9th International Electronic Conference on Synthetic Organic Chemistry. ECSOC-9
1-30 November 2005. http://www.usc.es/congresos/ecsoc/9/ECSOC9.HTM & http://www.mdpi.net/ecsoc/

[G004]

Non-Stochastic and Stochastic Atom-based 3D-Chiral Linear Indices and their Applications to Central Chirality Codification.

Juan A. Castillo Garit[a,b] & Yovani Marrero Ponce[b,c]

[a]Applied Chemistry Research Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba. **Fax:** 53-42-281130, 281455; **Telephone:** 53-42-281192, 281473; email: juancg@uclv.edu.cu; juancg.22@gmail.com or jacgarit@yahoo.es
[b]Department of Pharmacy, Faculty of Chemical-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, Dr. Moliner 50, E-46100 Burjassot (València), Spain.

Abstract

Non-stochastic and stochastic 2D linear indices have been generalized to codify chemical structure information for chiral drugs, making use of a trigonometric 3D-chirality correction factor. These descriptors circumvent the inability of conventional 2D non-stochastic linear indices to distinguish σ-stereoisomers. 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 σ-stereoisomers combinatorial library. Two linear discriminant analysis models, using non-stochastic and stochastic linear indices, were obtained. The models shown an accuracy of 100% and 96.65% for the training set; and 88.88% and 100% in the external test set, respectively. Canonical regression analysis corroborated the statistical quality of these models (R_{can} of 0.78 and of 0.77) and was also used to compute biology activity canonical scores for each compound. After that, the prediction of the σ-receptor antagonists of chiral 3-(3-hydroxyphenyl)piperidines by linear multiple regression analysis was carried out. Two statistically significant QSAR models were obtained when non-stochastic (R^2 = 0.982 and s = 0.157) and stochastic (R^2 = 0.941 and s = 0.267) 3D-chiral linear indices were used. The predictive power was assessed by the leave-one-out cross-validation experiment, yielding values of q^2 = 0.982 (s_{cv} = 0.186) and q^2 = 0.90 (s_{cv} = 0.319), respectively. Finally, the prediction of the corticosteroid-binding globulin binding affinity of steroids set was performed. The best results obtained in the cross-validation procedure with non-stochastic (q^2 = 0.904) and stochastic (q^2 = 0.88) 3D-chiral linear indices 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 provide a powerful alternative to 3D-QSAR.

Keywords: non-Stochastic and Stochastic 3D-Chiral Linear Indices, 3D-QSAR, Angiotesin-converting enzyme inhibitors, σ-Receptor antagonists, Binding Affinity of Steroids.

* To receive all correspondence

Introduction

Asymmetry of atomic configurations is very important feature in determining the physical, chemical and biological properties of chemicals substances [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 as referred to as σ-diastereomers [3]. Most of the physical as well as chemical properties of chiral molecules are similar. At the same time, it is well know that many biological molecules are chiral and that the chirality plays an essential role in defining biological activity [1]. The case of thalidomide is an example of a problem that was, at least, complicated by the ignorance of stereochemical effects [4]. 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 [5].

Several quantitative measures of chirality have been developed in the past and were extensively reviewed [6-8]. Buda and Mislow distinguished between two classes of measures [6]. 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. A different idea was to incorporate R/S labels into conventional topological indices (TIs) [9]. Derived chirality descriptors were correlated with biological activity by Julián-Ortiz *et al.* [10], Golbraikh *et al.* [1] and more recently by González-Díaz *et al.* [11]. These indices are refereed 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 [12].

Recently, a novel scheme to the rational –*in silico*- molecular design and to QSAR/QSPR has been introduced by one of the present authors ***TOMOCOMD*** **(**acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign). It calculates several new families of molecular descriptors. In this sense, quadratic and linear indices have been defined in analogy to the quadratic and linear

mathematical maps [13,14]. This approach has been successfully employed in QSPR [13,15-17] and QSAR [14,18-22] studies, including studies related to nucleic acid-drug interactions [23,24], and central chirality codification [25]. Finally, an alternative formulation of our approach for structural characterization of proteins was carried out recently [26,27].

The main aim of the present paper is to extend 2D linear indices of the "molecular pseudograph's atom adjacency matrix" in order to codify chirality related structural features. The problem of classification of ACE (Angiotesin-Converting Enzime) 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 matter of comparison with other CTIs, 3D and quantum chemical descriptors as well.

Theoretical framework

2D non-Stochastic and Stochastic linear indices

The atom, atom-type and total 2D non-stochastic and stochastic linear indices of the "molecular pseudograph`s atom adjacency matrix" for small-to-medium sized organic compounds have been explained in some detail elsewhere [13,14,20]. However, an overview of this approach will be given.

For a given molecule composed of n atoms, the "molecular vector" (X) is constructed and the k^{th} atom linear indices, $\boldsymbol{f_k}(x_i)$, are calculated as a linear maps on $\Re^n$ [$\boldsymbol{f_k}(x_i)$: $\Re^n \rightarrow \Re^n$; thus $\boldsymbol{f_k}(x_i)$: Endomorphism on $\Re^n$] in canonical basis as shown in Eq. **1**,

$$f_k(x_i) = \sum_{j=1}^{n} {}^k a_{ij} X_j \qquad (\mathbf{1})$$

where, ${}^k a_{ij} = {}^k a_{ji}$ (symmetric square matrix), n is the number of atoms of the molecule, and $X_1,\ldots,X_n$ are the coordinates or components of the "molecular vector" (X) in a system of canonical basis vectors of $\Re^n$. The components of the "molecular" vector are numeric values, which can be considered as weights (atom-labels) for the vertices of the pseudograph. Certain atomic properties (electronegativity, density, atomic radius, etc) can be used with this propose. In this work Pauling electronegativity was selected as atom weights [28].

The coefficients ${}^k a_{ij}$ are the elements of the k^{th} power of the symmetric square matrix **M**(G) of the molecular pseudograph (G) and are defined as follows: [14,16,20,22]

$a_{ij} = P_{ij}$ if $i \neq j$ and $\exists\, e_k \in$ E(G) **(2)**

$= L_{ii}$ if $i = j$

$= 0$ otherwise

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

Note that linear indices's matrices, $\mathbf{M}^k$, are graph–theoretic electronic–structure models; like an "extended Hückel MO model". The $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ- and π-networks) in one step and their power (k = 0, 1, 2, 3…) 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 [10].

The present approach is based on a simple model for the intramolecular movement of all outer-shell 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,…) throughout the chemical-bonding network.

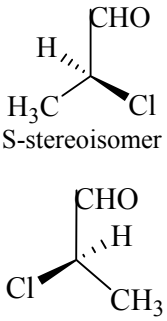
The k^{th} stochastic molecular pseudograph´s atom adjacency matrix [$\mathbf{S}^k$(G)] can be obtained from $\mathbf{M}^k$. Here, $\mathbf{S}^k$(G) = $\mathbf{S}^k$ = [$^k\mathbf{s}_{ij}$], is a squared table of order n (n = number of atoms) and the elements $^k\mathbf{s}_{ij}$ are defined as follows:

$$^k s_{ij} = \frac{^k a_{ij}}{^k SUM_i} = \frac{^k a_{ij}}{^k \delta_i} \qquad \textbf{(3)}$$

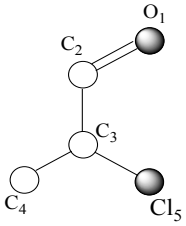
where, $^k a_{ij}$ are the elements of the k^{th} power of $\mathbf{M}$ and the SUM of the i^{th} row of $\mathbf{M}^k$ are named the k-order vertex degree of atom i, $^k\delta_i$. The k^{th} $\mathbf{s}_{ij}$ elements are the transition probabilities with the electrons move from atom i to j in the discrete time periods t_k. Note, that k^{th} element $\mathbf{s}_{ij}$ take into consideration the molecular topology in k step throughout of the chemical-bonding (σ- and π-) network.

Table 1 depict the calculation of the linear indices of the molecular pseudograph's atom adjacency matrix for 2-chloro-propionaldehyde.

Table 1. Definition and calculation of non-stochastic and stochastic total (whole-molecule) and local (atom) 3D-chiral and simple 2D-linear indices of the molecular pseudograph's atom adjacency matrix of the molecule 2-chloro-propionaldehyde.



Molecular Structure



Molecular H-Suppressed Pseudograph

$^*\mathbf{X} = [O_1\ C_2,\ C_3,\ C_4,\ Cl_5,]$

Chiral Molecular Vector: $^*\mathbf{X} \in \Re^5$

In the definition of the $^*\mathbf{X}$, 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 $\chi(O)$ (oxygen Pauling electronegativity) + $\sin((\omega_A+4\Delta)\pi/2)$. Therefore, if we use the canonical basis of $\Re^5$, the coordinates of any vector $^*\mathbf{X}$ coincide with the components of that chiral molecular vector.

$\sin((\omega_A+4\Delta)\pi/2)$ 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) linear indices ones for molecules without specific 3D characteristics.

[*X]: Column vector of coordinates of $^*\mathbf{X}$ in the canonical basis of R^5 (a *nx*1 matrix)
[*X] = [3.44, 2.55, ***2.55***, 2.55, 3.16] for chirality insensitive linear indices
[*X] = [3.44, 2.55, ***3.55***, 2.55, 3.16] for R-stereoisomer
[*X] = [3.44, 2.55, ***1.55***, 2.55, 3.16] for S-stereoisomer

non-Stochastic Linear Indices

$$f_1(x_i)=\sum_{j=1}^{n}{}^1a_{ij}X_j = \mathbf{M^1}[^*\mathrm{X}] = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} O \\ C \\ C \\ C \\ Cl \end{bmatrix} = \begin{bmatrix} 2C_2 \\ 2O_1 + 1C_3 \\ 1C_2 + 1C_4 + 1C1_5 \\ 1C_3 \\ 1C_3 \end{bmatrix}$$

Stochastic Linear Indices

$$f_1(x_i)=\sum_{j=1}^{n}{}^1s_{ij}X_j = \mathbf{S^1}[^*\mathrm{X}] = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.67 & 0 & 0.33 & 0 & 0 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} O \\ C \\ C \\ C \\ Cl \end{bmatrix} = \begin{bmatrix} 1C_2 \\ 0.67O_1 + 0.33C_3 \\ 0.33C_2 + 0.33C_4 + 0.33C1_5 \\ 1C_3 \\ 1C_3 \end{bmatrix}$$

3D-chiral linear indices of first order is a linear form; $^*f_1(x)$: $\Re^n \rightarrow \Re^n$ *such that,*
using non-stochastic linear indices: $^*f_1(O_1, C_2, C_3, C_4, Cl_5) = (2C_2, 2O_1+1C_3, 1C_2+1C_4+1Cl_5, 1C_3, 1C_3)$
using stochastic linear indices: $^*f_1(O_1, C_2, C_3, C_4, Cl_5) = (C_2, 0.67O_1+0.33C_3, 0.33C_2+0.33C_4+0.33Cl_5, 1C_3, 1C_3)$

	3D (R)-stereoisomer			**'Classical' 2D-indices**			**3D (S)-stereoisomer**		
Atom	***Local and total non-stochastic chiral linear indices of order 0-2*** (k = 0-2)								
(*i*)	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$
O_1	3.440	5.100	20.860	3.440	5.100	18.860	3.440	5.100	16.860
C_2	2.550	10.430	18.460	2.550	9.430	18.460	2.550	8.430	18.460
***C_3**	3.550	8.260	17.530	2.550	8.260	14.530	1.550	8.260	11.530
C_4	2.550	3.550	8.260	2.550	2.550	8.260	2.550	1.550	8.260
Cl_5	3.160	3.550	8.260	3.160	2.550	8.260	3.160	1.550	8.260
Total	15.250	30.890	73.370	14.250	27.890	68.370	13.250	24.890	63.370
Atom	***Local and total stochastic chiral linear indices of order 0-2*** (k = 0-2)								
(*i*)	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$	$^*f_0(x_i)$	$^*f_1(x_i)$	$^*f_2(x_i)$
O_1	3.440	2.550	3.477	3.440	2.550	3.143	3.440	2.550	2.810
C_2	2.550	3.477	2.637	2.550	3.143	2.637	2.550	2.810	2.637
***C_3**	3.550	2.753	3.506	2.550	2.753	2.906	1.550	2.753	2.306
C_4	2.550	3.550	2.753	2.550	2.550	2.753	2.550	1.550	2.753
Cl_5	3.160	3.550	2.753	3.160	2.550	2.753	3.160	1.550	2.753
Total	15.250	15.880	15.126	14.250	13.547	14.193	13.250	11.213	13.260

The k^{th} total [and local (atom and atom-type) stochastic linear indices], ${}^{s}\boldsymbol{f_k}(x)$ [${}^{s}\boldsymbol{f_k}(x_i)$] are calculated in the same way that the linear indices (non-stochastic), but using k^{th} stochastic molecular pseudograph´s atom adjacency matrix, $\mathbf{S}^k(G)$, like mathematical linear maps' matrices. On the other hand, the defining equation (**1**) for $\boldsymbol{f_k}(x_i)$ may be written as the single matrix equation:

$$\boldsymbol{f_k}(x_i) = [X']^k = \mathbf{M}^{\boldsymbol{k}}[X] \qquad (\mathbf{4})$$

where [X] is a column vector (a nx1 matrix) of the coordinates of X in the canonical basis of $\Re^{\boldsymbol{n}}$ and M^k the k^{th} power of the matrix M of the molecular pseudograph (map's matrix).

Total (whole-molecule) linear indices are *linear functional* (some mathematicians use the term *linear form*, which means the same as linear functional) on $\Re^n$. That is, the k^{th} total linear index is a linear map from $\Re^n$ to the scalar $\Re$ [$\boldsymbol{f_k}$ (x): $\Re^n \rightarrow \Re$] The mathematical definition of these molecular descriptors is the following:

$$f_k(x) = \sum_{i=1}^{n} f_k(x_i) \qquad (\mathbf{5})$$

where n is the number of atoms and $\boldsymbol{f_k}(x_i)$ are the atom's linear indices (linear maps) obtained by Eq. 1. Then, a linear form $\boldsymbol{f_k}(x)$ can be written in matrix form,

$$\boldsymbol{f_k}(x) = [u]^t [X']^k \qquad (\mathbf{6})$$

or

$$\boldsymbol{f_k}(x) = [u]^t \mathbf{M}^{\boldsymbol{k}}[X] \qquad (\mathbf{7})$$

for each molecular vector $X \in \Re^n$. $[u]^t$ is a n-dimensional unitary row vector. As can be seen, the k^{th} total linear index is calculated by summing the local (atom) linear indices of all atoms in the molecule.

3D-Chiral linear indices.

The total and local linear indices, as defined above, can not 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 [25]. In these sense, a chirality molecular vector is obtained (${}^{*}X$), where the components of X (for instance, Pauling electronegativity (X_A) [28] 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, ω_A and an integer parameter, Δ: [25].

ω_A = 1 and Δ is an odd number when *A* has R (rectus), E (entgegen), or *a* (axial) notation according to Cahn-Ingold-Prelog rules (**8**)

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

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

Thus, this 3D-chirality factor $\sin((\omega_A+4\Delta)\pi/2)$ takes different values in order to codify specific stereochemical information such as chirality, Z/E isomerism, and so on. This factor therefore takes values in the following order 1 > 0 > -1 for atoms that have specific 3D environments. The chemical idea here is not that the attraction of electrons by an atom depends on 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 [29]. This correction has principally a mathematical means and must not be source of any misunderstanding. That is to say, this approach can be seen as a simplification of molecular structure. However, in other level of the 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 [30].

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. In this sense, Gonzalez *et al.* [11] 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 ω_A and Δ (GBT-like chirality scale and other alternative chirality scales). In Table 2 clearly shown 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, 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 lest for linear scales [1,10,11].

Table 2. Values of trigonometric 3D-chirality correction factor [sin((ω$_a$+4Δ)π/2)] within the allowed domain.

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

A very interesting point is that the present 3D-chiral descriptor reduces to simples (2D) linear indices ones for molecules without specific 3D characteristics because sin(0+4Δ)π/2 = 0, being Δ zero or any even number. That is, when all the atoms in the molecule are not chiral, the ***TOMOCOMD-CARDD*** (*C*omputed-*A*ided '*R*ational' *D*rug *D*esign) 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, $^{*}\boldsymbol{f_k}(x) = \boldsymbol{f_k}(x)$.

Methods

TOMOCOMD-CARDD approach

For computation of 3D-chiral linear indices we used ***TOMOCOMD*** software [31]. It's 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). In this paper, we used the module ***CARDD*** for the calculation of non-stochastic and stochastic total 3D-chiral linear indices considering and not considering H-atoms in the molecular pseudograph (G).

Chemometric analysis

Statistical analysis was carried out with the STATISTICA software [32]. 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. In connection, we selected the model with a high 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 using an external prediction (test) set. In developing classification models the values of 1 and -1 were assigned to active and inactive compounds, respectively. By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(Active) - P(Inactive)]x100$ or as inactive otherwise. P(Active) and P(Inactive) are the probabilities with which the equations classify a compound as active and inactive, correspondingly.

Finally, the calculation of percentages of global good classification (accuracy) and Matthews's correlation coefficient (MCC) in the training and test sets permitted the assessment of the model [33]. 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. That is to say, MCC quantifies the strength of the linear relation between the molecular descriptors and the classifications, [33] and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages.

We also developed the linear discriminant canonical analysis by checking the following statistic: Canonical regression coefficient (R_{can}), Chi-squared and its *p*-level [$p(\chi^2)$] [34].

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 examining the regression's statistic parameters and of the cross-validation procedures [35,36]. In this sense, the quality of models was determined by examining the determination coefficients (also know as squared regression 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}) [35,37].

QSAR Applications and comparison with other theoretical studies

To evaluate the effectiveness of 3D-chiral linear indices, we have tested their ability to predict pharmacological properties in groups with a known stereochemical influence. First a data set of 32 perindoprilate stereoisomers, an angiotensin-converting enzyme (ACE) inhibitors, was used to test the applicability of the method [11,38]. ACE acts in plasma and blood vessels, removing the C-terminal dipeptide of undecapeptide Angiotesin I to produce the potent blood vessel constricting octapeptide Angiotesin II. In addition, ACE inactivates the hypotensive nonapeptide Bradykinin. For these reasons, ACE is the biological target of many important antihypertensive drugs called ACE inhibitors (ACEIs) [38]. Is this study active is taken to a mean a compound that has an IC_{50} value no higher than 110 nm.

After that, 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 linear indices application. The σ-receptors mediate severe side effects induced by various dopamine antagonists [10].

Finally, in order to validate even more 3D-chiral linear indices in QSAR studies, we select a molecular set that is well-know to QSAR researchers, the so-called Cramer's steroid data set. This data set was introduced by Cramer et al in 1988 [39] using Comparative Molecular Field Analysis (CoMFA) methodology and since then has become a benchmark for the assessment of novel QSAR methods [40,41]. Various groups used this data set to compare the quality of their 3D-QSAR methodologies. Hence, this data set has become one of the most often discussed ones and can be seen as point of reference data set for novel molecular descriptors [42]. Even though this data set is not the ideal 3D benchmark data set, [42] it was used for the shake of comparability [43]. We use this molecular set, because all compounds in this data set contain chiral atoms, and binding affinities of these compounds are available [39]. Some structures of these compounds were drawn incorrectly in the original paper and were corrected in a recent work [41].

Different methods were used to develop 3D-QSAR models for this data set, including CoMFA [39], Comparative Molecular Similarity Indices Analysis (CoMSIA) [44], Molecular Quantum Similarity Measures (MQSM) [45], Topological Quantum Similarity Indices (TQSI) [46], and

Comparative Molecular Moment Analysis (CoMMA) [41], Mapping Property Distributions of Molecular Surfaces (MAP) [43], and so on [47-50].

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

We tested the predictive power of 3D-chiral linear indices in the classification of perindopirilate stereoisomers. The classification obtained models are given below together with the LDA statistical parameters:

$$\textbf{ACEiactv} = 10.818 + 2.85\text{x}10\text{-}5 * {}^{\mathbf{H}}f_{11}(\text{x}) - 2.02\text{x}10\text{-}6 * f_{15}(\text{x}) \quad (\mathbf{9})$$

N = 23 λ = 0.398 $D^2 = 7.82$ F(2, 20) = 15.080 $p < 0.0001$

$$\textbf{ACEiactv} = 64.6484 + 7.5052 \, {}^{*}f_{6}(x) - 8.4588 \, {}^{*}f_{14}(x) \quad (\mathbf{10})$$

N = 23 λ = 0.399 $D^2 = 7.789$ F(2, 20) = 15.020 $p < .0001$

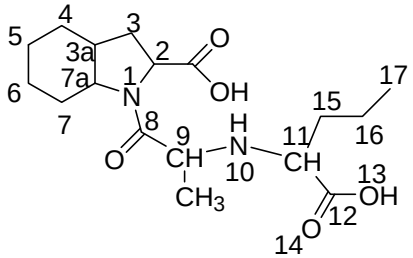
where N is the number of compounds, λ is the Wilks' statistic, D^2 is the squared Mahalanobis distance, F is the Fisher ratio and *p*-value is the significance level.

The model (9), which includes non-stochastic indices, has an accuracy of 100% for the training set. This model showed a high Matthews' correlation coefficient (MCC) of 1. The most important criterion for the acceptance or not of a discriminant model is based on the statistic for external prediction set. Model (9) classifies correctly 100.00% of active (isomers 1, 2 and 4) and 83.33% of inactive (isomers 12, 16, 20, 24 and 28) compounds in the test set, for an accuracy of 88.88% (MCC = 0.79).

In Table 3 we give the basic structure of perindoprilate stereoisomer and their classification in the training and prediction set together with their canonical scores and their posterior probabilities calculated from the Mahalanobis distance.

A very similar behavior was obtained with stochastic linear indices (Eq.10). In this case, the model 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. In addition, model 10 shown an accuracy of 100%, yielding a MCC of 1 for the test set.

Table 3. Basic structure and chirality notation of active and non-active perindoprilate stereoisomers with their posterior probabilities in data split in training and test sets and the canonical scores, using non-stochastic and stochastic linear indices.



No	Comp.[a]	Class[b]	IC_{50}[c]	Class	$\Delta P\%$[d]	Score[e]	Class	$\Delta P\%$[d]	Score[e]
					Eq. 9 (non-Stochastic)			Eq. 10 (Stochastic)	
active compounds									
1	SSRSS*	+	1.1	+	95.43	-2.00	+	76.96	1.79
2	SRSSS*	+	1.2	+	99.60	-2.92	+	97.03	2.60
3	*SSSSS*	+	1.5	+	98.45	-2.41	+	93.66	2.31
4	SRRSS*	+	3.3	+	98.81	-2.51	+	88.67	2.08
5	*SSSSR*	+	12.2	+	97.02	-2.16	+	94.07	2.34
6	*SSRSR*	+	29.4	+	91.34	-1.75	+	78.32	1.82
7	*SRRSR*	+	39.8	+	97.70	-2.26	+	89.39	2.11
8	*SRSSR*	+	54	+	99.22	-2.67	+	97.22	2.63
9	*RRSSS*	+	108	+	0.11	-0.59	-	-48.90	0.63
Non-active compounds									
10	*SSSRS*	-	$1.1x10^3$	-	-41.42	-0.26	-	-88.42	-0.02
11	*RSSSS*	-	$1.9x10^3$	-	-59.03	-0.09	-	-72.69	0.34
12	SSRRR*	-	$2.6x10^3$	-	-86.64	0.39	-	-96.74	-0.51
13	*RRSSR*	-	$5.5x10^3$	-	-31.80	-0.35	-	-46.24	0.65
14	*SSRRS*	-	$7.1x10^3$	-	-75.65	0.15	-	-96.96	-0.54
15	*RRSRS*	-	$7.8x10^3$	-	-99.35	1.55	-	-99.86	-1.70
16	RSRRR*	-	$23x10^3$	-	-99.97	2.72	-	-99.98	-2.48
17	*SRRRR*	-	$33x10^3$	-	-56.44	-0.12	-	-93.06	-0.22
18	*RSSSR*	-	$36x10^3$	-	-76.52	0.16	-	-71.02	0.36
19	*RSRSR*	-	$47x10^3$	-	-91.48	0.57	-	-91.83	-0.16
20	RSRSS*	-	$60x10^3$	-	-84.13	0.32	-	-92.35	-0.18
21	*RRRRR*	-	10^5	-	-99.89	2.21	-	-99.96	-2.19
22	*SRRRS*	-	10^5	-	-29.93	-0.36	-	-93.51	-0.24
23	*RRRSS*	-	10^5	-	-49.77	-0.19	-	-84.11	0.11
24	SRSRR*	-	10^5	-	-9.17	-0.53	-	-75.00	0.30
25	*RRRRS*	-	10^5	-	-99.78	1.96	-	-99.97	-2.22
26	*RRSRR*	-	10^5	-	-99.67	1.80	-	-99.85	-1.67
27	*SSSRR*	-	10^5	-	-64.76	-0.02	-	-87.65	0.01
28	RSSRS*	-	10^5	-	-99.83	2.06	-	-99.94	-1.99
29	*RRRSR*	-	10^5	-	-70.48	0.06	-	-83.07	0.14
30	*RSSRR*	-	10^5	-	-99.91	2.31	-	-99.93	-1.97
31	*RSRRS*	-	10^5	-	-99.94	2.47	-	-99.98	-2.51
32	SRSRS*	-	10^5	+	23.42	-0.77	-	-76.47	0.27

*Compounds used in the Test set. [a]Notation of the chiral centres in each perindoprilate derivative in the following order C_2, C_{3a}, C_{7a}, C_9, C_{11}. [b]Classification according to the value of the IC_{50}. [c]Values of the IC_{50}, of the compound, for ACE in nM taken from the references 11, 25 and 38. [d]ΔP Posterior probability predicted for each compound using Eq. 9 and Eq. 10. [e]Canonical scores predicted using canonical analysis.

Table 4 depicts the obtained results in our study as well as the achieved with other cheminformatic approaches. First, it is remarkable that our model contain one variable less than the model obtained with MARCH-INSIDE molecular descriptors [11] and the same number of variables that Marrero-Ponce *et al*. [25] used for develop their model using other 3D-chiral ***TOMOCOMD-CARDD*** descriptors. However, the accuracy of the model 9 for the training set is the best of all equations for this data set. In the model 10 this parameter, for the training and test set, are equal to the obtained when the 3D-chiral quadratic indices [25] were used and both are better than obtained for González-Díaz *et al.* (see Table 4). [11]

On the other hand, canonical analysis is used here to test both the ability of 3D-chiral non-stochastic and stochastic linear indices to discriminate between the two groups of stereoisomers and also to order these compounds accordingly with their stability profile.

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 non-stochastic and stochastic linear indices & LDA ACEinhibitory activity canonical analysis principal root are given below:

$$\textbf{ACEroot} = -4.643 - 1.1x10^{-5} * f^{H}_{11}(x) + 7.54x10^{-7} * f_{15}(x) \quad (\textbf{11})$$

$N = 23 \;\; \lambda = 0.398 \;\; R_{can} = 0.78 \;\; \chi^2 = 18.39 \;\; mean(+) = -1.98 \;\; mean(-) = 0.70 \;\; p < 0.0001$

$$\textbf{ACEroot} = 25.27 + 2.81^{*} f_{6}(x) - 3.172^{*} f_{14}(x) \quad (\textbf{12})$$

$N = 23 \;\; \lambda = 0.399 \;\; R_{can} = 0.77 \;\; \chi^2 = 18.34 \;\; mean(+) = 1.97 \;\; mean(-) = -0.70 \;\; p < 0.0001$

The canonical transformation of the LDA results with non-stochastic and stochastic 3D-chiral linear indices gives rise to canonical roots with good canonical regression coefficients of 0.78 and 0.77, respectively. Chi-squared test permits us to asses the statistical signification of this analysis as having a *p*-level <0.0001.

When LDA analysis is applied to solve the two-group classification problem we ever find two classification functions. However, we cannot use these two classification functions to evaluate all the compounds and obtain a bivariate stability map because they are not orthogonal [34]. To solve this problem we used canonical analysis in this case the dimensional reduction caused by canonical analysis makes possible to obtain a 1-dimension stability map [34].

Table 4. 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 Linear indices (Eq. 9)	2	0.398	7.82	100.00%	88.88%	15.08
Stochastic Linear indices (Eq. 10)	2	0.399	7.789	95.65%	100.00%	15.02
MARCH-INSIDE molecular descriptors[11]	3	0.38	8.43	91.30%	88.88%	10.30
Non-Stochastic Quadratic indices[25]	2	0.42	7.12	95.65%	100.00%	13.73

N: number of used compounds. **n**: number of parameter in the obtained model.

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. For example, we can detect an overall ascendant tendency of canonical scores of equation (11) when they are plotted in the same order in which IC_{50} increases (activity decreases). As it is expected, the over all mean of canonical root scores for the group of active isomers (lowest IC_{50} values) has an opposite sign (-) with respect to the other group [(+); highest IC_{50} values] [34].

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

We will now discuss the ability of 3D-chiral linear indices to predict σ receptor antagonist activities. 3D-linear indices are non-symmetric and reduce to classical descriptors when symmetry is not codified (see Table 1). Moreover, Gónzalez-Díaz *et al*. conclude that σ receptor antagonist activities are not a pseudoscalar property [11] and we can expect at least a good correlation with 3D-linear indices.

This experiment also permitted us to compare our method with others previously reported approaches. The MLR analysis was used to develop QSAR models for the σ receptor antagonist activities. The obtained models using non-stochastic linear indices are the follow:

$$\mathbf{logIC_{50}(\sigma)} = -8.9207(\pm 0.8388) + 0.5304(\pm 0.0695)^{*}\boldsymbol{f_0}(x) - 0.0065(\pm 0.0011)^{*}\boldsymbol{f}^{H}_{3}(x) \qquad (\mathbf{13})$$

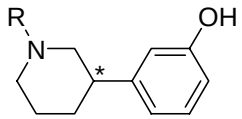
$N = 14 \quad R^2 = 0.939 \quad q^2_{LOO} = 0.909 \quad F(2, 11) = 84.876 \quad s = 0.271 \quad s_{cv} = 0.305 \quad p < 0.0001$

$$\mathbf{logIC_{50}(\sigma)} = -9.4831(\pm 0.4984) + 0.5886(\pm 0.0419)^{*}\boldsymbol{f_0}(x) - 0.0074(\pm 0.0007)^{*}\boldsymbol{f}^{H}_{3}(x) \qquad (\mathbf{14})$$

$N = 13 \quad R^2 = 0.982 \quad q^2_{LOO} = 0.966 \quad F(2, 10) = 265.66 \quad s = 0.157 \quad s_{cv} = 0.186 \quad p < 0.0001$

where, N is the size of the data set, R^2 is the squared regression coefficient (determination coefficient), *s* is the standard deviation of the regression, F is the Fischer ratio and q^2 (s_{cv}) are the squared correlation coefficient (standard deviation) of the cross-validation performed by the LOO procedure. This statistics indicate that these models are appropriate for the description of chemicals studied here. In the Table 5 are show the structure and values of experimental and calculated Log IC_{50} for this data set.

Table 5. Results of multivariate regression analysis of the log IC_{50} of a group of *n*-alkylated 3-(3-hidroxyphenyl)piperidines for the σ-receptor.



Compound	**Log IC_{50} (σ-receptor)**						
(Alkyl group)[a]	Obs.[b]	Cal.[c]	Res.[d]	Cal.[e]	Res.[d]	Cal.[f]	Res.[d]
(R)-3-HPP							
H	-0.66	-0.54	-0.12	-0.54	-0.12	-0.48	-0.18
CH_3	0.43	0.13	0.30	0.18	0.25	0.28	0.15
C_2H_5	0.95	0.72	0.23	0.81	0.14	0.70	0.25
n-C_3H_7	1.52	1.32	0.20	1.45	0.07	1.45	0.07
i-C_3H_7	0.61	1.30	-0.69	*outlier*	-	0.84	-0.23
n-C_4H_9	2.05	1.93	0.12	2.09	-0.04	1.89	0.16
2-Phenylethyl	2.10	2.22	-0.12	2.24	-0.14	2.41	-0.31
(S)-3-HPP							
H	-1.19	-1.09	-0.10	-1.13	-0.06	-0.80	-0.39
CH_3	-0.28	-0.42	0.14	-0.42	0.14	-0.56	0.28
C_2H_5	-0.01	0.17	-0.18	0.21	-0.22	0.19	-0.20
n-C_3H_7	0.81	0.77	0.04	0.85	-0.04	0.57	0.24
i-C_3H_7	0.68	0.75	-0.07	0.83	-0.15	0.62	0.06
n-C_4H_9	1.51	1.37	0.14	1.49	0.02	1.18	0.33
2-Phenylethyl	1.80	1.67	0.13	1.65	0.15	2.03	-0.23

[a]Alkylic (R) group at nitrogen ring. [b]Observed values of the Log IC_{50} for the σ-receptor taken from Ref. 10, 11 and 25. [c]Values calculated from Eq. 13. [d]Residual, defined as [Log IC_{50} (σ)Obs – Log IC_{50} (σ)Cal]. [e]Values calculated from Eq. 14. [f] Values calculated from Eq. 15. Abbreviations: HPP, *N*-alkylated 3-Hydroxyphenyl piperidines.

In the development of the first quantitative model for description of activities (Eq.13), one compound was detected as statistical outlier. Once rejected the statistical outlier, the Eq. 14 was obtained with better statistical parameters.

When the stochastic linear indices were used, the obtained model for the σ receptor antagonist activities is given below:

$$\mathbf{logIC_{50}(\sigma)} = -5.9421(\pm0.5197) +0.8067(\pm0.2739)\,{}^{*}f^{H}_{14}(x) -0.7329(\pm0.2741)\,{}^{*}f^{H}_{11}(x) \qquad (\mathbf{15})$$

$N = 14 \quad R^2 = 0.941 \quad q^2_{LOO} = 0.90 \quad F(2, 11) = 87.932 \quad s = 0.267 \quad s_{cv} = 0.319 \quad p<0.0001$

The comparison with other methods previously reported for the same activity is shown in Table 6. As it can be seen, our models have statistical parameter slightly better than models obtained with MARCH-INSIDE molecular descriptors [11] and other chiral TIs [10], and our statistics are very similar that obtained by Marrero-Ponce et al. [25] when 3D-chiral quadratic indices were used. Once rejected the statistical outlier our model show better predictive abilities ($R^2 = 0.982$, $s = 0.157$, $q^2 = 0.966$ and $s_{cv} = 0.186$) than model built with 3D-chiral quadratic indices ($R^2 = 0.977$, $s = 0.175$, $q^2 = 0.957$ and $s_{cv} = 0.211$) [25].

Table 6. 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**	$\mathbf{R^2}$	***s***	$\boldsymbol{q^2}$	$\boldsymbol{s_{CV}}$	**F**
Non-Stochastic Linear Indices (Eq. 13)	14	2	0.939	0.271	0.909	0.305	84.87
Non-Stochastic Linear Indices (Eq. 14)	13	2	0.982	0.157	0.966	0.186	256.66
Stochastic Linear Indices (Eq. 15)	14	2	0.941	0.267	0.90	0.319	87.93
Chiral TIs[10]	14	3	0.931	0.301	*	*	45.70
MARCH-INSIDE molecular descriptors[11]	14	2	0.922	0.295	*	0.32	71.17
Non-Stochastic Quadratic indices[25]	14	2	0.940	0.270	0.912	0.289	85.82
Non-Stochastic Quadratic indices[25]	13	2	0.977	0.175	0.957	0.211	211.20

*Values are not reported in the literature.

Prediction of the Corticosteroid-Binding Globulin (CBG) binding affinity of a Steroid family.

The training set used to validate our methodology is made up of 31 molecules. Table 7 gathers the entire studied set with the actual binding affinities, taken from Robert *et al.* [45]. Due to the studied steroid molecular structures have been already depicted in several papers, they will not be included here. For more details see, for example Figure 1 in reference 39 or Figure 1 in reference 41.

This study also permitted us to compare our method with others 3D QSAR methods such as MQMS, MaP, CoMMA, TQSAR and so on. The MLR analysis was used to develop QSAR models for the corticosteroid-binding globulin binding affinity. The obtained models using non-stochastic linear indices are the follow:

Table 7. Results of the steroids data set used for QSAR study.

		Observed CBG affinity (pKa)[a]	Pred. value[b]	% E[c]	%E_{cv}[d]	Pred. value[b]	% E[c]	%E_{cv}[d]
1	Aldosterone	-6.279	-6.149	2.063	2.396	-6.222	0.902	2.497
2	Androstanediol	-5.000	-5.161	-3.225	-5.187	-4.984	0.324	0.394
3	Androstenediol	-5.000	-4.965	0.692	0.875	-4.930	1.401	1.721
4	Androstenedione[e,f]	-5.763	-6.691	-16.096	-20.067	-6.583	-14.231	-17.342
5	Androsterone	-5.613	-5.265	6.197	7.865	-5.342	4.826	6.399
6	Corticosterone	-7.881	-7.283	7.588	8.857	-7.535	4.389	5.397
7	Cortisol	-7.881	-7.380	6.351	7.955	-7.794	1.100	1.475
8	Cortisone	-6.892	-6.892	0.004	0.006	-7.222	-4.793	-6.438
9	Dehydroepiandrosterone	-5.000	-5.094	-1.879	-2.296	-5.033	-0.652	-0.750
10	Deoxycorticosterone[f]	-7.653	-7.307	4.522	5.294	-6.820	10.885	12.194
11	Deoxycortisol	-7.881	-7.522	4.560	5.089	-7.202	8.618	9.710
12	Dihydrtestosterone	-5.919	-5.700	3.697	4.672	-6.025	-1.783	-2.380
13	Estradiol	-5.000	-4.803	3.946	5.880	-4.888	2.232	3.936
14	Estriol	-5.000	-5.194	-3.884	-5.544	-5.071	-1.421	-2.536
15	Estrone	-5.000	-4.960	0.808	1.679	-4.954	0.912	1.723
16	Ethiocholanolone	-5.255	-5.265	-0.194	-0.246	-5.342	-1.658	-2.198
17	Pregnenolone	-5.255	-5.450	-3.720	-4.537	-5.529	-5.220	-5.980
18	17-Hydroxyregnenolone	-5.000	-5.463	-9.264	-13.865	-5.405	-8.107	-10.835
19	Progesterone	-7.380	-6.730	8.814	9.652	-6.889	6.649	7.622
20	17-Hydroxyprogesterone[f]	-7.740	-7.025	9.238	10.883	-6.954	10.150	11.731
21	Testosterone	-6.724	-6.535	2.810	3.316	-6.480	3.630	4.159
22	Prednisolone	-7.512	-7.735	-2.972	-4.857	-7.687	-2.335	-3.273
23	Cortisolacetate	-7.553	-7.700	-1.943	-2.751	-7.647	-1.247	-3.642
24	4-Pregnene-3,11,20-trione	-6.779	-6.441	4.983	6.873	-7.007	-3.358	-4.393
25	Epicorticosterone	-7.200	-7.441	-3.344	-3.965	-7.695	-6.877	-9.164
26	19-Nortestosterone[e]	-6.144	-6.858	-11.616	-14.222	-6.758	-9.991	-12.091
27	16a,17a-Dihydroxyprogesterone[e]	-6.247	-7.439	-19.079	-21.199	-6.118	2.060	3.135
28	16a-Methylprogesterone	-7.120	-6.793	4.588	5.352	-7.239	-6.195	-7.372
29	19-Norprogesterone	-6.817	-7.019	-2.967	-3.570	-7.927	-3.108	-4.072
30	2a-Methylcortisol	-7.688	-7.773	-1.100	-1.374	-5.864	-1.148	-2.083
31	2a-Methyl-9a-fluorocortisol	-5.797	-5.940	-2.459	-4.541	-6.824	4.152	4.755

[a]Observed CBG affinity values taken from ref 45; [b]Predicted CBG affinity values using Eq.16; [b]Predicted CBG affinity values using Eq.18; [c]Percent of relative error; %E = 100x[Obs-Pred/Obs]. [d]Percent of relative error in leave-one-out cross-validation procedure; $\%E_{cv}$ = 100x[Obs-$Pred_{LOO\text{-}CV}$/Obs]. [e]Compounds detected as outlier in Eq. 16. [f]Compounds detected as outlier in Eq. 18.

$$\mathbf{CBG} = -6.396(\pm0.087) \; -7.596(\pm0.999)^{*}\boldsymbol{f}_{L14}(x_{\mathbf{E}}) \; -4.528(\pm1.816)^{*}\boldsymbol{f}_{4}(x) \; -6.696(\pm2.399)^{*}\boldsymbol{f}_{2}(x) \; +16.289(\pm2.908)^{*}\boldsymbol{f}_{L11}(x_{\mathbf{E}}) \; -9.603(\pm2.380)^{*}\boldsymbol{f}_{L7}(x_{\mathbf{E}}) \; -2.269(\pm0.662)^{*}\boldsymbol{f}_{0}(x) \qquad (\mathbf{16})$$

$N = 31 \quad R^2 = 0.84 \quad q^2_{LOO} = 0.77 \quad F(6, 24) = 21.060 \quad s = 0.48 \quad s_{cv} = 0.52 \quad p<0.0001$

$$\mathbf{CBG} = -6.511(\pm0.057) \; -2.297(\pm0.423)^{*}\boldsymbol{f}_{0}(x) \; -8.329(\pm1.512)^{*}\boldsymbol{f}_{2}(x) \; -5.782(\pm1.143)^{*}\boldsymbol{f}_{4}(x) \; -12.424(\pm1.527)^{*}\boldsymbol{f}_{L7}(x_{\mathbf{E}}) \; +19.908(\pm1.877)^{*}\boldsymbol{f}_{L11}(x_{\mathbf{E}}) \; -8.790(\pm0.651)^{*}\boldsymbol{f}_{L14}(x_{\mathbf{E}}) \qquad (\mathbf{17})$$

$N = 28 \quad R^2 = 0.946 \quad q^2_{LOO} = 0.904 \quad F(6, 21) = 61.765 \quad s = 0.296 \quad s_{cv} = 0.349 \quad p<0.0001$

In the development of the quantitative model (Eq.16), three compounds (**4**, **26** and **27**) were detected as statistical outlier. Once rejected the statistical outliers, a new model (Eq. 17) was obtained with better statistical parameters. As can be seen this new model explains more than the 94% of the variance of the experimental CBG values. These two models uses six variables each one to describe 31 and 28 steroids, correspondingly.

In addition, using stochastic linear fingerprints to describe the CBG binding affinity we obtained two models which are given below:

$$\mathbf{CBG} = -6.408(\pm 0.080) \; -6.218(\pm 1.388)^{*}\boldsymbol{f}_{L8}(x_{\mathbf{E}}) \; +5.024(\pm 1.000)^{*}\boldsymbol{f}_{L9}(x_{\mathbf{E}}) \; -4.647(\pm 1.060)^{*}\boldsymbol{f}_{L2}(x_{\text{E-H}}) + 1.172(\pm 0.628)^{*}\boldsymbol{f}_{L4}(x_{\mathbf{E}}) + 13.850(\pm 3.568)^{*}\boldsymbol{f}_{L4}(x_{\text{E-H}}) - 13.145(\pm 3.569)^{*}\boldsymbol{f}_{L6}(x_{\text{E-H}}) + 3.386(\pm 1.013)^{*}\boldsymbol{f}_{L9}(x_{\mathbf{E}}) \quad (\mathbf{18})$$

$N = 31 \quad R^2 = 0.87 \quad q^2_{LOO} = 0.787 \quad F(7, 23) = 22.863 \quad s = 0.437 \quad s_{cv} = 0.52 \quad p<0.0001$

$$\mathbf{CBG} = -6.383(\pm 0.066) \; -5.605(\pm 1.088)^{*}\boldsymbol{f}_{L8}(x_{\mathbf{E}}) \; +4.491(\pm 0.786)^{*}\boldsymbol{f}_{L9}(x_{\mathbf{E}}) \; -4.894(\pm 0.841)^{*}\boldsymbol{f}_{L2}(x_{\text{E-H}}) + 1.107(\pm 0.497)^{*}\boldsymbol{f}_{L4}(x_{\mathbf{E}}) + 15.003(\pm 2.789)^{*}\boldsymbol{f}_{L4}(x_{\text{E-H}}) - 14.277(\pm 2.780)^{*}\boldsymbol{f}_{L6}(x_{\text{E-H}}) + 3.679(\pm 0.788)^{*}\boldsymbol{f}_{L9}(x_{\mathbf{E}}) \quad (\mathbf{19})$$

$N = 28 \quad R^2 = 0.92 \quad q^2_{LOO} = 0.88 \quad F(7, 20) = 35.773 \quad s = 0.\ 338 \quad s_{cv} = 0.368 \quad p<0.0001$

In the development of the quantitative model (Eq.18), three compounds were also detected as statistical outlier. Once rejected these chemicals (**4**, **10** and **20**), a new model (Eq.19) was obtained with better statistical parameters. Notice that this new model explains more than the 92% of the variance of the experimental CBG values. These two models uses seven variables each one to describe 31 and 28 steroids, respectively.

All these results are summarized in Table 8, where a comparison with other computational scheme can be more easily performed. Nevertheless notice that the present QSAR method, non-stochastic and stochastic 3D-chiral linear indices, obtains comparable results to other highly predictive QSAR models; even when they use more sophisticated statistic methods such as: partial least squared, principal components analysis, non-linear neural network techniques and so on. Many of the models objects of comparison were obtained from different procedures based on quantum mechanics and/or geometric principles as well as molecular mechanic approaches.

Table 8. Comparison of ***TOMOCOMD-CARDD*** descriptors prediction for the steroid data set with other 3D QSAR approaches.

QSAR Method	N	n	Statistic Method	q^2	ref.
Similarity matrixes-based molecular descriptors	31	6	genetic NN	0.940	49
TOMOCOMD-CARDD *non-stochastic*	28	6	MLR	0.904	Eq. 17
TOMOCOMD-CARDD *stochastic*	28	7	MLR	0.882	Eq. 19
MaP	29	4	PCR-VS	0.880	43
TQSAR	31	6	MLR after PCA	0.842	45
TOMOCOMD-CARDD *stochastic*	31	7	MLR	0.788	Eq. 18
TQSI	31	3	MLR	0.775	46
TOMOCOMD-CARDD *non-stochastic*	31	6	MLR	0.767	Eq. 16
Similarity indices	31	1	PLS	0.734	48
MQMS	31	3	MLR and PCA	0.705	46
CoMMA	31	6	PCR	0.689	41
MaP	29	4 (168)	PLS	0.630	43
Wagener's	31	-	k-NN and FNN	0.630	47
MaP	29	5 (168)	PCR	0.530	43

N: number of steroids. **n:** number of variables. q^2**:** leave-one-out cross-validated coefficient of determination.

Final conclusions

Our studies demonstrated that 3D-chiral linear indices can be successfully applied in QSAR studies which include chiral molecules. Therefore, we suggest that 2D-QSAR methods enhanced by chirality descriptors present a powerful alternative to popular 3D-QSAR approaches.

We have shown here that 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 σ-receptors, and to predict the corticosteroid-binding globulin binding affinity of the Cramer's steroid data set. This result is only a preliminary conclusion and a deeper analysis of the potential of the 3D-chiral linear indices is necessary. However, we show that for three data sets chiral-QSAR models that use 3D-chiral linear indices had better or similar predictive ability as compared to other previously reported chiral and/or 3D-QSAR Methods.

References

[1] Golbraikh, A., Bonchev, D., Tropsha, A., J. Chem. Inf. Comput. Sci., 41 (2001) 147.

[2] de Julián-Ortiz, J.V., García-Doménech, R., Gálvez, J., Soler-Roca, R., Garcia-March, F.J., Anton-Fos, G.M., J. Cromat., 719 (1996) 37.

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

[4] Schumacher, H., Blake, D.A., Gurian, J.M., Gillette, J.R., J. Pharmacol. Exp. Ther., 160 (1968) 189.

[5] Stinson, S.C., Chem. Eng. News., 78 (2000) 43.

[6] Buda, A.B., Mislow, K., J. Mol. Struct., (Theochem) 232 (1991) 1.

[7] Avnir, D., Hel-Or, H.Z., Mezey, P.G., In: Schleyer, P.V.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer III, H.F., Schreiner P.R., (Eds.) Symmetry and Chirality: Continuos Measures, The Encyclopedia of Computational Chemistry, Vol. 4, Wiley, Chichester, 1998.

[8] Zabrodsky, H., Avnir, D., J. Am. Chem. Soc., 117 (1995) 462.

[9] Schultz, H.P., Schultz, E.B., Schultz, T.P., J. Chem. Inf. Comput. Sci., 35 (1995) 864.

[10] de Julián-Ortiz, J.V., de Alapont, C.G., Ríos-Santamarina, I., García-Doménech, R., Gálvez, J., J. Mol. Graphics Mod., 16 (1998) 14.

[11] González-Díaz, H., Hernández-Sánchez, I., Uriarte, E., Santana, L., Comput. Biol. Chem., 27 (2003**)** 217.

[12] Estrada, E., Uriarte, E., Curr. Med. Chem., 8 (2001**)** 1699.

[13] Marrero-Ponce, Y., Molecules., 8 (2003) 687.

[14] Marrero-Ponce, Y., J. Chem. Inf. Comput. Sci., 44 (2004) 2010.

[15] Marrero-Ponce, Y., Cabrera, M.A., Romero, V., Ofori, E., Montero L.A., Int. J. Mol. Sci., 4 (2003**)** 512.

[16] Marrero-Ponce, Y., Castillo-Garit, J. A., Torrens, F., Romero-Zaldivar, V., Castro, E. Molecules., 9 (2004) 1100.

[17] Marrero-Ponce, Y. Bioorg. Med. Chem., 12 (2004) 6351.

[18] Marrero-Ponce, Y., Cabrera, M.A., Romero, V., González, H.D., Torrens, F., J. Pharm. Pharm. Sci., 7 (2004) 186.

[19] Marrero-Ponce, Y., Castillo-Garit, J.A., Olazábal, E., Serrano, H.S., Morales, A., Castañedo, N., Ibarra-Velarde, F., Huesca-Guillen, A., Jorge, E., del Valle, A., Torrens, F., Castro, E.A., J. Comput. Aided Mol. Des., 18 (2004) 615

[20] Marrero-Ponce, Y., Montero-Torres, A., Romero-Zaldivar, C., Iyarreta-Veitía, M., Mayón-Peréz, M., García-Sánchez, R., Bioorg. Med. Chem., 13 (2005) 1293.

[21] Marrero-Ponce, Y., Cabrera, M. A., Romero-Zaldivar, V., Bermejo, M., Siverio, D., Torrens, F. Internet Electronic J. Mol. Des., 4 (2005) 124

[22] Marrero-Ponce, Y., Castillo-Garit, J.A., Olazábal, E., Serrano, H. S., Morales, A., Castañedo, N., Ibarra-Velarde, F., Huesca-Guillen, A., Sánchez, A. M., Torrens, F., Castro, E. A. Bioorg. Med. Chem., 13 (2005) 1005.

[23] Marrero-Ponce, Y., Nodarse, D., González-Díaz, H., Ramos de Armas, R., Romero-Zaldivar, V., Torrens, F., Castro, E., Int. J. Mol. Sci., 5 (2004**)** 276.

[24] Marrero-Ponce, Y., Castillo-Garit, J.A., Nodarse, D., Bioorg. Med. Chem., 13 (2005) 3397.

[25] Marrero-Ponce, Y., González-Díaz, H., Romero, V., Torrens, F., Castro, E.A., Bioorg. Med. Chem., 12 (2004**)** 5331.

[26] Marrero-Ponce, Y., Medina, R., Castro, E. A., de Armas, R., González, H., Romero, V., Torrens, F., Molecules., 9 (2004) 1124.

[27] Marrero-Ponce, Y., Medina, R., Castillo-Garit, J.A., Romero, V., Torrens, F., Castro, E.A., Bioorg. Med. Chem., 13 (2005) 3003.

[28] Pauling, L., The Nature of Chemical Bond, Cornell University Press, New York, 1939.

[29] Eliel, E., Wilen, S., Mander, L., Stereochemistry of Organic Compounds, John Wiley & Sons Inc, 1994.

[30] M. J. S. Dewar, J., Phys. Chem., 89 (1985) 2145.

[31] Marrero-Ponce, Y., Romero, V., ***TOMOCOMD*** software. Central University of Las Villas, 2002. ***TOMOCOMD*** (**TO**pological **MO**lecular **COM**puter **D**esign) 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

[32] STATISTICA version. 6.0, Statsoft, Inc.

[33] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., Nielsen. H., Bioinformatics, 16 (2000) 412.

[34] Ford, M.-G., Salt, D.-W. The Use of Canonical Correlation Analysis; In Chemometric methods in molecular design; van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995, p. 283-292.

[35] Wold, S., Erikson, L. Statistical Validation of QSAR Results. Validation Tools; In Chemometric Methods in Molecular Design, van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995, p. 309-318.

[36] Belsey, D. A.; Kuh, E.; Welsch, R. E. Regression Diagnostics, Wiley: New York, 198

[37] Golbraikh, A.; Tropsha, A., J. Mol. Graph. Modell., 20 (2002) 269.

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

[39] Cramer, R.D. III., Patterson, D.E., Bunce, J.D., J. Amer. Chem. Soc., 110 (1988) 5959.

[40] Coats, E.A. In 3D QSAR in Drug Design.V.3. Kubinyi, H., Folkers, G., Martin, Y.C., Eds., Kluwer/ESCOM: Dordrecht, 1998, pp 219-213.

[41] Silverman, B.D., Quant. Struct.-Act. Relat., 19 (2000) 237.

[42] Coats, E.A., Perspect. Drug Discovery Des., 12-14 (1998) 199.

[43] Stief, N., Baumann, Knutt., J. Med. Chem., 46 (2003) 1390.

[44] Klebe, G., Abraham, U., Mietzner, T., J. Med. Chem., 37 (1994) 4130.

[45] Robert, D., Amat, L., Carbo-Dorca, R., J. Chem. Inf. Comp. Sci., 39 (199) 333.

[46] Lobato, M, Amat, L., Besalu, E., Carbo-Dorca, R., J. Chem. Inf. Comp. Sci., 39 (1998) 465.

[47] Wagener, M., Sadowski, J., GAsteiger, J. J. Am. Chem. Soc., 117 (1995) 7769.

[48] Parretti, M.F., Kroemer, R.T., Rothman, J.H., Richards, W.G. J. Comp. Chem., 18 (1997) 1344.

[49] So, S.S., Karplus, M., J. Med. Chem., 40 (1997) 4347.

[50] Chen, H., Zhou, J., Xie, G., J. Chem. Inf. Comp. Sci., 39 (1998) 243.