ELSEVIER

Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification

Juan A. Castillo-Garit [a,b,*], Yovani Marrero-Ponce [b,c], Francisco Torrens [c], Richard Rotondo [d]

[a] Applied Chemistry Research Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba
[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] Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P.O. Box 22085, 46071 Valencia, Spain
[d] Mediscovery Inc., Suite 1050, 601 Carlson Parkway, Minnetonka, MN 55305, USA

Abstract

Non-stochastic and stochastic 2D bilinear indices have been generalized to codify chemical structure information for chiral drugs, making use of a trigonometric 3D-chirality correction factor. In order to evaluate the effectiveness of this novel approach in drug design we have modeled the angiotensin-converting enzyme inhibitory activity of perindoprilate's σ-stereoisomers combinatorial library. Two linear discriminant analysis models, using non-stochastic and stochastic linear indices, were obtained. The models had shown an accuracy of 95.65% for the training set and 100% for the external prediction set. Next the prediction of the σ-receptor antagonists of chiral 3-(3-hydroxyphenyl)piperidines by multiple linear regression analysis was carried out. Two statistically significant QSAR models were obtained when non-stochastic ($R^2 = 0.953$ and $s = 0.238$) and stochastic ($R^2 = 0.961$ and $s = 0.219$) 3D-chiral bilinear indices were used. These models showed adequate predictive power (assessed by the leave-one-out cross-validation experiment) yielding values of $q^2 = 0.935$ ($s_{cv} = 0.259$) and $q^2 = 0.946$ ($s_{cv} = 0.235$), respectively. Finally, the prediction of the corticosteroid-binding globulin binding affinity of steroids set was performed. The obtained results are rather similar to most of the 3D-QSAR approaches reported so far. The validation of this method was achieved by comparison with previous reports applied to the same data set. The non-stochastic and stochastic 3D-chiral linear indices appear to provide a very interesting alternative to other more common 3D-QSAR descriptors.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Non-stochastic and stochastic 3D-chiral bilinear indices; 3D-QSAR; Angiotensin-converting enzyme inhibitor; σ-Receptor antagonist; Steroid binding affinity; Discriminant analysis; Multiple linear regression

1. Introduction

It is well known that many biological molecules are asymmetric (or chiral), i.e., that they contain atoms in one of the two possible spatial configurations, which are mirror images of each other [1]. The non-superimposable mirror-image isomers are called enantiomers, but may also be referred to as enantiomorphs, optical isomers or optical antipodes [2]. The molecules with identical 2D structural formulas containing more than one asymmetric atom are referred to as σ-diastereomers [3]. In the literature, the asymmetric atoms are often referred to as chiral atoms and molecules containing chiral atoms are referred to as chiral molecules [4].

Chirality is a fundamental aspect of molecular structure, sometimes having profound influence on the properties of compounds. Enantiomers quite often exhibit different chemical and physical properties as well as different biological activities [5]. 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 process and phenomena are stereospecific. For

* Corresponding author at: Applied Chemistry Research Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba.
Tel.: +53 42 281192/281473; fax: +53 42 281130/281455.
E-mail addresses: juancg@uclv.edu.cu, jacgarit@yahoo.es, juancg.22@gmail.com (J.A. Castillo-Garit).

instance, L- and D-enantiomers of amino acids have different tastes [6,7], enantiomers of some compounds have different odours [8,9], and many medicinal preparations have physiological properties different from those of their enantiomers [10–12].

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 [2,5,13–21]. 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].

Some years ago, Buda and Mislow distinguished between two classes of measures [23]. In the first class 'the degree of chirality expresses the extent to which a chiral object differs from an achiral reference object'. In the second one 'it expresses the extent to which two enantiomorphs differ from one another'. Either method yields a single real value, usually an absolute quantity that is the same for both enantiomorphs. A different idea was to incorporate R/S labels into conventional topological indices [20]. Derived chirality descriptors were correlated with biological activity by de Julian-Ortiz et al. [16], Golbraikh et al. [18] and more recently by Diaz et al. [24]. One of the first approaches to in this field was introduced by de Julian-Ortiz et al. [16], in a study of the pharmacological activity of different pairs of enantiomers on dopamine D_2 and the σ-receptor. Fortunately, the so-called chiral topological indices (CTIs) are inexpensive in terms of computational time in comparison to grid dependent methods like CoMFA [25]. In any case, when chirality is considered many 3D-TIs become 'hard to interpret' in physical terms. For example, Golbraikh, Bonchev, and Tropsha's work generated even complex numbers that are incompatible with standard statistical software [18].

In two recent works of Aires-de-Sousa and Gasteiger [5,14], two different kinds of chirality codes were designed, named "conformation-independent chirality code" (CICC) and "conformation-dependent chirality code" (CDCC). The chirality code is a molecular transform that *represents* the chirality of a molecule by using a spectrum-like, fixed-length code and includes information about the geometry of chiral centers, properties of the atoms in the neighborhoods, and bonds lengths. The code distinguishes between enantiomers and yields descriptors with symmetrical values for opposite enantiomers [21].

More recently, one of the present authors (M.-P. Y.) has introduced new sets of total and atom-level molecular descriptors (MDs) relevant to QSAR/QSPR studies and 'rational' drug design, atom-based quadratic $q_k(x)$ and linear indices $f_k(x)$ [26–29]. These MDs are based on the calculation of quadratic and linear maps similar to those defined in linear algebra [30]. Our research team has focused its efforts mainly on their application in drug design, so that many new drug(lead)-like compounds within different pharmacological skills have been found out [31–37]. This approach has been successfully applied to the prediction of several physicochemical and pharmacokinetical properties of organic compounds [27,38–42]. In addition, these MDs have been extended to consider three-dimensional features of small/medium-sized molecules based on the trigonometric-3D-chirality-correction factor approach [4,28,43]. Other applications included studies related to nucleic acid–drug interactions [44,45] and related with structural characterization of proteins [46,47].

Moreover, other new set of total and atom-level molecular descriptors relevant to QSAR/QSPR studies and 'rational' drug design has been introduced [48]. The atom-based bilinear indices were defined in analogy to the bilinear mathematical maps. These new MDs have been successfully employed in the prediction of the chem-physical properties of organic compounds [48], and in the fast-track experimental discovery of novel tyrosinase inhibitors [49] as well as in the rational discovery of new trichomonacidal drugs.

The main aim of the present report is to extend 2D atom-based bilinear indices of the "non-stochastic and stochastic graph–theoretical electronic-density matrices", in order to codify chirality related structural features. The problem of classification of ACE (Angiotesin-converting enzyme) inhibitors, the prediction of σ-receptor antagonist activities and corticosteroid-binding globulin binding affinity of the Cramer's steroid data set are selected as 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.

2. Theoretical framework

In earlier publications, we outlined outstanding features concerned with the theory of 2D atom-based *TOMOCOMD-CARDD* descriptors. This method codifies the molecular structure by means of mathematical quadratic, linear and bilinear transformations. In order to calculate these algebraic maps for a molecule, the atom-based molecular vector, $\bar{x}$ (vector representation) and kth "non-stochastic and stochastic graph–theoretical electronic-density matrices", $\mathbf{M}^k$ and $\mathbf{S}^k$ correspondingly (matrix representations), are constructed [26–42,44]. Such atom-adjacency relationships and chemical-information codification will be applied in the present report to generate a series of atom-based MDs, *atom*, *group and atom-type as well as total bilinear indices*, to be used in drug design and chemoinformatic studies.

Therefore the structure of this section will be as follows: (1) a background in atom-based molecular vector and non-stochastic and stochastic graph–theoreticalal electronic-density matrices will be described in Sections 2.1 and 2.2, respectively; (2) an outline of the mathematical definition of bilinear maps will be develop in Sections 2.3–2.5, and (3) an introduction of a *trigonometric 3D-chirality correction factor* in molecular vector $\mathbf{X}$ in order to codify chirality-related structural features will be described in Section 2.6.

2.1. Chemical information and atom-based molecular vector

The atom-based molecular vector ($\bar{x}$) used to represent small-to-medium size organic chemicals has been explained elsewhere

in some detail [26–42,44]. The components (x) of $\bar{x}$ are numerical values, which represent a certain standard atomic property (atom label). Therefore, these weights correspond to different atomic properties for organic molecules. Thus, a molecule having 5, 10, 15, ..., n atomic nuclei can be represented by means of vectors, with 5, 10, 15, ..., n components, belonging to the spaces $\Re^5$, $\Re^{10}$, $\Re^{15}$, ..., $\Re^n$, respectively; where n is the dimension of the real set ($\Re^n$). Furthermore, $\bar{x}$ is the n-dimensional property vector of the atoms (atomic nuclei) in a molecule.

This approach allows us to encode organic molecules such as 3-mercapto-pyridine-4-carbaldehyde through the molecular vector $\bar{x} = [x_{N1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}, x_{C6}, x_{C7}, x_{O8}, x_{S9}]$ (see also Table 1 for molecular structure). This vector belongs to the product space $\Re^n$. However, diverse kinds of atomic weights (x) can be used for codifying information related to each atomic nucleus in the molecule. These atomic labels are chemically meaningful numbers such as atomic log P [50], surface area contributions of polar atoms [51], atomic contributions to molar refractivity [52], atomic hybrid polarizabilities [53], Gasteiger–Marsilli atomic charges [54], atomic masses (M) [55], the van der Waals volumes (V) [55], the atomic polarizabilities (P) [55], atomic electronegativities (E) in Pauling scale [56] and so on.

Now, if we are interested in codifying the chemical information by means of two different molecular vectors, for instance, $\bar{x} = [x_1, \ldots, x_n]$ and $\bar{y} = [y_1, \ldots, y_n]$; then different combinations of molecular vectors ($\bar{x} \neq \bar{y}$) are possible when a weighting scheme is used. In the present report, we characterized each atomic nucleus with the following parameters: atomic masses (M) [55], the van der Waals volumes (V) [55], the atomic polarizabilities (P) [55], and atomic electronegativities (E) in Pauling scale [56]. The values of these atomic labels are shown in Table 2. From this weighting scheme, 6 (or 12 if $\bar{x}_M$-$\bar{y}_V$ $\neq \bar{x}_V$-$\bar{y}_M$) combinations (pairs) of molecular vectors ($\bar{x}, \bar{y}; \bar{x} \neq \bar{y}$) can be computed, $\bar{x}_M$-$\bar{y}_V, \bar{x}_M$-$\bar{y}_P, \bar{x}_M$-$\bar{y}_E, \bar{x}_V$-$\bar{y}_P, \bar{x}_V$-$\bar{y}_E$, and $\bar{x}_P$-$\bar{y}_E$. Here, we used the symbols $\bar{x}_W$-$\bar{y}_Z$, where the subscripts W and Z mean any pair of atomic properties from our weighting scheme and a hyphen (-) expresses the combination (pair) of two selected atom-label chemical properties. In order to illustrate this, let us consider the same organic molecule as in the example above, (3-mercapto-pyridine-4-carbaldehyde) and the following weighting scheme: M and V ($\bar{x}_M$-$\bar{y}_V = \bar{x}_V$-$\bar{y}_M$). The following molecular vectors, $\bar{x} = [14.01, 12.01, 12.01, 12.01, 12.01, 12.01, 12.01, 16.0, 32.07]$ and $\bar{y} = [15.599, 22.449, 22.449, 22.449, 22.449, 22.449, 22.449, 11.494, 24.429]$ are obtained when we use M and V as chemical weights for codifying each atom in the example molecule in the vectors $\bar{x}$ and $\bar{y}$, respectively.

2.2. Background in non-stochastic and stochastic graph–theoretical electronic-density matrices

In molecular topology, molecular structure is expressed, generally, by the hydrogen-suppressed graph. Therefore, a molecule is represented by a graph. Informally a graph G is a collection of vertices (points) and edges (lines or bonds) connecting these vertices [57–59]. In more formal terms, a simple graph G is defined as an ordered pair [V(G), E(G)] which consists of a nonempty set of vertices V(G) and a set

E(G) of unordered pairs of elements of V(G), called edges [57–59]. In this particular case we are not dealing with a simple graph but with a so-called pseudograph (G). Informally, a pseudograph is a graph with multiple edges or loops between the same vertices or the same vertex. Formally: a pseudograph is a set V of vertices along a set E of edges, and a function f from E to $\{\{u,v\}|\ u,v\ \text{in}\ V\}$ (The function f shows which pair of vertices are connected by which edge). An edge is a loop if $f(e) = \{u\}$ for some vertex u in V [26,27,60].

In earlier publications we have introduced new molecular matrices that describe changes along the time in the electronic distribution throughout the molecular backbone. The $n \times n$ kth non-stochastic graph–theoretical electronic-density matrix of the molecular pseudograph (G), $\mathbf{M}^k$, is a square and symmetric matrix, where n is the number of atoms (atomic nuclei) in the molecule [26–42,44]. The coefficients $^k m_{ij}$ are the elements of the kth power of $\mathbf{M}$(G) and are defined as follows [26–42,44]:

$$m_{ij} = \begin{cases} P_{ij} & \text{if } i \neq j \text{ and } \exists\, e_k \in \text{E(G)} \\ L_{ii} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where E(G) represents the set of edges of G. P_{ij} is the number of edges (bonds) between vertices (atomic nuclei) v_i and v_j, and L_{ii} is the number of loops in v_i.
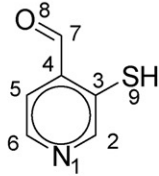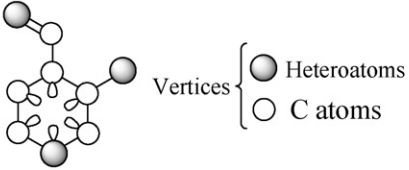
The elements $m_{ij} = P_{ij}$ of such a matrix represent the number of chemical bonds between an atomic nucleus i and other j. The matrix $\mathbf{M}^k$ provides the numbers of walks of length k that link every pair of vertices v_i and v_j. For this reason, each edge in $\mathbf{M}^1$ represents 2 electrons belonging to the covalent bond between atomic nuclei i and j; e.g. the inputs of $\mathbf{M}^1$ are equal to 1, 2 or 3 when a single, double or triple bond, correspondingly, appears between vertices v_i and v_j. On the other hand, molecules containing aromatic rings with more than one canonical structure are represented by a pseudograph. It happens for substituted aromatic compounds such as pyridine, naphthalene, quinoline, and so on, where the presence of pi (π) electrons is accounted by means of loops in each atomic nucleus of the aromatic ring. Conversely, aromatic rings having only one canonical structure, such as furan, thiophene and pyrrol are represented by a multigraph. In order to illustrate the calculation of these matrices, let us consider the same molecule selected in the previous section. Table 1 depicts the molecular structure of this compound and its labeled molecular pseudograph. The zero ($k = 0$), first ($k = 1$), second ($k = 2$) and third ($k = 3$) powers of the non-stochastic graph–theoretical electronic-density matrices are also given in this table.

As can be seen, $\mathbf{M}^k$ is graph–theoretical electronic–structure models, like in "extended Hückel theory (EHT) model". The $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ- and π-networks) in one step and its powers ($k = 0, 1, 2, 3, \ldots$) can be considered as interacting–electron chemical–network models in the k step. The complete model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas [61].

The present approach is based on a simple model for the intramolecular movement of all outer-shell electrons. Let us

Table 1
(A) Chemical structure of 3-mercapto-pyridine-4-carbaldehyde and its labeled molecular pseudograph, G. (B) and (C) the zero ($k = 0$), first ($k = 1$), second ($k = 2$) and third ($k = 3$) powers of the non-stochastic and stochastic graph–theoretical electronic-density matrices of G, respectively

A)	molecular structure	molecular pseudograph (H-atoms-suppressed pseudograph)[a]

B) k^{th} non-stochastic graph–theoretical electronic-density matrices, $\mathbf{M}^k$ ($k = 0$-3)

zero order ($k = 0$)[b]

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

first order ($k = 1$)

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

second order ($k = 2$)

$$\begin{bmatrix} 3 & 2 & 1 & 0 & 1 & 2 & 0 & 0 & 0 \\ 2 & 3 & 2 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 2 & 4 & 2 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

third order ($k = 3$)

$$\begin{bmatrix} 7 & 6 & 3 & 2 & 3 & 6 & 0 & 0 & 1 \\ 6 & 7 & 7 & 3 & 2 & 3 & 1 & 0 & 2 \\ 3 & 7 & 9 & 8 & 3 & 2 & 2 & 2 & 4 \\ 2 & 3 & 8 & 9 & 7 & 3 & 8 & 2 & 2 \\ 3 & 2 & 3 & 7 & 7 & 6 & 2 & 2 & 1 \\ 6 & 3 & 2 & 3 & 6 & 7 & 1 & 0 & 0 \\ 0 & 1 & 2 & 8 & 2 & 1 & 1 & 10 & 1 \\ 0 & 0 & 2 & 2 & 2 & 0 & 10 & 0 & 0 \\ 1 & 2 & 4 & 2 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

C) k^{th} stochastic graph–theoretical electronic-density matrices, $\mathbf{S}^k$ ($k = 1$-3)[c]

first order ($k = 1$)

$$\begin{bmatrix} 0.3333 & 0.3333 & 0 & 0 & 0 & 0.3333 & 0 & 0 & 0 \\ 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0.3333 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0 & 0.6666 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

second order ($k = 2$)

$$\begin{bmatrix} 0.3333 & 0.2222 & 0.1111 & 0 & 0.1111 & 0.2222 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0 & 0.1 & 0 & 0 & 0.1 \\ 0.0833 & 0.166 & 0.3333 & 0.1666 & 0.0833 & 0 & 0.0833 & 0 & 0.0833 \\ 0 & 0.0714 & 0.1429 & 0.2857 & 0.1429 & 0.0714 & 0.0714 & 0.1429 & 0.0714 \\ 0.1 & 0 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0 & 0 \\ 0.2222 & 0.1111 & 0 & 0.1111 & 0.2222 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 0 & 0.625 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0 & 0.6666 & 0 \\ 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}$$

third order ($k = 3$)

$$\begin{bmatrix} 0.25 & 0.2142 & 0.1071 & 0.0714 & 0.1071 & 0.2143 & 0 & 0 & 0.0357 \\ 0.1935 & 0.2258 & 0.2258 & 0.0967 & 0.0645 & 0.0967 & 0.0323 & 0 & 0.0645 \\ 0.075 & 0.175 & 0.225 & 0.2 & 0.075 & 0.05 & 0.05 & 0.05 & 0.1 \\ 0.0455 & 0.0682 & 0.1818 & 0.2045 & 0.1591 & 0.0682 & 0.1818 & 0.0455 & 0.0455 \\ 0.0909 & 0.0606 & 0.0909 & 0.2121 & 0.2121 & 0.1818 & 0.0606 & 0.0606 & 0.0303 \\ 0.2143 & 0.1071 & 0.0714 & 0.1971 & 0.2143 & 0.25 & 0.0357 & 0 & 0 \\ 0 & 0.0385 & 0.0769 & 0.3076 & 0.0769 & 0.0385 & 0.0385 & 0.3846 & 0.0385 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 0 & 0.625 & 0 & 0 \\ 0.0833 & 0.1666 & 0.333 & 0.1666 & 0.0833 & 0 & 0.0833 & 0 & 0.0833 \end{bmatrix}$$

[a] Each edge in $\mathbf{M}^1$ represents 2 electrons belonging to the covalent bond between atoms (vertices) v_i and v_j; e.g. the inputs of $\mathbf{M}^1$ are equal to 1, 2 or 3 when single, double or triple bonds, correspondingly, appears between vertices v_i and v_j. The presence of pi (π) electrons in aromatic systems such as benzene is accounted by means of loops in each atom of the aromatic ring. Therefore, the $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ- and π-networks) in one step and their powers ($k = 0, 1, 2, 3,\ldots$) can be considered as an interacting–electron chemical–network model in k step. [b] The zero power ($k = 0$) of the stochastic graph–theoretical electronic-density matrix, $\mathbf{S}^0$, coincides with non-stochastic matrix one ($\mathbf{M}^0 = \mathbf{S}^0$). [c] The values of the elements of kth matrices $\mathbf{S}^k$ ($^k s_{ij}$) have been approximated.

Table 2
Values of the atomic weights used for bilinear indices calculation [55,56,76,77]

ID	Atomic mass	VdW[a] volume (Å^3)	Polarizability (Å^3)	Pauling electronegativity
H	1.01	6.709	0.667	2.2
B	10.81	17.875	3.030	2.04
C	12.01	22.449	1.760	2.55
N	14.01	15.599	1.100	3.04
O	16.00	11.494	0.802	3.44
F	19.00	9.203	0.557	3.98
Al	26.98	36.511	6.800	1.61
Si	28.09	31.976	5.380	1.9
P	30.97	26.522	3.630	2.19
S	32.07	24.429	2.900	2.58
Cl	35.45	23.228	2.180	3.16
Fe	55.85	41.052	8.400	1.83
Co	58.93	35.041	7.500	1.88
Ni	58.69	17.157	6.800	1.91
Cu	63.55	11.494	6.100	1.9
Zn	65.39	38.351	7.100	1.65
Br	79.90	31.059	3.050	2.96
Sn	118.71	45.830	7.700	1.96
I	126.90	38.792	5.350	2.66

[a] VdW: van der Waals.

consider a hypothetical situation in which a set of atoms is free in space at an arbitrary initial time (t_0). At this time, the electrons are distributed around the atomic nuclei. Alternatively, these electrons can be distributed around cores in discrete intervals of time t_k. Therefore, an the electron in an arbitrary atom i can move (step-by-step) to other atoms at different discrete time periods t_k ($k = 0, 1, 2, 3,...$) throughout the chemical-bonding network.

On the other hand, the kth stochastic graph–theoretical electronic-density matrix of **G**, $\mathbf{S}^k$, can be obtained directly from $\mathbf{M}^k$. Here, $\mathbf{S}^k = [^k s_{ij}]$, is a square matrix of order n (n = number of atomic nuclei) and the elements $^k s_{ij}$ are defined as follows [32,33,35,36]:

$$^k\mathbf{s}_{ij} = \frac{^k m_{ij}}{^k\text{SUM}_i} = \frac{^k m_{ij}}{^k\delta_i} \tag{2}$$

where $^k m_{ij}$ are the elements of the kth power of M and the SUM of the ith row of $\mathbf{M}^k$ is named the k-order vertex degree of atom i, $^k\delta_i$. It should be remarked that the matrix $\mathbf{S}^k$ in Eq. (2) has the property that *the sum of the elements in each row* is 1. An $n \times n$ matrix with nonnegative entries having this property is called a "stochastic matrix" [30]. The kth s_{ij} elements are the transition probabilities with the electrons moving from atom i to j in the discrete time periods t_k. It should be also pointed out that the kth element s_{ij} takes into consideration the molecular topology in the k step throughout the chemical-bonding (σ- and π) network. Therefore, the $^2 s_{ij}$ values can distinguish between hybrid states of atoms in bonds. For instance, the self-return probability of second order ($^2 s_{ii}$) [i.e., the probability with which an electron returns to its at t_2], varies regularly according to the different hybrid states of atom i in the molecule, e.g. an electron will have a higher probability of returning to the sp C atom than to the sp^2 (or sp^3) C atom in t_2 [$p(C_{sp})$

$> p(C_{sp}^2) > p(C_{sp\ arom}^2) > p(C_{sp}^3)]$ (see Table 1 for more details). This is a logical result if the electronegativity scale of these hybrid states is taken into account.

2.3. Mathematical bilinear forms: a theoretical framework

In mathematics, a *bilinear form* in a real vector space is a mapping $b: V \times V \to \mathfrak{R}$, which is linear in both arguments [62–67]. Therefore, this function satisfies the following axioms for any scalar α and any choice of vectors $\bar{v}, \bar{w}, \bar{v}_1, \bar{v}_2, \bar{w}_1$ and $\bar{w}_2$:

(i) $b(\alpha\bar{v}, \bar{w}) = b(\bar{v}, \alpha\bar{w}) = \alpha b(\bar{v}, \bar{w})$
(ii) $b(\bar{v}_1 + \bar{v}_2, \bar{w}) = b(\bar{v}_1, \bar{w}) + b(\bar{v}_2, \bar{w})$
(iii) $b(\bar{v}, \bar{w}_1 + \bar{w}_2) = b(\bar{v}, \bar{w}_1) + b(\bar{v}, \bar{w}_2)$

Namely, b is *bilinear* if it is linear in each parameter, taken separately.

Let V be a real vector space in $\mathfrak{R}^n$ ($V \in \mathfrak{R}^n$) and consider that the following vector set, $\{\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n\}$ is a basis set of $\mathfrak{R}^n$. This basis set permits us to write in unambiguous form any vectors $\bar{x}$ and $\bar{y}$ of V, where $(x^1, x^2, \ldots, x^n) \in \mathfrak{R}^n$ and $(y^1, y^2, \ldots, y^n) \in \mathfrak{R}^n$ are the coordinates of the vectors $\bar{x}$ and $\bar{y}$, respectively. Therefore,

$$\bar{x} = \sum_{i=1}^{n} x^i \bar{e}_i \tag{3}$$

and,

$$\bar{y} = \sum_{i=1}^{n} y^j \bar{e}_j \tag{4}$$

subsequently,

$$b(\bar{x}, \bar{y}) = b(x^i \bar{e}_i, y^j \bar{e}_j) = x^i y^j b(\bar{e}_i, \bar{e}_j) \tag{5}$$

if we take the a_{ij} as the $n \times n$ scalars $b(\bar{e}_i, \bar{e}_j)$, i.e.,

$$a_{ij} = b(\bar{e}_i, \bar{e}_j), \text{to } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, n \tag{6}$$

Then,

$$b(\bar{x}, \bar{y}) = \sum_{i,j}^{n} a_{ij} x^i y^j = [X]^{\text{T}} A[Y]$$

$$= [x^1 \quad \ldots \quad x^n] \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix} \tag{7}$$

As it can be seen, the defined equation for b can be written in matrix notation (see Eq. (7)), where $[Y]$ is a column vector (an $n \times 1$ matrix) of the coordinates of $\bar{y}$ in a basis set of $\mathfrak{R}^n$, and $[X]^{\text{T}}$ (a $1 \times n$ matrix) is the transpose of $[X]$, where $[X]$ is a column vector (an $n \times 1$ matrix) of the coordinates of $\bar{x}$ in the same basis of $\mathfrak{R}^n$.

Finally, we introduce the formal definition of symmetric bilinear form. Let V be a real vector space and b be a bilinear function in $V \times V$. The bilinear function b is called symmetric if $b(\bar{x}, \bar{y}) = b(\bar{y}, \bar{x})$, $\forall \bar{x}, \bar{y} \in V$ [62–67]. Then,

$$b(\bar{x}, \bar{y}) = \sum_{i,j}^{n} a_{ij} x^i y^j = \sum_{i,j}^{n} a_{ji} x^j y^i = b(\bar{y}, \bar{x}) \tag{8}$$

2.4. Non-stochastic and stochastic atom-based bilinear indices: total definition

The kth non-stochastic and stochastic bilinear indices for a molecule, $b_k(\bar{x}, \bar{y})$ and $^s b_k(\bar{x}, \bar{y})$, are computed from these kth non-stochastic and stochastic graph–theoretical electronic-density matrices, $\mathbf{M}^k$ and $\mathbf{S}^k$ as shown in Eqs. (9) and (10), respectively:

$$b_k(\bar{x}, \bar{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k m_{ij} x^i y^j \tag{9}$$

$$^s b_k(\bar{x}, \bar{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k s_{ij} x^i y^j \tag{10}$$

where n is the number of atoms in the molecule, and $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$ are the coordinates or components of the molecular vectors $\bar{x}$ and $\bar{y}$ in a canonical basis set of $\Re^n$.

The defined Eqs. (9) and (10) for $b_k(\bar{x}, \bar{y})$ and $^s b_k(\bar{x}, \bar{y})$ can also be written in matrix notation:

$$b(\bar{x}, \bar{y}) = [X]^{\mathrm{T}} \mathbf{M}^k [Y] \tag{11}$$

$$^s b(\bar{x}, \bar{y}) = [X]^{\mathrm{T}} \mathbf{S}^k [Y] \tag{12}$$

where $[Y]$ is a column vector (an $n \times 1$ matrix) of the coordinates of $\bar{y}$ in the canonical basis set of $\Re^n$, and $[X]^{\mathrm{T}}$ is the transpose of $[X]$, where $[X]$ is a column vector (an $n \times 1$ matrix) of the coordinates of $\bar{x}$ in the canonical basis of $\Re^n$. Therefore, if we use the canonical basis set, the coordinates $[(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)]$ of any molecular vectors ($\bar{x}$ and $\bar{y}$) coincide with the components of those vectors $[(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)]$. For that reason, those coordinates can be considered as weights (atomic labels) of the vertices of the molecular pseudograph, due to the fact that components of the molecular vectors are values of some atomic property that characterizes each kind of atomic nuclei in molecule.

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

2.5. Non-stochastic and stochastic atom-based bilinear indices: local (atomic, group, and atom-type) definition

In the last decade, Randić [68] proposed a list of desirable attributes for a MD. Therefore, this list can be considered as a methodological guide for the development of new topological indices (TIs). One of the most important criteria is the possibility of defining the descriptors locally. This attribute refers to the fact that the index could be calculated for the molecule as a whole but also over certain fragments of the structure itself.

Sometimes, the properties of a group of molecules are more related to a certain zone or fragment than to the molecule as a whole. Thereinafter, the global definition never satisfies the structural requirements needed to obtain a good correlation in QSAR and QSPR studies. The local indices can be used in certain problems such as:

- Research on drugs, toxics or generally any organic molecules with a common skeleton, which is responsible for the activity or property under study.
- Study on the reactivity of specific sites of a series of molecules, which can undergo a chemical reaction or enzymatic metabolism.
- In the study of molecular properties such as spectroscopic measurements, which are obtained experimentally in a local way.

In any general case where it is necessary to study not the molecule as a whole, but rather some local properties of certain fragments, then the definition of local descriptors could be necessary.

Therefore, in addition to *total bilinear indices* computed for the whole molecule, local-fragment (atomic, group or atom-type) formalism can be developed. These descriptors are termed *local non-stochastic and stochastic bilinear indices*, $b_{kL}(\bar{x}, \bar{y})$ and $^s b_{kL}(\bar{x}, \bar{y})$, respectively. The definition of these descriptors is as follows:

$$b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k m_{ijL} x^i y^j \tag{13}$$

$$^s b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k s_{ijL} x^i y^j \tag{14}$$

where $^k m_{ijL}$ [$^k s_{ijL}$] is the kth element of the row "i" and column "j" of the local matrix $\mathbf{M}_{\mathrm{L}}^k [\mathbf{S}_{\mathrm{L}}^k]$. This matrix is extracted from the $\mathbf{M}^k [\mathbf{S}^k]$ matrix and contains information referred to the pairs of vertices (atomic nuclei) of the specific molecular fragments and also of the molecular environment in k step.

The matrix $\mathbf{M}_{\mathrm{L}}^k [\mathbf{S}_{\mathrm{L}}^k]$ with elements $^k m_{ijL}$ [$^k s_{ijL}$] is defined as follows:

$$^k m_{ijL} [{}^k s_{ijL}]$$
$$= {}^k m_{ij} [{}^k s_{ijL}] \text{ if both } v_i \text{ and } v_j \text{ are atomic nuclei}$$
$$\text{contained within the molecular fragment}$$
$$= (1/2)^k m_{ij} [{}^k s_{ijL}] \text{ either } v_i \text{ or } v_j \text{ is an atomic nucleus}$$
$$\text{contained within the molecular fragment}$$
$$= 0 \text{ otherwise} \tag{15}$$

These local analogues can also be expressed in matrix form by the expressions:

$$b_{\mathrm{L}}(\bar{x}, \bar{y}) = [X]^{\mathrm{T}} \mathbf{M}_{\mathrm{L}}^k [Y] \tag{16}$$

$$^s b_{\mathrm{L}}(\bar{x}, \bar{y}) = [X]^{\mathrm{T}} \mathbf{S}_{\mathrm{L}}^k [Y] \tag{17}$$

It should be remarked that the scheme above follows the spirit of a Mulliken population analysis [69]. It should be also pointed out that for every partitioning of a molecule into Z molecular fragments there will be Z local molecular fragment matrices. Therefore, if a molecule is partitioned into Z molecular fragments, the matrix $\mathbf{M}^k$ [$\mathbf{S}^k$] can be correspondingly partitioned into Z local matrices $\mathbf{M}_L^k$[$\mathbf{S}_L^k$], $L = 1, \ldots, Z$, and the kth power of matrix $\mathbf{M}$ [$\mathbf{S}$] is exactly the sum of the kth power of the local Z matrices. Moreover, the total non-stochastic and stochastic bilinear indices are the sum of the non-stochastic and stochastic bilinear indices, respectively, of the Z molecular fragments:

$$b(\bar{x}, \bar{y}) = \sum_{L=1}^{Z} b_{kL}(\bar{x}, \bar{y}) \qquad (18)$$

$$^s b(\bar{x}, \bar{y}) = \sum_{L=1}^{Z} {}^s b_{kL}(\bar{x}, \bar{y}) \qquad (19)$$

Atomic, group and atom-type bilinear fingerprints are specific cases of local bilinear indices. Atomic bilinear indices, $b_{kL}(\bar{x}_i, \bar{y}_i)$ and $^s b_{kL}(\bar{x}_i, \bar{y}_i)$, can be computed for each atom i in the molecule and contain electronic and topological structural information from all other atoms within the structure. The atom-level bilinear indices values for the common scaffold atoms can be used directly as variables in seeking a QSPR/QSAR model as long as these atoms are numbered in the same way in all molecules in the database.

In addition, the atom-type bilinear indices can also be calculated. In the same way as atom-type E-state values [70], for all data sets (including those with a common skeletal core as well as those with very diverse structures), these novel local MDs provide much useful information. Therefore, this approach provides the basis for application to a wider range of problems to which the atomic bilinear indices formalism is applicable without the need for superposition [71,72]. Moreover, the present method represents a significant advantage over traditional QSAR methods. The atom-type bilinear descriptors are calculated by adding the kth atomic bilinear indices for all atoms of the same type in the molecule. This atom type index lends itself to be use in a group additive-type scheme in which an index appears for each atom type in the molecule. In the atom-type bilinear indices formalism, each atom in the molecule is classified into an atom type (fragment), such as –F, –OH, =O, –CH$_3$, and so on [70–72]. Therefore, each atom in the molecule is categorized according to a valence-state classification scheme including the number of attached H-atoms [70]. The atom-type descriptors combine three important aspects of structural information: (1) collective electronic and topological accessibility to the atoms of the same type (for each structural feature: atom or hybrid group such as –Cl, =O, –CH$_2$–, etc.), (2) presence/absence of the atom type (structural features), and (3) count of the atoms in the atom-type sets.

Finally, these local MDs can be calculated by a chemical (or functional) group in the molecule, such as heteroatoms (O, N and S in all valence states and including the number of attached H-atoms), hydrogen bonding (H-bonding) to heteroatoms (O, N and S in all valence states), halogen atoms (F, Cl, Br and I), all

aliphatic carbon chains (several atom types), all aromatic atoms (aromatic rings), and so on. The group-level bilinear indices are the sum of the individual atom-level bilinear indices for a particular group of atoms. For all data set structures, the kth group-based bilinear indices provide important information for QSAR/QSPR studies.

2.6. 3D-chiral bilinear indices

The total and local bilinear 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 $\bar{x}$ [4]. Therefore, a chirality molecular vector is obtained ($^*\bar{x}$), where the components of $\bar{x}$ (for instance, Pauling electronegativity (x_A) [56] of the A atom) are substituted by the following term [$x_A + \sin((\omega_A + 4\Delta)\pi/2)$].

The *trigonometric 3D-chirality correction factor* uses a dummy variable, ω_A and an integer parameter, Δ [4,24,28]:

$$
\begin{aligned}
\omega_A &= 1 \text{ and } \Delta \text{ is an odd number when A has R(rectus)}, \\
&\quad \text{E(entgegen), or } a \text{ (axial) notation according} \\
&\quad \text{to Cahn-Ingold} - \text{Prelog(CIP)} - \text{IUPAC rules} \\
&= 0 \text{ and } \Delta \text{ is an even number, if A does not have} \\
&\quad \text{3D specific enviroment} \\
&= -1 \text{ and } \Delta \text{ is an odd number when A has S(sinister)}, \\
&\quad \text{Z(zusammen), or e(ecuatorial) notation according} \\
&\quad \text{to CIP rules} \qquad (20)
\end{aligned}
$$

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 the fact that chirality does not change the electronegativities of atoms in the molecule in an isotropic environment in an observable way [73]. This correction has principally a mathematical means and must not be source of any misunderstanding. Therefore, this approach can be seen as a simplification of molecular structure. However, in other level of theoretical chemistry this procedure has also been used. As was recalled by Dewar almost 20 years ago, the Schrödinger equation is not exact; it is only an approximation where electron spin is incorporated in the results only as an artifact [74].

A severe limitation of the Golbraikh–Bonchev–Tropsha (GBT) approach is the existence of different chirality corrections and we had great difficulty in selecting one of these. Therefore, Diaz et al. [24] introduced an exponential chirality factor ($\exp(\omega_A\Delta)$), which eliminated indetermination in the selection of chirality and 3D scales for stochastic topological 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 topological indices (GBT-like

Table 3

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

ω_A	Δ														
	−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

ones). Table 3 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). Table 3 clearly shows that the trigonometric 3D-chirality factor is invariant with respect to the selection of all possible real scales. Therefore, 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-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 [16,18,24].

A very interesting point is that the present 3D-chiral descriptor reduces to simple two-dimensional (2D) bilinear indices for molecules without specific 3D characteristics because $\sin(0 + 4\Delta)\pi/2 = 0$, being Δ zero or any even number. Therefore, when all the atoms in the molecule are not chiral, the *TOMOCOMD-CARDD* (acronym of *TO*pological *MO*lecular *COM*puter *D*esign) *C*omputed-*A*ided '*R*ational' *D*rug *D*esign) molecular descriptors or any GBT-like chiral topological index do not change upon the introduction of this factor. This means that $^*\bar{x} = \bar{x}$ and thus, $^*\boldsymbol{b}_k(\bar{x}, \bar{y}) = \boldsymbol{b}_k(\bar{x}, \bar{y})$.

3. Methods

3.1. TOMOCOMD-CARDD approach

For computation of atom-based 3D-chiral (2.5) bilinear indices we used *TOMOCOMD* software [75]. It is an interactive program for molecular design and bioinformatics research, which contains four subprograms: *CARDD*, *CAMPS* (*C*omputed-*A*ided *M*odeling in *P*rotein *S*cience), *CANAR* (*C*omputed-*A*ided *N*ucleic *A*cid *R*esearch), and *CABPD* (*C*omputed-*A*ided *B*io-*P*olymers *D*ocking). It is composed of four subprograms; each one of them allows both drawing the structures (drawing mode) and calculating molecular 2D/3D descriptors (calculation mode). 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 3D-chiral atom-based bilinear indices considering and not considering H-atoms in the molecular pseudograph (G).

The main steps for the application of the present method in QSAR/QSPR and drug design can be summarized briefly in the following algorithm:

(1) Draw the molecular structure for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol

belonging to the different groups in the periodic table of the elements;

(2) Use appropriate weights in order to differentiate the atoms in the molecule. The weights used in this work are those previously proposed for the calculation of the DRAGON descriptors [55,76,77], i.e., atomic mass (M), atomic polarizability (P), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (E). The values of these atomic labels are shown in Table 2 [55,56,76,77];

(3) Compute the total and local (atomic, group and atom-type) non-stochastic and stochastic bilinear indices. It can be carried out in the software calculation mode, where one can select the atomic properties and the descriptor family before calculating the molecular indices. This software generates a table in which the rows correspond to the compounds, and columns correspond to the atom-based (both total and local) bilinear maps or other MDs family implemented in this program;

(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 a property P and the bilinear fingerprints having, for instance, the following appearance,

$$\mathbf{P} = a_0\boldsymbol{b}_0(\bar{x}, \bar{y}) + a_1\boldsymbol{b}_1(\bar{x}, \bar{y}) + a_2\boldsymbol{b}_2(\bar{x}, \bar{y}) + \cdots$$
$$+ a_k\boldsymbol{b}_k(\bar{x}, \bar{y}) + c \qquad (21)$$

where **P** is the measured property, $\boldsymbol{b}_k(\bar{x}, \bar{y})$, are the *k*th non-stochastic total bilinear indices, and the $\boldsymbol{a}_k$'s and c are the coefficients obtained by the MRA;

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

The following descriptors were calculated in this work:

(i) *k*th total 3D-chiral bilinear indices not considering and considering H atoms in the molecular pseudograph (G) [$^*\boldsymbol{b}_k(\bar{x}, \bar{y})$] and $^*\boldsymbol{b}_k{}^H(\bar{x}, \bar{y})$, respectively].

(ii) *k*th local (atomic group = heteroatoms: S, N, O) 3D-chiral bilinear indices not considering and considering H atoms in the molecular pseudograph (G) [$^*\boldsymbol{b}_{kL}(\bar{x}_E, \bar{y}_E)$ and $^*\boldsymbol{b}_{kL}^H(\bar{x}_E, \bar{y}_E)$], correspondingly]. These local descriptors are putative H-bonding acceptors; in addition, charge as well as dipole moment.

(iii) *k*th local (atomic group = H-atoms bonding to heteroatoms: S, N, O) 3D-chiral bilinear indices considering H

atoms in the molecular pseudograph (G) $[^{*}\boldsymbol{b}_{kL}^{H}(\bar{x}_{\text{E-H}}, \bar{y}_{\text{E-H}})]$. These local descriptors are putative H-bonding donors.

The kth stochastic total $[^{*s}\boldsymbol{b}_{k}(\bar{x}, \bar{y})$ and $^{*s}\boldsymbol{b}_{k}^{H}(\bar{x}, \bar{y})]$ and local $[^{*s}\boldsymbol{b}_{kL}(\bar{x}_{\text{E}}, \bar{y}_{\text{E}}), \, ^{*s}\boldsymbol{b}_{kL}^{H}(\bar{x}_{\text{E}}, \bar{y}_{\text{E}}),$ and $^{*s}\boldsymbol{b}_{kL}^{H}(\bar{x}_{\text{E-H}}, \bar{y}_{\text{E-H}})]$ 3D-chiral bilinear indices were also computed.

3.2. Chemometric analysis

Statistical analysis was carried out with the STATISTICA software [78]. 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 parsimony (Occam's razor) was taken into account as strategy for model selection. Therefore, we selected the model with a highest statistical signification but having as few parameters (a_k) as possible.

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 $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$ or as inactive otherwise. $P(\text{active})$ and $P(\text{inactive})$ 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 [79]. MCC is always 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 respect to random baseline. Therefore, MCC quantifies the strength of the linear relation between the molecular descriptors and the classifications [79], and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages.

On the other hand, *multiple linear regression* (MLR) was carried out to predict σ-receptor antagonist activities of 3-(3-hydroxyphenyl)piperidines and the corticosteroid-binding globulin (CBG) binding affinity of a steroid data set. The quality of the models was determined by examining the regression's statistic parameters and those of the cross-validation procedures [80,81]. Therefore, the quality of models was determined by examining the determination coefficients (also know as square correlation coefficient; R^2), Fisher-ratio's p-level [$p(F)$], standard deviations of the regression (s) and the leave-*one*-out (LOO) press statistics (q^2, s_{cv}) analogues to R^2 and s [80,82].

4. QSAR applications and comparison with other approaches

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

4.1. Classification of the ACE inhibitory activity of 32 perindopirilate's σ-stereoisomers

A recently introduced data set of 32 perindoprilate stereoisomers, angiotensin-converting enzyme (ACE) inhibitors [24,83], was used to test the applicability of the method. ACE acts in plasma and blood vessels, removing the C-terminal dipeptide residue of undecapeptide 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) [83]. In this study, "active" is taken to mean a compound that has an IC_{50} value not higher than 110 nM.

In this application, we tested the predictive power of 3D-chiral bilinear indices in the classification of perindopirilate stereoisomers. The obtained classification models are given below together with the LDA statistical parameters:

ACEi-actv
$$= -155.66 + 7.55 \times 10^{-5 \text{M-P}} \boldsymbol{b}_{8}^{H}(\bar{x}, \bar{y})$$
$$- 3.8 \times 10^{-7 \text{M-E}} \boldsymbol{b}_{14}(\bar{x}, \bar{y}), \qquad N = 23, \quad \lambda = 0.398,$$
$$D^2 = 7.164, \quad F(2, 20) = 15.128, \quad p < 0.0001 \qquad (22)$$

ACEi-actv $= 247.41 - 0.664^{*\text{M-E}s} \boldsymbol{b}_{14}(\bar{x}, \bar{y})$
$$+ 0.746^{*\text{M-P}s} \boldsymbol{b}_{10}^{H}(\bar{x}, \bar{y}),$$
$$N = 23, \quad \lambda = 0.393, \quad D^2 = 7.30,$$
$$F(2, 20) = 15.417, \quad p < 0.0001 \qquad (23)$$

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

The model (22), which includes non-stochastic 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 statistic for external

Table 4

Basic structure and chirality notation of active and non-active perindoprilate stereoisomers with their posterior probabilities in data split into training and test sets, using non-stochastic and stochastic bilinear indices
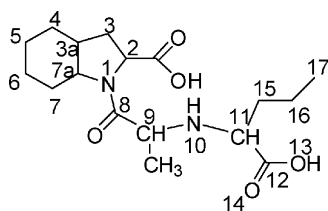
No.	Comp.[a]	Class[b]	IC_{50}[c]	Eq. (22) (non-stochastic)		Eq. (23) (stochastic)	
				Class	$\Delta P\%$[d]	Class	$\Delta P\%$[d]
Active compounds							
1	SSRSS[e]	+	1.1	+	84.00	+	75.80
2	SRSSS[e]	+	1.2	+	98.27	+	99.25
3	SSSSS	+	1.5	+	93.96	+	94.66
4	SRRSS[e]	+	3.3	+	95.00	+	95.96
5	SSSSR	+	12.2	+	91.91	+	90.38
6	SSRSR	+	29.4	+	78.76	+	59.15
7	SRRSR	+	39.8	+	92.97	+	92.13
8	SRSSR	+	54	+	97.57	+	98.53
9	RRSSS	+	108	−	−54.84	−	−19.11
Non-active compounds							
10	SSSRS	−	1.1×10^3	−	−80.23	−	−80.21
11	RSSSS	−	1.9×10^3	−	−86.78	−	−84.10
12	SSRRR[e]	−	2.6×10^3	−	−94.33	−	−97.69
13	RRSSR	−	5.5×10^3	−	−65.97	−	−48.98
14	SSRRS	−	7.1×10^3	−	−92.80	−	−95.91
15	RRSRS	−	7.8×10^3	−	−99.83	−	−99.66
16	RSRRR[e]	−	23×10^3	−	−99.99	−	−100.00
17	SRRRR	−	33×10^3	−	−84.37	−	−88.14
18	RSSSR	−	36×10^3	−	−90.15	−	−91.13
19	RSRSR	−	47×10^3	−	−96.74	−	−98.40
20	RSRSS[e]	−	60×10^3	−	−95.53	−	−97.01
21	RRRRR	−	10^5	−	−99.96	−	−99.97
22	SRRRS	−	10^5	−	−79.69	−	−78.60
23	RRRSS	−	10^5	−	−83.51	−	−80.20
24	SRSRR[e]	−	10^5	−	−58.29	−	−46.48
25	RRRRS	−	10^5	−	−99.95	−	−99.95
26	RRSRR	−	10^5	−	−99.87	−	−99.82
27	SSSRR	−	10^5	−	−84.09	−	−88.26
28	RSSRS[e]	−	10^5	−	−99.95	−	−99.95
29	RRRSR	−	10^5	−	−88.22	−	−89.63
30	RSSRR	−	10^5	−	−99.96	−	−99.97
31	RSRRS	−	10^5	−	−99.99	−	−99.99
32	SRSRS[e]	−	10^5	−	−48.35	−	−18.65

[a] Notation of the chiral centers in each perindoprilate derivative in the following order C_2, C_{3a}, C_{7a}, C_9, C_{11}.

[b] Classification according to the value of the IC_{50}.

[c] Values of the IC_{50}, of the compound, for ACE in nM were take from previous works [4,24,28,83].

[d] ΔP posterior probability predicted for each compound using Eqs. (22) and (23).

[e] Compounds used in the test set.

prediction sets. Therefore, model (22) 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 4 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.

A very similar behavior was obtained with stochastic linear indices (Eq. (23)). In this case, the model correctly classifies 22

of 23 compounds for accuracy of 95.65% and a high MCC of 0.887 for the training set. In addition, model (23) shows a maximal accuracy of 100.00%, yielding a maximal MCC of 1.000 for the test set.

Table 5 depicts the results obtained in our study as well as those achieved with other cheminformatic approaches. It should be remarked that our model contain one variable less than that model obtained with MARCH-INSIDE molecular descriptors [24] and the same number of variables that were

Table 5
Classification of 32 perindopirilate's stereoisomers and the statistical parameters of the QSAR models obtained using different molecular descriptors

Index	n	λ	D^2	Accuracy (training) (%)	Accuracy (test) (%)	F
Non-stochastic bilinear indices (Eq. (22))	2	0.398	7.164	95.65	100.00	15.13
Stochastic bilinear indices (Eq. (23))	2	0.393	7.30	95.65	100.00	15.42
Non-stochastic linear indices [28]	2	0.398	7.82	100.00	88.88	15.08
Stochastic linear indices [28]	2	0.399	7.789	95.65	100.00	15.02
MARCH-INSIDE molecular descriptors [24]	3	0.38	8.43	91.30	88.88	10.30
Non-stochastic quadratic indices [4]	2	0.42	7.12	95.65	100.00	13.73

n: number of parameters in the obtained model.

used for develop their models by Marrero-Ponce and co-workers [4,43], to develop their models. However, the accuracy of these models is the same as that for the highest values obtained for this data set for the training set and test set.
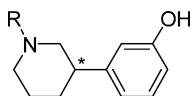
4.2. Modeling σ-receptor antagonist activities of 3-(3-hydroxyphenyl)piperidines

In a second application we investigate the ability of 3D-chiral bilinear indices to predict σ receptor antagonist activities. A short data set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl)piperidines that bind to σ-receptors, are also selected as illustrative example of the 3D-chiral bilinear indices application. This data set was introduced in QSAR studies by de Julian-Ortiz

et al. [16] in 1998, and after that it has been repeatedly used by some authors [4,24,28] in recent years, to validate new chiral topological indices. The σ-receptors mediate severe side effects induced by various dopamine antagonists [16].

3D-chiral bilinear indices are non-symmetric and reduce to classical descriptors when symmetry is not codified (see Table 3). Besides, Diaz et al. concluded that σ-receptor antagonistic activities are not a pseudoscalar property [24] and we can expect at least a good correlation with 3D-bilinear indices. This experiment also permitted us to compare our method with other previously reported approaches. The MLR analysis was used to develop QSAR models for the σ receptor antagonist activities. The models obtained using non-stochastic and stochastic bilinear indices for the σ-receptor antagonistic activities are given below:

Table 6
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]
(*R*)-3-HPP					
H	−0.66	−0.50	−0.16	−0.71	0.05
CH$_3$	0.43	0.06	0.37	0.35	0.08
C$_2$H$_5$	0.95	0.67	0.28	0.75	0.20
n-C$_3$H$_7$	1.52	1.34	0.18	1.55	−0.03
i-C$_3$H$_7$	0.61	1.10	−0.49	1.11	−0.50
n-C$_4$H$_9$	2.05	2.04	0.01	1.83	0.22
2-Phenylethyl	2.10	2.20	−0.10	2.21	−0.11
(*S*)-3-HPP					
H	−1.19	−1.02	−0.17	−1.25	0.06
CH$_3$	−0.28	−0.44	0.16	−0.19	−0.09
C$_2$H$_5$	−0.01	0.17	−0.18	0.19	−0.20
n-C$_3$H$_7$	0.81	0.85	−0.04	0.98	−0.17
I-C$_3$H$_7$	0.68	0.62	0.06	0.53	0.15
n-C$_4$H$_9$	1.51	1.54	−0.03	1.27	0.24
2-Phenylethyl	1.80	1.70	0.10	1.70	0.10

Abbreviations: HPP, *N*-alkylated 3-hydroxyphenyl piperidines.
[a] Alkylic (*R*) group at nitrogen ring.
[b] Observed values of the log IC$_{50}$ for the σ-receptor were take from the literature [4,16,24,28].
[c] Values calculated from Eq. (24).
[d] Residual, defined as [log IC$_{50}$(σ)Obs − log IC$_{50}$(σ)Cal].
[e] Values calculated from Eq. (25).

Table 7
Statistical parameters of the QSAR models obtained using different molecular descriptors 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
Non-stochastic bilinear indices (Eq. (24))	14	2	0.953	0.238	0.935	0.259	111.93
Stochastic bilinear indices (Eq. (25))	14	2	0.961	0.219	0.946	0.235	134.02
Non-stochastic linear indices [28]	14	2	0.939	0.271	0.909	0.305	84.87
Stochastic linear indices [28]	14	2	0.941	0.267	0.90	0.319	87.93
Chiral TIs [16]	14	3	0.931	0.301	a	a	45.70
MARCH-INSIDE molecular descriptors [24]	14	2	0.922	0.295	a	0.32	71.17
Non-stochastic quadratic indices [4]	14	2	0.940	0.270	0.912	0.289	85.82

[a] Values are not reported in the literature.

$$\log \mathrm{IC}_{50}(\sigma) = -5.934(\pm 0.451) + 1.449$$
$$\times 10^{-2}(\pm 0.129 \times 10^{-2})^{*\mathrm{M\text{-}E}}\boldsymbol{b}_1(\bar{x}, \bar{y}) - 3.810$$
$$\times 10^{-4}(\pm 0.429 \times 10^{-4})^{*\mathrm{M\text{-}E}}\boldsymbol{b}_4(\bar{x}, \bar{y}),$$
$$N = 14, \quad R^2 = 0.953, \quad q^2_{\mathrm{LOO}} = 0.935,$$
$$(2, 11) = 111.93, \quad s = 0.238, \quad s_{cv} = 0.259,$$
$$< 0.0001 \tag{24}$$

$$\log \mathrm{IC}_{50}(\sigma) = -16.975(\pm 2.338) + 5.369$$
$$\times 10^{-2}(\pm 0.405 \times 10^{-2})^{*\mathrm{P\text{-}Es}}\boldsymbol{b}^{\mathrm{H}}_{15\mathrm{L}}(\bar{x}, \bar{y}) + 3.462$$
$$\times 10^{-2}(\pm 0.679 \times 10^{-2})^{*\mathrm{M\text{-}Vs}}\boldsymbol{b}^{\mathrm{H}}_{11\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}}),$$
$$N = 14, \quad R^2 = 0.961, \quad q^2_{\mathrm{LOO}} = 0.946,$$
$$F(2, 11) = 134.02, \quad s = 0.219, \quad s_{cv} = 0.235,$$
$$p < 0.0001 \tag{25}$$

where N is the size of the data set, R^2 the square correlation coefficient (determination coefficient), s the standard deviation of the regression, F the Fischer ratio and q^2 (s_{cv}) is the square correlation coefficient (standard deviation) of the cross-validation performed by the LOO procedure. These statistics indicate that both models are appropriate for the description of chemicals studied here. In Table 6, they are shown the structure and values of experimental and calculated log IC$_{50}$(50% inhibitory concentration) for this data set.

The comparison with other methods reported previously for the same activity is shown in Table 7. As can be seen, our models have better statistical parameters than models obtained with MARCH-INSIDE molecular descriptors [24] and other chiral TIs [16], and slightly better than those obtain by us in earlier publications [4,28], when 3D-chiral quadratic and linear indices were used.

4.3. Prediction of the corticosteroid-binding globulin (CBG) binding affinity of a steroid family

Finally, in order to validate even more 3D-chiral bilinear indices in QSAR studies, we select a molecular set that is well known to QSAR researchers, the so-called Cramer's steroid data set. This data set was introduced by Cramer et al. [25] in 1988 using Comparative Molecular Field Analysis (CoMFA) method and, since then, it has become a benchmark for the assessment of novel QSAR methods [84,85]. 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 molecular descriptors [86]. Even though this data set is not the ideal 3D benchmark data set [86], it was used for the shake of comparability [87]. We use this molecular set, because all the compounds in this data set contain chiral atoms, and the binding affinities of these compounds are available [25]. Some structures of these compounds were drawn incorrectly in the original paper and were corrected in a recent report [85].

Different methods were used to develop 3D-QSAR models for this data set, including CoMFA [25], Comparative Molecular Similarity Indices Analysis (CoMSIA) [88], Molecular Quantum Similarity Measures (MQSM) [89], Topological Quantum Similarity Indices (TQSI) [90], Comparative Molecular Moment Analysis (CoMMA) [85], Mapping Property Distributions of Molecular Surfaces (MAP) [87], and so on [91–94]. Table 8 gathers the entire studied set with the actual binding affinities, taken from Robert et al. [89] The obtained models are given below together with their statistical parameters:

$$\mathrm{CBG} = -6.39(\pm 0.08) - 2.01(\pm 0.6)^{*\mathrm{P\text{-}E}}\boldsymbol{b}_1(x, y)$$
$$+ 2.07(\pm 0.53)^{*\mathrm{P\text{-}E}}\boldsymbol{b}_2(\bar{x}, \bar{y})$$
$$+ 8.30(\pm 1.32)^{*\mathrm{P\text{-}E}}\boldsymbol{b}_{2\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$- 10.16(\pm 1.31)^{*\mathrm{P\text{-}E}}\boldsymbol{b}_{4\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$+ 3.22(\pm 0.60)^{*\mathrm{P\text{-}E}}\boldsymbol{b}_{7\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$- 2.02(\pm 0.67)^{*\mathrm{P\text{-}E}}\boldsymbol{b}^{\mathrm{H}}_{2\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}}),$$
$$N = 31, \quad R = 0.937, \quad R^2 = 0.878,$$
$$F(6, 24) = 28.64, s = 0.42, \quad q^2 = 0.799,$$
$$s_{cv} = 0.48, \quad p < 0.0001 \tag{26}$$

$$\mathrm{CBG} = -6.39(\pm 0.07) + 4.56(\pm 0.64)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}_0(\bar{x}, \bar{y})$$
$$- 3.12(\pm 0.48)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}_1(\bar{x}, \bar{y}) - 1.49(\pm 0.33)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}^{\mathrm{H}}_0(\bar{x}, \bar{y})$$
$$+ 2.05(\pm 0.56)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}_{5\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$- 5.16(\pm 0.64)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}_{12\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$- 4.81(\pm 0.72)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}^{\mathrm{H}}_{1\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}})$$
$$+ 7.08(\pm 1.05)^{*\mathrm{P\text{-}Es}}\boldsymbol{b}^{\mathrm{H}}_{3\mathrm{L}}(\bar{x}_{\mathrm{E}}, \bar{y}_{\mathrm{E}}),$$
$$N = 31, \quad R = 0.955, \quad R^2 = 0.913, \quad F(7, 23) = 34.35,$$
$$s = 0.37, \quad q^2 = 0.833, \quad s_{cv} = 0.44, \quad p < 0.0001$$
$$\tag{27}$$

Table 8
Results of the steroids data set used for QSAR study

No.	Compounds	Observed CBG affinity (pK_a)[a]	Pred. value[b]	%E[c]	%E_{cv}[d]	Pred. value[e]	%E[c]	%E_{cv}[d]
1	Aldosterone	−6.279	−6.596	−5.050	−6.081	−6.289	−0.157	−0.195
2	Androstanediol	−5.000	−5.102	−2.040	−3.205	−5.056	−1.124	−1.941
3	Androstenediol	−5.000	−5.122	−2.439	−3.104	−4.999	0.027	0.033
4	Androstenedione	−5.763	−6.207	−7.710	−9.216	−6.444	−11.824	−16.737
5	Androsterone	−5.613	−5.409	3.628	5.165	−5.581	0.564	0.653
6	Corticosterone	−7.881	−7.322	7.097	8.294	−7.354	6.693	8.037
7	Cortisol	−7.881	−7.618	3.337	3.905	−7.603	3.531	4.155
8	Cortisone	−6.892	−7.850	−13.904	−16.679	−7.385	−7.156	−8.674
9	Dehydroepiandrosterone	−5.000	−4.864	2.725	3.143	−4.859	2.816	3.406
10	Deoxycorticosterone	−7.653	−7.300	4.611	5.200	−7.351	3.950	4.707
11	Deoxycortisol	−7.881	−7.343	6.830	7.867	−7.447	5.504	6.719
12	Dihydrtestosterone	−5.919	−5.647	4.594	6.614	−5.704	3.636	4.890
13	Estradiol	−5.000	−4.984	0.328	0.433	−5.181	−3.618	−4.460
14	Estriol	−5.000	−4.913	1.735	2.775	−4.888	2.230	5.967
15	Estrone	−5.000	−4.729	5.419	8.026	−4.533	9.345	16.410
16	Ethiocholanolone	−5.255	−5.409	−2.937	−4.181	−5.581	−6.210	−7.196
17	Pregnenolone	−5.255	−5.609	−6.742	−7.807	−5.471	−4.109	−6.082
18	17-Hydroxyregnenolone	−5.000	−5.688	−13.751	−20.909	−5.507	−10.138	−14.497
19	Progesterone	−7.380	−6.948	5.852	6.712	−7.060	4.342	5.247
20	17-Hydroxyprogesterone	−7.740	−7.029	9.189	11.758	−7.221	6.703	8.605
21	Testosterone	−6.724	−6.463	3.880	4.371	−6.866	−2.116	−2.559
22	Prednisolone	−7.512	−7.461	0.685	0.901	−7.785	−3.628	−4.588
23	Cortisolacetate	−7.553	−7.263	3.842	7.234	−7.754	−2.664	−4.244
24	4-Pregnene-3,11,20-trione	−6.779	−7.017	−3.516	−4.652	−6.953	−2.570	−3.296
25	Epicorticosterone	−7.200	−7.402	−2.800	−3.776	−7.279	−1.093	−1.317
26	19-Nortestosterone	−6.144	−6.651	−8.247	−10.181	−6.524	−6.189	−9.163
27	16a,17a-Dihydroxyprogesterone	−6.247	−6.437	−3.039	−3.301	−6.246	0.013	0.016
28	16a-Methylprogesterone	−7.120	−7.237	−1.644	−2.144	−7.187	−0.942	−1.354
29	19-Norprogesterone	−6.817	−6.773	0.643	0.758	−6.386	6.316	8.773
30	2a-Methylcortisol	−7.688	−8.016	−4.272	−5.250	−7.899	−2.744	−3.884
31	2a-Methyl-9a-fluorocortisol	−5.797	−5.564	4.018	6.957	−5.579	3.756	8.527

[a] Observed CBG affinity values taken from Robert et al. [89].
[b] Predicted CBG affinity values using Eq. (26).
[c] Percent of relative error; %$E = 100 \times$ [Obs − Pred/Obs].
[d] Percent of relative error in leave-one-out cross-validation procedure; %$E_{cv} = 100 \times$ [Obs − Pred$_{loo-CV}$/Obs].
[e] Predicted CBG affinity values using Eq. (27).

Table 9
Comparison between of prediction for the steroid data set 3D-chiral bilinear indices and other 3D QSAR approaches

QSAR method	N	n	Statistical method	q^2	Reference
Similarity matrices-based molecular descriptors	31	6	genetic NN	0.940	[93]
TQSAR	31	6	MLR after PCA	0.842	[89]
3D-chiral bilinear indices (*stochastic*)	31	7	MLR	0.833	Eq. (27)
3D-chiral bilinear indices (*non-stochastic*)	31	6	MLR	0.799	Eq. (26)
3D-chiral linear indices (*stochastic*)	31	7	MLR	0.788	[28]
3D-chiral quadratic indices (*non-stochastic*)	31	6	MLR	0.781	[43]
MEDV	31	5	GA and RLM	0.777	[95]
TQSI	31	3	MLR	0.775	[90]
3D-chiral linear indices (*non-stochastic*)	31	6	MLR	0.767	[28]
MEDV	31	6	GA and RLM	0.765	[95]
3D-chiral quadratic indices (*stochastic*)	31	7	MLR	0.735	[43]
Similarity indices	31	1	PLS	0.734	[92]
E-state and kappa shape index	31	4	MLR[a]	0.730	[96]
MQSM	31	4	MLR after PLS	0.727	[97]
E-state and kappa shape index	31	4	MLR	0.720	[96]
MQMS	31	3	MLR and PCA	0.705	[90]
CoMMA	31	6	PCR	0.689	[98]
MEDV	31	4	GA and RLM	0.648	[95]
Wagener's	31	–	k-NN and FNN	0.630	[91]

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

As it can be seen, the non-stochastic model (Eq. (26)) explains almost the 88% of the variance of the experimental CBG values using six variables to describe the 31 steroids, while the stochastic model (Eq. (27)) explains more than 91% of the variance using seven variables. Both models showed a small value of standard deviation $s = 0.42$ and $s = 0.37$, respectively.

An important aspect of QSAR modeling is the development of a way to validate the model. 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 0.799 and 0.833, respectively. Both values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models [80–82]. 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 9, where the results were arranged in decreasing order of q^2 and the comparison can be easily carried out.

5. Final conclusions

In these studies we demonstrated that atom-based 3D-chiral bilinear indices can be successfully applied in QSAR studies which 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 shown in the present work, the generalized atom-based 3D-chiral bilinear indices 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 σ-receptors, as well as to predict the corticosteroid-binding globulin binding affinity of the Cramer's steroid data set. This result is only a preliminary conclusion and an extensive analysis of the potential of the 3D-chiral bilinear indices is necessary. However, we show that for the three data sets chiral-QSAR models obtained with 3D-chiral bilinear indices had better or similar predictive ability as compared to other previously reported chiral and/or 3D-QSAR.

Acknowledgements

References

[1] A. Golbraikh, A. Tropsha, QSAR modeling using chirality descriptors derived from molecular topology, J. Chem. Inf. Comput. Sci. 43 (2003) 144–154.

[2] J.V. de Julian-Ortiz, R. Garcia-Domenech, J. Galvez, R. Soler, F.J. García-March, G.M. Antón-Fos, Use of topological descriptors in chromatographic chiral separations, J. Chromatogr. A 719 (1996) 37–44.

[3] V.M. Potapov, Stereochemistry, Khimia, Moscow, 1988.

[4] Y. Marrero-Ponce, H.G. Díaz, V. Romero, F. Torrens, E.A. Castro, 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 r-receptor antagonist activities, Bioorg. Med. Chem. 12 (2004) 5331–5342.

[5] J. Aires-de-Sousa, J. Gasteiger, Prediction of enantiomeric selectivity in chromatography. Application of conformation-dependent and conformation-independent descriptors of molecular chirality, J. Mol. Graph. Model. 20 (2002) 373–388.

[6] J. Solms, L. Vuataz, R.H. Egli, The taste of L- and D-amino acids, Experientia 21 (1965) 692–694.

[7] S.S. Schiffman, T.B. Clark III, J. Gagnon, Influence of chirality of amino acids on the growth of perceived taste intensity with concentration, Physiol. Behav. 28 (1982) 457–465.

[8] M. Laska, P. Teubner, Olfactory discrimination ability of human subjects for ten pairs of enantiomers, Chem. Sens. 24 (1999) 161–170.

[9] E.H. Polak, A.M. Fombon, C. Tilquin, P.H. Punter, Sensory evidence for olfactory receptors with opposite chiral selectivity, Behav. Brain Res. 31 (1989) 199–206.

[10] W.H. DeCamp, The FDA, Chirality 1 (1989) 2.

[11] A.J. Hutt, S.C. Tan, Drug chirality and its clinical significance, Drugs 52 (Suppl. 5) (1996) 1–12.

[12] S. Wnendt, K. Zwingenberger, Thalidomide's chirality, Nature (London) 385 (1997) 303–304.

[13] J. Aires-de-Sousa, J. Gasteiger, I. Gutman, D. Vidovic, Chirality codes and molecular structure, J. Chem. Inf. Comput. Sci. 44 (2004) 831–836.

[14] J. Aires-de-Sousa, J. Gasteiger, New description of molecular chirality and its application to the prediction of the preferred enantiomer in stereoselective reactions, J. Chem. Inf. Comp. Sci. 41 (2001) 369–375.

[15] E. Ruch, Algebraic aspects of the chirality phenomenon in chemistry, Acc. Chem. Res. 5 (1972) 49–56.

[16] J.V. de Julian-Ortiz, C. de Gregorio Alapont, I. Rios-Santamarina, R. Garcia-Domenech, J. Galvez, Prediction of properties of chiral compounds by molecular topology, J. Mol. Graph. Model. 16 (1998) 14–18.

[17] R. Benigni, M. Cotta-Ramusino, G. Gallo, F. Giorgi, A. Giuliani, M.R. Vari, Deriving a quantitative chirality measure from molecular similarity indices, J. Med. Chem. 43 (2000) 3699–3703.

[18] A. Golbraikh, D. Bonchev, A. Tropsha, Novel chirality descriptors derived from molecular topology, J. Chem. Inf. Comput. Sci. 41 (2001) 147–158.

[19] S.A. Wildman, G.M. Crippen, Validation of DAPPER for 3D QSAR: conformational search and chirality metric, J. Chem. Inf. Comput. Sci. 43 (2003) 629–636.

[20] H.P. Schultz, E.B. Schultz, T.P. Schultz, Topological organic chemistry. 9. Graph theory and molecular topological indices of stereoisomeric organic compounds, J. Chem. Inf. Comput. Sci. 35 (1995) 864–870.

[21] J. Aires-de-Sousa, J. Gasteiger, Prediction of enantiomeric excess in a combinatorial library of catalytic enantioselective reactions, J. Comb. Chem. 7 (2005) 298–301.

[22] E. Estrada, E. Uriarte, Recent advances on the role of topological indices in drug discovery research, Curr. Med. Chem. 8 (2001) 1573–1588.

[23] A.B. Buda, K. Mislow, On geometric measures of chirality, J. Mol. Struct. (Theochem.) 232 (1991) 1–12.

[24] H.G. Diaz, I.H. Sanchez, E. Uriarte, L. Santana, Symmetry considerations in Markovian chemicals 'in silico' design (MARCH-INSIDE). I. Central chirality codification, classification of ACE inhibitors and prediction of sigma-receptor antagonist activities, Comput. Biol. Chem. 27 (2003) 217–227.

[25] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative Molecular Field Analysis (CoMFA). 1. Effects of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[26] Y. Marrero-Ponce, Linear indices of the "molecular pseudograph's atom adjacency matrix": definition significance-interpretation, and application to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors, J. Chem. Inf. Comput. Sci. 44 (2004) 2010–2026.

[27] Y. Marrero-Ponce, Total and local quadratic indices of the molecular pseudograph's atom adjacency matrix: applications to the prediction of physical properties of organic compounds, Molecules 8 (2003) 687–726.

[28] Y. Marrero-Ponce, J.A. Castillo-Garit, 3D-chiral atom, atom-type, and total non-stochastic and stochastic molecular linear indices and their applications to central chirality codification, J. Comput.-Aided Mol. Des. 19 (2005) 369–383.

[29] Y. Marrero-Ponce, A. Huesca-Guillen, F. Ibarra-Velarde, Quadratic indices of the "molecular pseudograph's atom adjacency matrix" and their stochastic forms: a novel approach for virtual screening and in silico discovery of new lead paramphistomicide drugs-like compounds, J. Mol. Struct. (Theochem.) 717 (2005) 67–79.

[30] C.H. Edwards, D.E. Penney, Elementary Linear Algebra, Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1988.

[31] Y. Marrero-Ponce, A. Montero-Torres, C.R. Zaldivar, M.I. Veitia, M.M. Perez, R.N. Sanchez, Non-stochastic and stochastic linear indices of the 'molecular pseudograph's atom adjacency matrix': application to 'in silico' studies for the rational discovery of new antimalarial compounds, Bioorg. Med. Chem. 13 (2005) 1293–1304.

[32] Y. Marrero-Ponce, M. Iyarreta-Veitia, A. Montero-Torres, C. Romero-Zaldivar, C.A. Brandt, P.E. Avila, K. Kirchgatter, Y. Machado, Ligand-based virtual screening and in silico design of new antimalarial compounds using nonstochastic and stochastic total and atom-type quadratic maps, J. Chem. Inf. Model. 45 (2005) 1082–1100.

[33] Y. Marrero-Ponce, R.M. Marrero, F. Torrens, Y. Martinez, M.G. Bernal, V.R. Zaldivar, E.A. Castro, R.G. Abalo, Non-stochastic and stochastic linear indices of the molecular pseudograph's atom-adjacency matrix: a novel approach for computational in silico screening and "rational" selection of new lead antibacterial agents, J. Mol. Model. 12 (2006) 255–271.

[34] Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A.M. Sanchez, F. Torrens, E.A. Castro, Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic, Bioorg. Med. Chem. 13 (2005) 1005–1020.

[35] Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castañedo, F. Ibarra-Velarde, A. Huesca-Guillen, E. Jorge, A. del Valle, F. Torrens, E.A. Castro, TOMOCOMD-CARDD, a novel approach for computer-aided 'rational' drug design. I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds, J. Comput.-Aided Mol. Des. 18 (2004) 615–634.

[36] Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E.A. Castro, Atom, atom-type, and total non-stochastic and stochastic quadratic fingerprints: a promising approach for modeling of antibacterial activity, Bioorg. Med. Chem. 13 (2005) 2881–2899.

[37] A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado-Tugores, A. Montero-Torres, D.M. Pereira, J.A. Escario, J.J. Nogal-Ruiz, C. Ochoa, V.J. Aran, A.R. Martinez-Fernandez, R.N. Garcia Sanchez, A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results, Bioorg. Med. Chem. Lett. 15 (2005) 3838–3843.

[38] Y. Marrero-Ponce, Total and local (atom and atom type) molecular quadratic indices: significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications, Bioorg. Med. Chem. 12 (2004) 6351–6369.

[39] Y. Marrero-Ponce, J.A. Castillo-Garit, F. Torrens, V. Romero-Zaldivar, E. Castro, Atom, atom-type, and total linear indices of the "molecular pseudograph's atom adjacency matrix": application to QSPR/QSAR studies of organic compounds, Molecules 9 (2004) 1100–1123.

[40] Y. Marrero-Ponce, M.A. Cabrera, V Romero, E. Ofori, L.A. Montero, Total and local quadratic indices of the "molecular pseudograph's atom adjacency matrix". Application to prediction of Caco-2 permeability of drugs, Int. J. Mol. Sci. 4 (2003) 512–536.

[41] Y. Marrero Ponce, M.A. Cabrera Perez, V. Romero Zaldivar, H. Gonzalez Diaz, F. Torrens, A new topological descriptors based model for predicting intestinal epithelial transport of drugs in Caco-2 cell culture, J. Pharm. Pharmaceut. Sci. 7 (2004) 186–199.

[42] Y. Marrero-Ponce, M.A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio, F. Torrens, Prediction of intestinal epithelial transport of drug in (Caco-2) cell culture from molecular structure using in silico approaches during early drug discovery, Internet Electron. J. Mol. Des. 4 (2005) 124–150.

[43] J.A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, Atom-based 3D-chiral quadratic indices. Part 2. Prediction of the corticosteroid-binding globulinbinding affinity of the 31 benchmark steroids data set, Bioorg. Med. Chem. 14 (2006) 2398–2408.

[44] Y. Marrero-Ponce, D. Nodarse, H.D. González, R. Ramos de Armas, V. Romero-Zaldivar, F. Torrens, E. Castro, Nucleic acid quadratic indices of the "macromolecular graph's nucleotides adjacency matrix". Modeling of footprints after the interaction of paromomycin with the HIV-1 Ψ-RNA packaging region, Int. J. Mol. Sci. 5 (2004) 276–293.

[45] Y. Marrero Ponce, J.A. Castillo Garit, D. Nodarse, Linear indices of the 'macromolecular graph's nucleotides adjacency matrix' as a promising approach for bioinformatics studies. Part 1. Prediction of paromomycin's affinity constant with HIV-1 psi-RNA packaging region, Bioorg. Med. Chem. 13 (2005) 3397–3404.

[46] Y. Marrero-Ponce, R. Medina, E.A. Castro, R. de Armas, H. González, V. Romero, F. Torrens, Protein quadratic indices of the "macromolecular pseudograph's α-carbon atom adjacency matrix". 1. Prediction of arc repressor alanine-mutant's stability, Molecules 9 (2004) 1124–1147.

[47] Y. Marrero-Ponce, R. Medina-Marrero, J.A. Castillo-Garit, V. Romero-Zaldivar, F. Torrens, E.A. Castro, Protein linear indices of the 'macromolecular pseudograph alpha-carbon atom adjacency matrix' in bioinformatics. Part 1. Prediction of protein stability effects of a complete set of alanine substitutions in Arc repressor, Bioorg. Med. Chem. 13 (2005) 3003–3015.

[48] Y. Marrero-Ponce, A. Meneses-Marcel, J.A. Castillo Garit, Y. Machado-Tugores, J.A. Escario, A.G. Barrio, D.M. Pereira, J.J. Nogal-Ruiz, V.J. Arán, A.R. Martínez-Fernández, F. Torrens, R. Rotondo, F. Ibarra-Velarde, Y.J. Alvarado, Predicting antitrichomonal acitivity: A computational screening using atom-based bilinear indices and experimental proofs, Bioorg. Med. Chem. 14 (2006) 6502–6524.

[49] Y. Marrero-Ponce, M.T.H. Khan, G.M. Casañola-Martín, A. Ather, M.N. Sultankhodzhaev, F. Torrens, Prediction of tyrosinase inhibition spectra for chemicals using novel atom-based bilinear indices, CheMedChem, submitted for publication.

[50] R. Wang, Y. Gao, L.P. Lai, Calculating partition coefficient by atom-additive method, Drug Discov. Des. 19 (2000) 47–66.

[51] P. Ertl, B. Rohde, P. Selzer, Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties, J. Med. Chem. 43 (2000) 3714–3717.

[52] A.K. Ghose, G.M. Crippen, Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions, J. Chem. Inf. Comput. Sci. 27 (1987) 21–35.

[53] K.J. Millar, Additivity methods in molecular polarizability, J. Am. Chem. Soc. 112 (1990) 8533–8542.

[54] J. Gasteiger, M. Marsilli, A new model for calculating atomic charge in molecules, Tetrahedron Lett. 34 (1978) 3181–3184.

[55] L.B. Kier, L.H. Hall, Molecular connectivity in structure–activity analysis, Research Studies Press, Letchworth, UK, 1986.

[56] L. Pauling, The Nature of Chemical Bond, Cornell University Press, Ithaca, NY, 1939.

[57] D.H. Rouvray, in: A.T. Balaban (Ed.), Chemical Applications of Graph Theory, Academic Press, London, 1976, pp. 180–181.

[58] N. Trinajstić, Chemical Graph Theory, 2nd ed., CRC Press, Boca Raton, FL, 1992 http://www.ecampus.com/book/0849342562.

[59] I. Gutman, O.E. Polansky, Mathematical Concepts in Organic Chemistry, Springer-Verlag, Berlin, 1986.

[60] E. Estrada, G. Patlewicz, On the usefulness of graph-theoretic descriptors in predicting theoretical parameters. Phototoxicity of polycyclic aromatic hydrocarbons (PAHs), Croat. Chim. Acta 77 (2004) 203–211.

[61] D.J. Klein, Graph theoretically formulated electronic–structure theory, Internet Electron. J. Mol. Des. 2 (2003) 814–834.

[62] N. Jacobson, Basic Algebra I, 2nd ed., W.H. Freeman and Company, New York, 1985.

[63] K.F. Riley, M.P. Hobson, S.J. Vence, Mathematical Methods for Physics and Engineering, Cambridge University Press, Cambridge, 1998.

[64] E. Hernández, Álgebra y Geometría, Universidad Autonoma de Madrid, Madrid, 1987.

[65] J. de Burgos-Román, Álgebra y Geometría Cartesíana, 2da Edición ed., McGraw-Hill Interamericana de España, Madrid, 2000.

[66] J. de Burgos-Román, Curso de Álgebra y Geometría, Alhambra Longman, Madrid, 1994.

[67] G. Werner, Linear Algebra, 4th ed., Springer-Verlag, New York, 1981.

[68] M. Randić, Generalized Molecular Descriptors, J. Math. Chem. 7 (1991) 155–168.

[69] P.D. Walker, P.G. Mezey, Molecular electron density Lego approach to molecule building, J. Am. Chem. Soc. 115 (1993) 12423–12430.

[70] L.B. Kier, L.H. Hall, Molecular Structure Description. The Electrotopological State, Academic Press, New York, 1999.

[71] L.H. Hall, C.T. Story, Boiling point and critical temperature of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks, J. Chem. Inf. Comput. Sci. 36 (1996) 1004–1014.

[72] J.D. Gough, L.H. Hall, Modeling antileukemic activity of carboquinones with electrotopological state and chi indices, J. Chem. Inf. Comput. Sci. 39 (1999) 356–361.

[73] E.L. Elliel, S. Wilen, L. Mander, Stereochemistry of Organic Compounds, John Wiley & Sons Inc., New York, 1994.

[74] M.J.S. Dewar, Applications of quantum mechanical molecular models to chemical problems. Part 70. Quantum mechanical molecular models, J. Phys. Chem. 89 (1985) 2145–2150.

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

[76] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors, J. Chem. Inf. Comput. Sci. 43 (2002) 682–692.

[77] R. Todeschini, P. Gramatica, New 3D molecular descriptors: The WHIM theory and QSAR applications, Persp. Drug Disc. Des. 9–11 (1998) 355–380.

[78] STATISTICA version. 6.0 Statsoft, I.

[79] P. Baldi, S. Brunak, Y. Chauvin, C.A. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview, Bioinformatics 16 (2000) 412–424.

[80] S. Wold, L. Erikson, Chemometric methods in molecular design, in: H. van de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH Publishers, Weinheim, 1995, pp. 309–318.

[81] D.A. Belsey, E. Kuh, R.E. Welsch, Regression Diagnostics, Wiley, New York, 1980.

[82] A. Golbraikh, A. Tropsha, Beware of q2! J. Mol. Graph. Model. 20 (2002) 269–276.

[83] M. Vincent, B. Marchand, G. Remond, S. Jaguelin-Guinamant, G. Damien, B. Portevin, J.Y. Baumal, J.P. Volland, J.P. Bouchet, P.H. Lambert, et al., Synthesis and ACE inhibitory activity of the stereoisomers of perindopril (S 9490) and perindoprilate (S 9780), Drug Des. Discov. 9 (1992) 11–28.

[84] E.A. Coats, 3D QSAR in Drug Design, Kluwer/ESCOM, Dordrecht, 1998

[85] B.D. Silverman, The thirty-one benchmark steroids revisisited: comparative molecular moment analysis (CoMMA) with principal component regression, Quant. Struct.-Act. Relat. 19 (2000) 237–246.

[86] E.A. Coats, The CoMFA steroids as a benchmark dataset for development of 3D QSAR methods, Persp. Drug Disc. Des. 12–14 (1998) 199–213.

[87] N. Stiefl, K. Baumann, Mapping property distributions of molecular surfaces: algorithm and evaluation of a novel 3D quantitative structure–activity relationship technique, J. Med. Chem. 46 (2003) 1390–1407.

[88] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.

[89] D. Robert, L. Amat, R. Carbo-Dorca, Three-dimensional quantitative structure–activity relationships from tuned molecular quantum similarity measures: prediction of the corticosteroid-binding globulin binding affinity for a steroid family, J. Chem. Inf. Comput. Sci. 39 (1999) 333–344.

[90] M. Lobato, L. Amat, E. Besalu, R. Carbo-Dorca, Structure–activity relationships of a steroid family using quantum similarity measures and topological quantum similarity indices, Quant. Struct.-Act. Relat. 16 (1997) 465–472.

[91] M. Wagener, J. Sadowski, J. Gasteiger, Autocorrelation of molecular-surface properties for modeling corticosteroid-binding globulin and cytosolic AH receptor activity by neural networks, J. Am. Chem. Soc. 117 (1995) 7769–7775.

[92] M.F. Parretti, R.T. Kroemer, J.H. Rothman, W.G. Richards, Alignment of molecules by the Monte Carlo optimization of molecular similarity indices, J. Comput. Chem. 18 (1997) 1334–1353.

[93] S.S. So, M. Karplus, Three-dimensional quantitative structure–activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations, J. Med. Chem. 40 (1997) 4347–4359.

[94] H. Chen, J. Zhou, G. Xie, PARM: a genetic algorithm to predict bioactivity, J. Chem. Inf. Comp. Sci. 38 (1998) 243–250.

[95] S.S. Liu, C.S. Yin, L.S. Wang, Combined MEDV-GA-MLR method for QSAR of three panels of steroids, dipeptides, and COX-2 inhibitors, J. Chem. Inf. Comput. Sci. 42 (2002) 749–756.

[96] H.H. Maw, L.H. Hall, E-state modeling of corticosteroids binding affinity validation of model for small data set, J. Chem. Inf. Comput. Sci. 41 (2001) 1248–1254.

[97] E. Besalu, X. Girones, L. Amat, R. Carbo-Dorca, Molecular quantum similarity and the fundamentals of QSAR, Acc. Chem. Res. 35 (2002) 289–295.

[98] B.D. Silverman, D.E. Platt, Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition, J. Med. Chem. 39 (1996) 2129–2140.