^* q_k(x) + c \quad (12)$$

where *A* is the measurement of the activity,  $q_k(x)$  [or  $q_{kL}(x)$ ] is the *k*th total [or local] 3D-chiral quadratic indices, and the  $a_k$ 's are the coefficients obtained by the linear regression analysis.

In this paper, LDA was employed to develop a simple linear QSAR model to fit the classification functions in order to discriminate between two degrees of ACE inhibitory activity. The biological activity was codified by a dummy variable (ACEiactv). This variable indicates the presence of either a very active compound (ACEiactv = 1) or a nonactive compound (ACEiactv = −1). In this study very active is taken to mean a compound that has an IC<sub>50</sub> value no higher than 110 nM.

The statistical analysis were carried out with the STATISTICA software.<sup>40</sup> The quality of the model was determined examining the statistics parameter of multivariable comparison (Wilk's  $\lambda$  statistic, the square of Mahalanobis distance, Fisher ratio *F*, the corresponding *p*-level [*p*(*F*))] as well as the percentage of good classification, the proportion between the cases and variables in the equation) and the cross-validation (leave-one-out) procedure. We also developed the linear discriminant canonical analysis by checking the following statistic: canonical regression coefficient ( $R_{can}$ ), chi-squared and its *p*-level [*p*( $\chi^2$ )].<sup>41</sup>

A linear multiple regression (LMR) analysis was used to obtain a quantitative model that related 3D-chiral quadratic indices and  $\sigma$ -receptors antagonist activities. The search for the best model can be processed in terms of the highest correlation coefficient (*R*) or *F*-test equations (Fisher-ratio's *p*-level [*p*(*F*)]), and the lowest standard deviation equations (*s*). The quality of models was also determined by examining the LOO press statistics ( $q^2, s_{cv}$ ).<sup>42</sup> In recent years, the LOO press statistics (e.g.,  $q^2$ ) have been used as a means of indicating predic-



tive ability. Many authors consider high  $q^2$  values (for instance,  $q^2 > 0.5$ ) as indicator or even as the ultimate proof of the high-predictive power of a QSAR model.

#### 4. Result and discussion

First, we tested the predictive power of 3D-chiral quadratic indices in the classification of the all 32 perindoprilate stereoisomers. The QSAR-LDA model selection was subjected to the principle of parsimony (Occam's razor). We then chose a function with high statistical significance, but with as few parameters as possible. Forward stepwise was established as the strategy for variable selection. The classification model obtained is given below together with the statistical parameters of the LDA:

$$\begin{aligned} \text{ACEiactv} &= 0.496315 + 0.0032217 \cdot q_7(x) \\ &\quad - 2.70579 \times 10^{-6} \cdot q_{14}(x) \\ N &= 32, \quad \lambda = 0.33, \quad D^2 = 10.03, \\ F(2.29) &= 29.393, \quad p < 0.0000 \end{aligned} \quad (13)$$

where  $N$  is the number of compounds,  $\lambda$  is Wilk's coefficient,  $F$  is the Fisher ratio,  $D^2$  is the squared Mahalanobis distance and  $p$ -value is the significance level. The Wilks'  $\lambda$  parameter can takes values in the range of 0 (perfect discrimination) to 1 (no discrimination) and the Mahalanobis distance indicates the separation between the respective groups. It shows whether the model has an appropriate discriminatory power for differentiating between the two respective groups. The classification of cases was carried out by means of the posterior classification probabilities. Using the Mahalanobis distances to do the classification, we can now derive probabilities. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from that group centroid. In summary, the posterior probability is the probability, based on our knowledge of the values of others variables, that the respective case belongs to a particular group.

This model (Eq. 13) classified correctly 88.89% of active compounds (8/9) and the 100.00% of inactive compounds. The global good classification for the data set was 96.87% (31/32). Only one inactive stereoisomer was bad classified.

In Table 4 are shown the results of perindoprilate stereoisomers classification and a posteriori probabilities for 32 compounds of the data set.

To assess the predictability of the classification model (Eq. 13), a LOO cross-validation was carried out. This methodology systematically removed one data point at a time from the data set. A discriminant model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data point. This procedure was repeated until a complete set of predicted classification was obtained. Using this approach, the model also classified correctly 88.89% and 100.00% of active and inactive compounds, respectively. The glo-

bal classification of the LOO cross-validation procedure was the same that for data set: 96.87%.

Canonical analysis is used here to test both the ability of 3D-chiral quadratic indices to discriminate between the two groups of stereoisomers and also to order these compounds accordingly with their stability profile.

3D-Chiral total quadratic indices & LDA ACE inhibitory activity canonical analysis principal root:

$$\begin{aligned} \text{ACEroot} &= -1.13859 - 0.00105 \cdot q_7(x) \\ &\quad + 8.82 \times 10^{-7} \cdot q_{14}(x) \\ N &= 32, \quad \lambda = 0.33, \quad R_{\text{can}} = 0.82, \quad \chi^2 = 32.12, \\ p &< 0.0000 \end{aligned} \quad (14)$$

The canonical transformation of the LDA results yields one canonical root with a good canonical regression coefficient (0.82). Chi-squared test permits us to test the statistical signification of this analysis with a  $p$ -level  $< 0.0001$ .

When LDA analysis is applied to solve the two-group classification problem we ever find two classification functions.<sup>43</sup> However, we cannot use these two classification functions to evaluate all the compounds and obtain a bivariate stability map because they are not orthogonal.<sup>41</sup> To solve this problem we used canonical analysis in this case the dimensional reduction caused by canonical analysis makes possible to obtain a one-dimension stability map.<sup>41</sup>

That is the same that we can order all compounds taking into account its canonical scores. The canonical scores of all stereoisomer of perindoprilate appear in Table 3. We can detect an overall ascendant tendency of canonical scores when they are plotted in the same order in which IC<sub>50</sub> increases (activity decreases) (data not shown). As it is expected, the over all mean of canonical root scores for the group of active isomers (lowest IC<sub>50</sub> values) has an opposite sign (–) with respect to the other group [(+); highest IC<sub>50</sub> values].<sup>41</sup>

In addition, to demonstrate the true merit of this approach, we developed a direct comparison with other approaches. In connection, the 32 stereoisomers of perindoprilate was split in training and test set. These series was taken equal that González-Díaz et al.<sup>17</sup> The QSAR-LDA and canonical analysis models obtained for training set (using the same 3D-chiral quadratic indices) of 23 compounds (sixth column in Table 3:  $P_{\text{Train}}(\%)$ ) are given below together with the statistical parameters:

$$\begin{aligned} \text{ACEiactv} &= 0.9464 + 0.00237 \cdot q_7(x) \\ &\quad - 1.9958 \times 10^{-6} \cdot q_{14}(x) \\ N &= 23, \quad \lambda = 0.42, \quad D^2 = 7.12, \\ F(2.20) &= 13.734, \quad p < 0.0002 \end{aligned} \quad (15)$$

$$\begin{aligned}\text{ACEroot} &= -1.38932 - 0.00093^*q_7(x) \\ &\quad + 7.83 \times 10^{-7}^*q_{14}(x) \\ N &= 23, \quad \lambda = 0.42, \quad R_{\text{can}} = 0.76, \\ \chi^2 &= 17.29, \quad p < 0.0000\end{aligned}\quad (16)$$

These models classified correctly 83.33% (5/6) of the active and 100.00% (17/17) of the inactive compounds in the training set, for a global good classification of 95.65% (22/23). In the LOO cross-validation procedure, the model showed a good robustness: 95.65% of global good classification. The most important criterion for the acceptance or not of a discriminant model, such as model (15), is based on the statistic for external prediction set. Model (15) classifies correctly 100.00% of active (three isomers) and inactive (six isomers) compounds in the test set (see seventh column in Table 3:  $P_{\text{Test}}\%$ ). If we considered the training set and the test set (*full set*) the percentage of good classification of model (15) was 96.87% (31/32); the same that the obtained using the data set (Eq. 13). In Table 3 we give the classification of stereoisomers in the training and prediction set together with their canonical scores and their posterior probabilities calculated from the Mahalanobis distance.

A similar equation was reported by González-Díaz et al. using three MARCH-INSIDE molecular descriptors.<sup>17</sup>

$$\begin{aligned}\text{ECAactv} &= 73.19 - 2.96.73^{\text{SR}}\pi_6(\omega) + 416.68^{\text{SR}}\pi_8(\omega) \\ &\quad - 113.53^{\text{SR}}\pi_{15}(\omega) \\ N &= 23, \quad \lambda = 0.38, \quad D^2 = 8.43, \quad F(3.19) = 10.3, \\ p &< 0.00\end{aligned}\quad (17)$$

$$\begin{aligned}\text{ECAroot} &= 27.42 - 106.94^{\text{SR}}\pi_6(\omega) + 150.17^{\text{SR}}\pi_8(\omega) \\ &\quad - 41.02^{\text{SR}}\pi_{15}(\omega) \\ N &= 23, \quad \lambda = 0.38, \quad R_{\text{can}} = 0.79, \\ \chi^2 &= 18.82, \quad p < 0.000\end{aligned}\quad (18)$$

These equations had very similar training and test statistical parameters that obtained above using 3D-chiral quadratic indices (Eqs. 15 and 16).

MARCH-INSIDE & LDA ACEiactv classification function classifies correctly 91.30% of compounds in the training series, that is, two compounds are misclassified out of a total of 23. Specifically, the model (17) classifies correctly 83.33% (5/6) of active compounds and 94.12% (16/17) of nonactive stereoisomers of perindoprilate. In the test set, this model recognizes correctly 83.33% (5/6) and 100.00% (3/3) of the inactive and active compounds, respectively. If we considered the training set and the test set (*full set*) the percentage of good classification of model (17) was 90.72% (29/32). Table 3 also depicted the classification of stereoisomers in the training and prediction set together by Eq. 17.<sup>17</sup>

Finally, we will now discuss the ability of 3D-quadratic indices to predict  $\sigma$  receptor antagonist activities. 3D-Quadratic indices are nonsymmetric and reduce to classical descriptors when symmetry is not codified (see above). Besides, González-Díaz et al. conclude that  $\sigma$  receptor antagonist activities is not a pseudoscalar property<sup>17</sup> and we can expect at least a good correlation with 3D-quadratic indices.

This experiment permitted us to solve two problems at once. We will be able either to test 3D-quadratic indices on class (3) properties or to compare our method with other previously reported approaches. The QSAR-LMR models obtained for the  $\sigma$  receptor antagonist activities are follow:

$$\begin{aligned}\text{Log IC}_{50}(\sigma) &= -6.4679(\pm 0.5584) \\ &\quad + 0.109523(\pm 0.0116)^*q_0(x) \\ &\quad - 0.0022(\pm 0.0004)^*q_3(x) \\ N &= 14, \quad R = 0.969, \quad R^2 = 0.940, \\ q^2 &= 0.912, \quad F(2, 11) = 85.826, \quad s = 0.270, \\ s_{\text{cv}} &= 0.289, \quad p < 0.000\end{aligned}\quad (19)$$

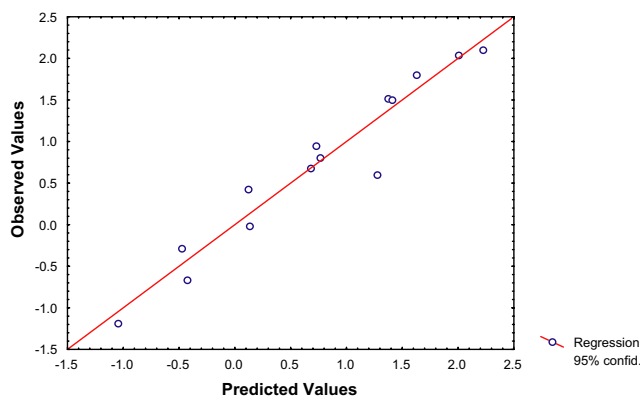
$$\begin{aligned}\text{Log IC}_{50}(\sigma) &= -6.66884(\pm 0.3661) \\ &\quad + 0.116049(\pm 0.0077)^*q_0(x) \\ &\quad - 0.0024(\pm 0.0002)^*q_3(x) \\ N &= 13, \quad R = 0.988, \quad R^2 = 0.977, \quad q^2 = 0.957, \\ F(2, 10) &= 211.20, \quad s = 0.175, \quad s_{\text{cv}} = 0.211, \\ p &< 0.0000\end{aligned}\quad (20)$$

where  $N$  is the size of the data set,  $R$  is the regression coefficient,  $s$  is the standard deviation of the regression,  $F$  is the Fischer ratio and  $q^2$ ,  $s_{\text{cv}}$  are the squared correlation coefficient and the standard deviation of the cross-validation performed by the LOO procedure, respectively. This statistics indicate that these models are appropriate for the description of chemicals studied here. In Table 4 are depicted the values of experimental and calculated Log IC<sub>50</sub> for data set (both models), and in Figures 1 and 2 are illustrated the linear relationships between them.

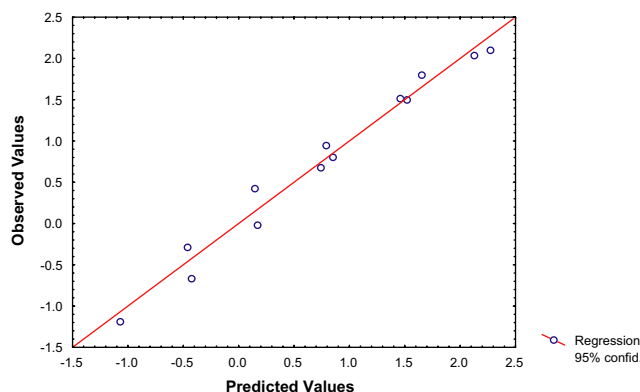
In the development of the first quantitative model for description of activities (Eq. 19), one compound was detected as statistical outlier. Outliers detection was carried out using the following standard statistical test: residual, standardized residuals, studentized residual and Cooks' distance.<sup>44</sup> Once rejected the statistical outlier, the Eq. 20 was obtained with better statistical parameters.

The correlation coefficient ( $R^2$ ), for Eqs. 19 and 20 were 0.940 and 0.977, respectively, so these models explained the 94% and 97.7% of the variance for the experimental values of Log IC<sub>50</sub>.<sup>44</sup>

Validation is a crucial aspect of any QSAR/QSPR modelling.<sup>45</sup> One of the most popular validation criteria is LOO cross-validated  $R^2$  (LOO  $q^2$ ; internal validation). For this reason, the quality of models was determined



**Figure 1.** Correlation between experimental and calculated (by Eq. 19) Log IC<sub>50</sub> of 14 compounds of the data set.



**Figure 2.** Correlation between experimental and calculated (by Eq. 20) Log IC<sub>50</sub> of 13 (one omitted) compounds of the data set.

by examining the LOO press statistics ( $q^2$ ,  $s_{cv}$ ).<sup>42</sup> Using this approach, the models 19 and 20 had a LOO  $q^2$  of 0.912 and a 0.957, respectively. These values of  $q^2$  ( $q^2 > 0.5$ ) can be considered as a proof of the high predictive ability of the models. In this context, the Eqs. 19 and 20 showed a cross-validation standard error of only 0.289 and 0.211, respectively.

Similar equations were reported for González-Díaz et al.<sup>17</sup> and De Julián-Ortiz et al.<sup>25</sup> using two [ $^A\pi_0(C^*, \omega)$ ,  $^A\pi_1(C^*, \omega)$ ] and three [ $^3D_C^*$ , PR2 and E] novel 3D-chiral TIs:

$$\begin{aligned} \text{Log IC}_{50}(\sigma) = & 6.437(\pm 0.508) \\ & - 134.35(\pm 11.33)^A\pi_0(C^*, \omega) \\ & + 9.408(\pm 2.28)^A\pi_1(C^*, \omega) \\ N = 14, \quad R = 0.96, \quad F(2, 11) = 71.17, \\ s = 0.295, \quad s_{cv} = 0.32, \quad p < 0.00000 \end{aligned} \quad (21)$$

$$\begin{aligned} \text{Log IC}_{50}(\sigma) = & 4.08 + 5.30^3D_C^* - 0.56\text{PR2} - 1.57E \\ N = 14, \quad R = 0.965, \quad F(3, 10) = 45.70, \\ s = 0.301, \quad p < 0.00000 \end{aligned} \quad (22)$$

All equations have rather similar data statistical parameters. It is remarkable that our models (Eqs. 19 and 20) uses one variable less than the one object of comparison. In order to avoid chance correlation, the number of compounds must be around one half minus one the number of investigated variables and five times the number of variables in the model. The TIs based model has three variables and barely obey this condition, Eq. 22.

## 5. Concluding remarks

Computer-aided molecular design has become in a very important tool in the development of novel chemicals to be used in different areas of human life. In this sense, the strategies for virtual screening have the aim of selecting a diverse subset of compounds from a large population by using an 'in silico' method. In these studies as well as in molecular diversity analysis of large data set of chemicals, many of traditional 2D-QSAR descriptors are not applicable as they apply to congeneric series. Thus, the continuous definition of novel molecular descriptors that could explain different biological properties by means of 2D- and 3D-QSAR is necessary.

As we have shown in the present work, the generalized TOMOCOMD-CARDD approach is not only able to discriminate between active and inactive perindoprilate stereoisomers, but also to codify information related to pharmacological property highly dependent on molecular symmetry of a set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)-piperidines that bind  $\sigma$ -receptors. This finding is only a preliminary conclusion and a more in-depth analysis of the potential of the 3D-chiral quadratic indices is necessary. However, we show that for two data sets chiral-QSAR models that use 3D-chirality total quadratic indices had better or similar predictive ability as compared to other previously reported approaches.

## Acknowledgements

The authors give thanks to the anonymous referees for their useful comments, which contribute to an improved presentation of these results.

## References and notes

- Potapov, V. M. *Stereochemistry*; Khimia: Moscow, 1988.
- Kislow, K. Fuzzy Restrictions and Inherent Uncertainties in Chirality Studies. In *Fuzzy Logic in Chemistry*; Rouvray, D. H., Ed.; Academic: San Diego, 1997; pp 65–90.
- Eliel, E.; Wilen, S.; Mander, L. *Stereochemistry of Organic Compounds*; John Wiley & Sons, 1994.
- Kelvin, W. T. *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light*; Clay, C. J.: London, 1904; p 619.
- Morrison, R. T.; Boyd, R. N. *Organic Chemistry*; Allyn and Bacon: New York, 1983.
- Jacques, J. *Sur la Dissymétrie Moléculaire*; Christian Bourgois: Paris, France, 1986.
- Pasteur L. Researches on the Molecular Asymmetry of Natural Organic Products; Alembic Club Reprint N0. 14, Alembic Club, Edinburgh, UK, 1905.

8. Solms, J.; Vuataz, L.; Egli, R. H. *Experiencia* **1965**, *21*, 692.
9. Schiffman, S. S.; Clark, T. B., III; Gagnon, J. *Physiol. Behav.* **1982**, *28*, 457.
10. Laska, M.; Teubner, P. *Chem. Senses* **1999**, *24*, 161.
11. Polak, E. H.; Fombon, A. M.; Tilquin, C.; Punter, P. H. *Behav. Brain Res.* **1989**, *31*, 199.
12. DeCamp, W. H. *Chirality* **1989**, *1*, 2.
13. Hutt, A. J.; Tan, S. C. *Drugs* **1996**, *52*, 1.
14. Wnendt, S.; Zwingenberger, K. *Nature* **1997**, *385*, 303.
15. Schumacher, H.; Blake, D. A.; Gurian, J. M.; Gillette J. R. *J. Pharmacol. Exp. Ther.* **1968**, *160*, 189.
16. Stinson, S. C. *Chem. Eng. News* **2000**, *78*, 43.
17. González-Díaz, H.; Hernández-Sánchez, I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
18. Guye, P.-A. *Compt. Rend. (Paris)* **1890**, *110*, 714; Guye, P.-A. *Compt. Rend. (Paris)* **1893**, *116*, 1378, 1451, 1454; See also: Crum Brown, A. *Proc. R. Soc. Edinburgh* **1890**, *17*, 181.
19. Buda, A. B.; Mislow, K. *J. Mol. Struct. Theochem.* **1991**, *232*, 1.
20. Avnir, D.; Hel-Or, H. Z.; Mezey, P. G. Symmetry and Chirality: Continuous Measures. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: Chichester, 1998; Vol. 4, pp 2890–2901.
21. Zabrodsky, H.; Avnir, D. *J. Am. Chem. Soc.* **1995**, *117*, 462.
22. Benigni, R.; Cotta-Ramusino, M.; Gallo, G.; Giorgi, F.; Giuliani, A.; Vari, M. R. *J. Med. Chem.* **2000**, *43*, 3699.
23. Moreau, G. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 929.
24. Schultz, H. P.; Schultz, E. B.; Schultz, T. P. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 864.
25. de Julián-Ortiz, J. V.; de Alapont, C. G.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. *J. Mol. Graphics Mod.* **1998**, *16*, 14.
26. Golbraikh, A.; Bonchev, D.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147.
27. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1699.
28. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
29. Goodford, P. J. The Properties of Force Fields. In *QSAR and Molecular Modelling: Concepts Computational Tools and Biological Applications*; Sanz, F., Giraldo, J., Manaut, F., Eds.; Prous Science: Barcelona, 1995; pp 199–205.
30. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687. <http://www.mdpi.org>.
31. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; Ofori, E.; Montero, L. A. *Int. J. Mol. Sci.* **2003**, *4*, 512 [www.mdpi.org/ijms/](http://www.mdpi.org/ijms/).
32. Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. *CPS: physchem/0401004*.
33. Browder, A. *Mathematical Analysis. An Introduction*; Springer: New York, 1996; pp 176–296.
34. Axler, S. *Linear Algebra Done Right*; Springer: New York, 1996, pp 37–70.
35. Ross, K. A.; Wright, C. R. B. *Matemáticas Discretas*; Prentice may Hispanoamericana: Mexico, 1990.
36. Randić, M. *J. Math. Chem.* **1991**, *7*, 155.
37. Cotton, F. A. *Advanced Inorganic Chemistry*; Editorial Revolucionaria: Havana, Cuba, 1970; p 103.
38. Marrero-Ponce, Y.; Romero, V. *tomocomd* software. Central University of Las Villas, 2002. *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](mailto:yovanimp@qf.uclv.edu.cu); [ymarrero77@yahoo.es](mailto:ymarrero77@yahoo.es).
39. Vicent, M.; Marchand, B.; Rémond, G.; Jaquelin-Guinamant, S.; Damien, G.; Portevin, B.; Baupal, J.; Volland, J.; Bouchet, J.; Lambert, P.; Serkiz, B.; Luitjen, W.; Lauibie, M.; Schiavi, P. *Drug Des. Discov.* **1992**, *9*, 11–28.
40. STATISTICA ver. 5.5, Statsoft, 1999.
41. Ford, M.-G.; Salt, D.-W. The Use of Canonical Correlation Analysis. In *Methods and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, L., Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 283–292.
42. Wold, S.; Erikson, L. Statistical Validation of QSAR Results. Validation Tools. In *Methods and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, L., Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 309–318.
43. van de Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Methods and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, L., Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 265–288.
44. Belsey, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics*; Wiley: New York, 1980.
45. Golbraikh, A.; Tropsha, A. *J. Mol. Graphic Modell.* **2002**, *20*, 269.