

3D-chiral (2.5) atom-based *TOMOCOMD-CARDD* descriptors: theory and QSAR applications to central chirality codification

Yovani Marrero-Ponce ·
Juan Alberto Castillo-Garit · Eduardo A. Castro ·
Francisco Torrens · Richard Rotondo

Received: 4 October 2007 / Accepted: 23 October 2007 / Published online: 20 May 2008
© Springer Science+Business Media, LLC 2008

Abstract The history of the use of chiral descriptors in Quantitative structure–activity relationships (QSAR) studies is described, with a particular emphasis on several series of novel chirality descriptors that have been introduced in this field. Specifically, chiral topological indices that circumvent the inability of conventional (chiral insensitive) topological molecular descriptors are reviewed. These modified descriptors were applied to several well-know data sets in order to validate each one of them. Particularly, Cramer’s steroid data set has become a benchmark for the assessment of novel QSAR methods. This data set has been used by several researches using 3D-QSAR approaches such as Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Indices Analysis (CoMSIA), Molecular Quantum Similarity Measures (MQSM), 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors,

Y. Marrero-Ponce (✉) · J. A. Castillo-Garit
Unit of Computer-Aided Molecular “Biosilico” Discovery and Bioinformatic Research
(CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Central University of Las Villas,
Santa Clara 54830, Villa Clara, Cuba
e-mail: ymarrero77@yahoo.es; yovani.marrero@uv.es; ymponce@gmail.com;
yovanimp@qf.uclv.edu.cu
URL: <http://www.uv.es/yoma/>

Y. Marrero-Ponce · J. A. Castillo-Garit · F. Torrens
Institut Universitari de Ciència Molecular, Universitat de València, Edifici d’Instituts de Paterna,
P. O. Box 22085, 46071 Valencia, Spain

J. A. Castillo-Garit
Applied Chemistry Research Center, Central University of Las Villas,
Santa Clara 54830, Villa Clara, Cuba

E. A. Castro
INIFTA, División Química Teórica, Suc. 4, C.C.16,
La Plata 1900, Buenos Aires, Argentina

R. Rotondo
Mediscovery, Inc., Suite 1050, 601 Carlson Parkway, Minnetonka, MN 55305, USA

Topological Quantum Similarity Indices (TQSI), similarity matrixes, Comparative Molecular Moment Analysis (CoMMA), E-state, Mapping Property Distributions of Molecular Surfaces (MAP), EVA and so on. For that reason, it was selected for the shake of comparability by us. An extensive comparison between all these approaches was updated. In addition, to evaluate the effectiveness of this novel approach in drug design we have modelled the angiotensin-converting enzyme inhibitory activity of perindoprilate's σ -stereoisomers combinatorial library as well as 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. The validation of this method was achieved by comparison with previous reports applied to the same data sets. The non-stochastic and stochastic 3D-chiral (2.5) bilinear indices appear to provide a very interesting alternative to other more common 3D-QSAR descriptors.

Keywords Non-Stochastic and Stochastic 3D-chiral (2.5) atom-based TOMOCOMD-CARDD descriptors · 3D-QSAR · Angiotensin-converting enzyme inhibitors · σ -Receptor antagonists · Binding affinity of steroids

1 Background

Asymmetry of atomic configurations is very important feature in determining the physical, chemical and biological properties of chemicals substances [1]. In the literature, the asymmetric atoms are often referred to as chiral atoms and molecules containing chiral atoms are referred to as chiral molecules. Two molecules with identical chemical formulas but different states of symmetry of only one atom are referred to as 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].

However, if a molecule contains chiral atoms, it can be an achiral molecule, because it may present a symmetric element such as symmetric centre or plane (i.e. meso-compounds: tartaric acid). In addition, chirality can be caused by a spatial isomerism resulting from the lack of free rotation around single or double bonds such as in derivatives of biphenyl or in allenes, rather than due to chiral atoms [3,4]. Thus, a necessary and sufficient condition to consider a compound as chiral is the absence of an element of symmetry which avoids chirality. That is to say, if there is not an inverse axis (S_n) of symmetry (including the plane and centre as self-containing cases of S_n) we can affirm that we are in the presence of a chiral molecule (having or not chiral atoms) [3,5]. In the original definition, Kelvin formulated the concept of chirality as an abstract property of geometric objects: “*I call any geometrical figure, or group of points, chiral ... if its image in a plane mirror, ideally realized cannot be brought to coincide with itself*” [6,7]. Although enantiomorphs cannot be placed upon each other (superimposed or overlapped) so that at least parts of them coincide.

Hence, chirality cannot be equated with asymmetry (i.e., the total absence of symmetry) because all molecules present at least one simple axis of symmetry. Anyhow, this element does not preclude chirality. For this reason, in the past the word

‘dissymmetry’ was often used as a synonym for what we now call chirality. Pasteur was well aware of the difference between ‘dissymmetry’ and asymmetry, as evidenced by the French title of his lecture ‘Recherches sur la Dissymétrie moléculaire des produits Organiques naturels’ [8]. Unfortunately, this was translated into English as ‘Researches on the Molecular Asymmetry of Natural Organic Products’ [9]. The word dissymmetry, in the sense of what we now call chirality seems to have been lost to the English language over time [5].

Most of the physical as well as chemical properties of chiral molecules are similar. At the same time, it is well known that many biological molecules are chiral and that the chirality plays an essential role in defining biological activity [1]. 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. In addition, they may show differences in behaviour (both in reaction rates and in product stereochemistry) in their interactions with a chiral reagent. In this sense, many biochemical processes and phenomena are stereospecific. For instance, L- and D-enantiomers of amino acids have different tastes [10, 11], enantiomers of some compounds have different odours [12, 13], and many medicinal preparations have physiological properties different from those of their enantiomers [14–16]. The case of thalidomide is an example of a problem that was, at least, complicated by the ignorance of stereochemical effects [17]. 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 [18].

Based on the experience of chemists, we can recognize at least three kinds of chemical properties or biological activities that depend on the symmetry properties of the molecule, the environment and the apparatus used to measure the property [5, 19]:

1. *Symmetry independent properties* (they have the same absolute value and sign for both enantiomers and are invariant to both proper and improper operations of symmetry, they are known as *scalar properties*). These properties are generally measured in an isotropic environment with symmetric apparatus, e.g. boiling points, density. These properties *do not need the use of chiral molecular descriptors* to be predicted.
2. *Molecular Symmetry dependent properties measured in isotropic environments with specific symmetric apparatus* (they have the same absolute value but opposite sign and are in general *pseudoscalar properties*). Pseudoscalar properties are those, which remain invariant to proper operations of symmetry (rotations) but change sign under an improper operation (reflections). These properties depend in their absolute magnitude on molecular symmetry and *need the use of chiral molecular symmetric descriptors* to be predicted, e.g. optical activity.
3. *Molecular symmetry dependent properties measured in non-isotropic environments* (they have different absolute values and could have or not opposite mathematical sign depending on the scale). They could be scalar or pseudoscalar with respect to the system as a whole (the molecule and the molecular environment). This specific group of properties *need the use of non-symmetric chiral descriptors*, e.g. retention

time of enantiomers in chiral chromatographic or σ -Receptors Antagonists Activities (see below).

Attempts to give quantitative meaning to molecular chirality can be dated almost as far back as van't Hoff's and LeBel's proposition to extend the structural formulas of chemistry into three-dimensional space (3D). In 1890 Guye introduced the first function designed to correlate a pseudoscalar property, i.e., optical rotation, with the molecular structure of chiroids—the first example of a chirality function in chemistry [20]. Chirality, however, is an inherent molecular property that depends only on symmetry and that is independent of its physical and chemical manifestations. It should therefore be possible to quantify chirality, i.e., to construct a chirality measure, without reference to any experimental data.

In view of the great importance of molecular chirality in chemistry, biochemistry, pharmacology, etc, much effort has been made to design theoretical methods by which enantiomeric species could be distinguished [1, 2, 21–29]. 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 [30]. Among the chiral topological indices (CTIs) published in the literature, Estrada et al. mention in a recent review about topological indices (TIs) [30]. Those derived by Pyka [31–33], Gutman and Pyka [34] have rationalized some of these indices from a mathematical point of view. The relationships between these indices and the Wiener index have been established. Moreover, Schultz et al. [28] modified a series of TIs in order to introduce information regarding the chirality of stereocenters in the molecules.

Some years ago, Buda and Mislow distinguished between two classes of measures [35]. In the first class 'the degree of chirality expresses the extent to which a chiral object differs from an achiral reference object'. In the second one 'it expresses the extent to which two enantiomorphs differ from one another'. These methods yield a single real value, usually an absolute quantity that is the same for both enantiomorphs. Recently, Benigni et al. proposed a chirality measure for molecules in a data set [26]. This measure is based on the comparison of the 3D structure for a molecule with all the others in a data set, in terms of electrostatic potential and shape indices. Moreau described a quantitative measure of the chirality of the environment of each atom [36]. However, applications of quantitative measures of chirality to the prediction of experimental observables have been quite limited.

A different idea was to incorporate R/S labels into conventional topological indices [28]. Derived chirality descriptors were correlated with biological activity by Julián-Ortiz et al. [25], Golbraikh et al. [1] and more recently by Díaz et al. [19]. One of the first approaches to in this field was introduced by de Julián-Ortiz et al. [25] in a study of the pharmacological activity of different pairs of enantiomers on the σ -receptor. Fortunately, the so-called CTIs are inexpensive in terms of computation time in comparison to grid dependent methods like CoMFA [37]. 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 statistical software [1].

In addition to CTIs, the characterization of symmetry, and specifically chiral structural features in computer-aided drug discovery (CADD), has become possible only

since the development of 3D-QSAR methods. Among these methods, special mention must be made of the use of CoMFA [37]. Evidently; the chirality in CoMFA is taken into account by default, since 3D field values of chiral isomers are different. Despite its wide popularity, CoMFA is not always applicable, especially in situation where compounds under investigation are highly flexible. Even almost these difficulties are solved by Grid (a CoMFA like last generation method) several drawbacks still remain when large data must be processed [38].

In recent times, a novel scheme to the rational *-in silico-* molecular design and to QSAR/QSPR has been introduced by our research group *TOMOCOMD* (acronym of *TO*pological *MO*lecular *COM*puter *DES*ign). It calculates several new families of 2D, 3D-Chiral (2.5) and 3D (geometric and topographic) non-stochastic and stochastic atom- and bond-based molecular descriptors based on algebraic theory and discrete mathematic. They are denominate quadratic, linear and bilinear indices and have been defined in analogy to the quadratic, linear and bilinear mathematical maps [39–44]. These approaches describe changes in the electron distribution with time throughout the molecular backbone and they have been successfully employed in the prediction of several physical, physicochemical, chemical biological and pharmacokinetical properties of organic compounds [45–61]. Applications included studies related to nucleic acid-drug interactions [62, 63] and related with structural characterization of proteins [64, 65]. Besides, these indices have been extended to considering three-dimensional features of small/medium-sized molecules based on the *trigonometric 3D-chirality correction factor approach* [66–68]. In recent works, we had obtained very promising results when stochastic and non-stochastic 3D-chiral (2.5) quadratic, linear and bilinear indices were applied to three of the most commonly used chiral data sets [66–69].

The present report is written with two objectives in mind, gather the definition of novel families of CTIs, namely 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors, and compare the results achieved with them and those obtained with other methods.

This review is structured as follows: firstly we will introduce the mathematical definition of the atom-based quadratic, linear and bilinear molecular descriptors which is necessary for understanding the nature of such descriptors and after that we will perform an exhaustive comparison between the 3D-chiral (2.5) *TOMOCOMD-CARRD* molecular descriptors and other CTIs as well as several 3D-QSAR methods. For the purpose of comparison, we are going to use three well-know chiral data sets: (i) angiotensin-converting enzyme inhibitory activity of perindoprilate's σ -stereoisomers combinatorial library, (ii) σ -receptor antagonist activities of chiral 3-(3-hydroxyphenyl)piperidines and (iii) corticosteroid-binding globulin binding affinity of the Cramer's steroid data set.

2 Theoretical scaffold about 3D-Chiral (2.5) atom-based *TOMOCOMD-CARDD* descriptors

In previous reports, we outline outstanding features concerned with the theory of 2D atom-and bond-based *TOMOCOMD-CARDD* molecular descriptors (MDs). This method codifies the molecular structure by means of mathematical quadratic, linear

and bilinear transformations [39–44]. In order to calculate these algebraic maps for a molecule, the atom-based molecular vector, \bar{x} (vector representation) and k th “non-stochastic and stochastic graph—theoretic electronic-density matrices”, \mathbf{M}^k and \mathbf{S}^k correspondingly (matrix representations), are constructed [39,40,45–58,60,62–67,70]. Such atom-adjacency relationships and chemical-information codification are applied to generate a series of atom-based *TOMOCOMD-CARDD* MDs, namely *atomic, group and atom-type as well as total quadratic, linear and 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—theoretic electronic-density matrices will be described in the next two subsections, and (2) an outline of the mathematical definition of quadratic, linear and bilinear maps and a definition of our procedures will be develop in the third subsection.

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 have been explained elsewhere in some detail [39,40,45–48,50–58,60,65–67]. The components (x) of \bar{x} are numeric values, which represent a certain standard atomic property (atomic label). That is to say, these weights correspond to different atom 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 \mathfrak{R}^5 , \mathfrak{R}^{10} , \mathfrak{R}^{15} , . . . , \mathfrak{R}^n , respectively; where n is the dimension of the real set (\mathfrak{R}^n). That is to say, \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 2 for molecular structure). This vector belongs to the product space \mathfrak{R}^9 . 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 masses, the atomic polarizabilities, and so on. In the present report, we characterized each atomic nucleus with the following parameters (weighting scheme): atomic masses (M), the van der Waals volumes (V), the atomic polarizabilities (P), atomic electronegativity in Pauling scale (E). The values of these atomic labels are shown in Table 1 [71–73].

Additionally, if we are interested to codify the chemical information by means of two different molecular vectors, for instance, $\bar{x} = [x_1, \dots, x_n]$ and $\bar{y} = [y_1, \dots, y_n]$; then different combinations of molecular vectors ($\bar{x} \neq \bar{y}$) are possible when a weighting scheme is used. For instance, in the present report, we characterized each atomic nucleus with the following parameters describe above (see Table 1): M, V, P, and E. From this weighting scheme, six (or twelve 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 two atomic properties from our weighting scheme and a minus (–) expresses the combination (pair) of two selected atom-label

Table 1 Values of the atomic weights used for TOMOCOMD-CARDD MDs

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

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 \bar{x} and \bar{y} vectors, respectively.

2.2 Background in non-stochastic and stochastic graph—theoretic electronic-density matrices

In molecular topology, molecular structure is expressed, generally, by the hydrogen-suppressed graph. That is, 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 [74–76]. 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 [74–76]. 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 [39, 40, 77].

In the earlier reports we have introduced new molecular matrices that describe changes along the time in the electronic distribution throughout the molecular back-

bone. The $n \times n$ k th non-stochastic graph—theoretic 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 [39,40,45–48,50–58,60,65–67]. The coefficients ${}^k m_{ij}$ are the elements of the k th power of $\mathbf{M}(G)$ and are defined as follows:

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

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 single, double or triple bonds, 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 (3-mercapto-pyridine-4-carbaldehyde). Table 2 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—theoretic electronic-density matrices are also given in this Table.

As can be seen, \mathbf{M}^k are graph—theoretic electronic-structure models, like an “extended Hückel theory (EHT) model”. The \mathbf{M}^1 matrix considers all valence-bond electrons (σ - and π -networks) in one step and its power ($k = 0, 1, 2, 3, \dots$) can be considered as interacting—electron chemical—network models in k step. The complete model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas [78].

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). 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 . In this sense, the electron in an arbitrary atom i can move (step-by-step) to other atoms at different discrete time periods t_k ($k = 0, 1, 2, 3, \dots$) throughout the chemical-bonding network.

On the other hand, the k th stochastic graph—theoretic electronic-density matrix of G , \mathbf{S}^k , can be directly obtained 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:

Table 2 (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—theoretic electronic-density matrices of G, respectively

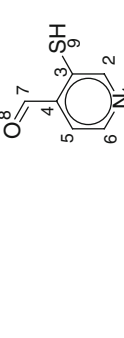

A)	molecular structure	molecular pseudograph (H-atoms-suppressed pseudograph) ^a
		
B)		
	k^{th} non-stochastic graph—theoretic electronic-density matrices, \mathbf{M}^k ($k = 0-3$)	
	zero order ($k = 0$)	first order ($k = 1$)
	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}$
	second order ($k = 2$)	third order ($k = 3$)
	$\begin{bmatrix} 3 & 2 & 1 & 0 & 1 & 2 & 0 & 0 & 0 \\ 2 & 3 & 2 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 2 & 4 & 2 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 7 & 6 & 3 & 2 & 3 & 6 & 0 & 0 & 1 \\ 6 & 7 & 7 & 3 & 2 & 3 & 1 & 0 & 2 \\ 3 & 7 & 9 & 8 & 3 & 2 & 2 & 2 & 4 \\ 2 & 3 & 8 & 9 & 7 & 3 & 8 & 2 & 2 \\ 3 & 2 & 3 & 7 & 7 & 6 & 2 & 2 & 1 \\ 6 & 3 & 2 & 3 & 6 & 7 & 1 & 0 & 0 \\ 0 & 1 & 2 & 8 & 2 & 1 & 1 & 10 & 1 \\ 0 & 0 & 2 & 2 & 2 & 0 & 10 & 0 & 0 \\ 1 & 2 & 4 & 2 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$
	stochastic graph—theoretic electronic-density matrices, \mathbf{S}^k ($k = 0-3$) ^{b,c}	
	first order ($k = 1$)	second order ($k = 2$)
	$\begin{bmatrix} 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0.3333 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0.6666 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.3333 & 0.2222 & 0.1111 & 0 & 0.1111 & 0.2222 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0 & 0.1 & 0 & 0 & 0.1 \\ 0.0833 & 0.166 & 0.3333 & 0.1666 & 0.0833 & 0 & 0.0833 & 0 & 0.0833 \\ 0 & 0.0714 & 0.1429 & 0.2857 & 0.1429 & 0.0714 & 0.0714 & 0.1429 & 0.0714 \\ 0.1 & 0 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0 & 0 \\ 0.2222 & 0.1111 & 0 & 0.1111 & 0.2222 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 0.125 & 0 & 0.625 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0 & 0.6666 & 0 \\ 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}$
	first order ($k = 1$)	second order ($k = 2$)
	$\begin{bmatrix} 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0.3333 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0.6666 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.25 & 0.2142 & 0.1071 & 0.0714 & 0.1071 & 0.2143 & 0 & 0 & 0.0357 \\ 0.1935 & 0.2258 & 0.2258 & 0.2258 & 0.0967 & 0.0645 & 0.0967 & 0.0323 & 0 & 0.0645 \\ 0.075 & 0.175 & 0.225 & 0.2 & 0.075 & 0.05 & 0.05 & 0.05 & 0.1 & 0 \\ 0.0455 & 0.0682 & 0.1818 & 0.2045 & 0.1591 & 0.0682 & 0.1818 & 0.0455 & 0.0455 & 0.0455 \\ 0.0909 & 0.0606 & 0.0909 & 0.2121 & 0.2121 & 0.1818 & 0.0606 & 0.0606 & 0.0303 & 0.0303 \\ 0.2143 & 0.1071 & 0.0714 & 0.1971 & 0.2143 & 0.25 & 0.0357 & 0 & 0 & 0 \\ 0 & 0.0385 & 0.0769 & 0.3076 & 0.0769 & 0.0385 & 0.0385 & 0.3846 & 0.0385 & 0.0385 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 0 & 0.625 & 0 & 0 & 0 \\ 0.0833 & 0.1666 & 0.333 & 0.1666 & 0.0833 & 0 & 0.0833 & 0 & 0.0833 & 0 \end{bmatrix}$

Table 2 continued

- ^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 ($k = 0, 1, 2, 3, \dots$) can be considered as an interacting–electron chemical–network model in k step
- ^b The zero power ($k = 0$) of the stochastic graph—theoretic 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 k th matrices \mathbf{S}^k (s_{ij}^k) have been approximated

$${}^k s_{ij} = \frac{{}^k m_{ij}}{{}^k \text{SUM}_i} = \frac{{}^k m_{ij}}{{}^k \delta_i} \quad (2)$$

where, ${}^k m_{ij}$ 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 , ${}^k \delta_i$. 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**” [70]. The k th 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 k th element s_{ij} takes into consideration the molecular topology in k step throughout the chemical-bonding (σ - and π -) network. In this sense, 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 electron returns to the original atom 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 sp² (or sp³) C atom in t_2 [$p(\text{C}_{sp}) > p(\text{C}_{sp}^2) > p(\text{C}_{sp}^{\text{arom}}) > p(\text{C}_{sp}^3)$] (see Table 2 for more details). This is a logical result if the electronegativity scale of these hybrid states is taken into account.

2.3 Calculation of atom-based 2D TOMOCOMD-CARDD MDs

2.3.1 Quadratic and bilinear indices for atoms, group, atom-types and the whole molecule

If a molecule consists of n atoms (*vector of* \mathcal{N}^n), then the k th total (whole) quadratic and bilinear indices are calculated as quadratic and bilinear forms on \mathcal{N}^n , respectively, in canonical basis set. Specifically, the k th non-stochastic and stochastic atom-based quadratic [bilinear] indices for a molecule, $q_k(\bar{x})[b_k(\bar{x}, \bar{y})]$ and ${}^s q_k(\bar{x})[{}^s b_k(\bar{x}, \bar{y})]$, are computed from these k th non-stochastic and stochastic graph—theoretic electronic-density matrices, \mathbf{M}^k and \mathbf{S}^k as shown in Eqs. 3 [3a] and 4 [4a], correspondingly [39,40,45–58,60,62–67,70]:

$$q_k(\bar{x}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ij} x^i x^j = [\mathbf{X}]^t \mathbf{M}^k [\mathbf{X}] \quad (3)$$

$$b_k(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ij} x^i y^j = [\mathbf{X}]^t \mathbf{M}^k [\mathbf{Y}] \quad (3a)$$

$${}^s q_k(\bar{x}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ij} x^i x^j = [\mathbf{X}]^t \mathbf{S}^k [\mathbf{X}] \quad (4)$$

$${}^s b_k(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ij} x^i y^j = [\mathbf{X}]^t \mathbf{S}^k [\mathbf{Y}] \quad (4a)$$

where, n is the number of atoms (atomic nuclei) of the molecule, and x^1, \dots, x^n and y^1, \dots, y^n are the coordinates or components of the molecular vectors \bar{x} and \bar{y} in a system of canonical ('natural') basis vectors of \mathfrak{N}^n . In this basis system, the coordinates $[(x^1, \dots, x^n)$ and $(y^1, \dots, y^n)]$ of any molecular vectors (\bar{x} and \bar{y}) coincide with the components of those vectors $[(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)]$ [79,80]. For that reason, those coordinates can be considered as weights (atomic labels) of the vertices of the molecular pseudograph. The coefficients ${}^k m_{ij}$ and ${}^k s_{ij}$ are the elements of the k^{th} power of the matrix $\mathbf{M}(\mathbf{G})$ and $\mathbf{S}(\mathbf{G})$, correspondingly, of the molecular pseudograph.

The defined Eqs. 3 [3a] and 4 [4a] for $q_k(\bar{x})[b_k(\bar{x}, \bar{y})]$ and ${}^s q_k(\bar{x})[{}^s b_k(\bar{x}, \bar{y})]$ may also be written as the single matrix equation, where $[\mathbf{X}]$ ($[\mathbf{Y}]$) is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{x} (\bar{y}) in the canonical basis of \mathfrak{N}^n , and $[\mathbf{X}]^t$ (a $1 \times n$ matrix) is the transpose of $[\mathbf{X}]$. Here, \mathbf{M}^k and \mathbf{S}^k denote the matrices of quadratic maps with respect to the natural basis set.

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, \mathbf{M} and \mathbf{V} are used as atomic weights to compute these MDs, two different sets of stochastic bilinear indices, ${}^{\mathbf{M}-\mathbf{V}s} b_k^{\mathbf{H}}(x, y)$ and ${}^{\mathbf{V}-\mathbf{M}s} b_k^{\mathbf{H}}(x, y)$ [because $\bar{x}_{\mathbf{M}} - \bar{y}_{\mathbf{V}} \neq \bar{x}_{\mathbf{V}} - \bar{y}_{\mathbf{M}}$] can be obtained and only one group of non-stochastic bilinear indices (${}^{\mathbf{M}-\mathbf{V}s} b_k^{\mathbf{H}}(x, y) = {}^{\mathbf{V}-\mathbf{M}s} b_k^{\mathbf{H}}(x, y)$ because in this case $\bar{x}_{\mathbf{M}} - \bar{y}_{\mathbf{V}} = \bar{x}_{\mathbf{V}} - \bar{y}_{\mathbf{M}}$) can be calculated.

In the last decade, Randić [81] proposed a list of desirable attributes for a MD. Therefore, this list can be considered as a methodological guide for the development of new 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 related more 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 quadratic and 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 quadratic [bilinear] indices*, $q_{kL}(\bar{x})[b_{kL}(\bar{x}, \bar{y})]$ and ${}^s q_{kL}(\bar{x}) [{}^s b_{kL}(\bar{x}, \bar{y})]$, respectively. The definition of these descriptors is as follows:

$$q_{kL}(\bar{x}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ijL} x^i x^j = [\mathbf{X}]^t \mathbf{M}_L^k [\mathbf{X}] \quad (5)$$

$$b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k m_{ijL} x^i y^j = [\mathbf{X}]^t \mathbf{M}_L^k [\mathbf{Y}] \quad (5a)$$

$${}^s q_{kL}(\bar{x}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ijL} x^i x^j = [\mathbf{X}]^t \mathbf{S}_L^k [\mathbf{X}] \quad (6)$$

$${}^s b_{kL}(\bar{x}, \bar{y}) = \sum_{i=1}^n \sum_{j=1}^n {}^k s_{ijL} x^i y^j = [\mathbf{X}]^t \mathbf{S}_L^k [\mathbf{Y}] \quad (6a)$$

where ${}^k m_{ijL} [{}^k s_{ijL}]$ is the k th element of the row “ i ” and column “ j ” of the local matrix $\mathbf{M}_L^k [\mathbf{S}_L^k]$. 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 $\mathbf{M}_L^k [\mathbf{S}_L^k]$ with elements ${}^k m_{ijL} [{}^k s_{ijL}]$ is defined as follows:

$$\begin{aligned} {}^k m_{ijL} [{}^k s_{ijL}] &= {}^k m_{ij} [{}^k s_{ijL}] \text{ if both } v_i \text{ and } v_j \text{ are atomic nuclei contained within} \\ &\quad \text{the molecular fragment} \\ &= \frac{1}{2} {}^k m_{ij} [{}^k s_{ijL}] \text{ if } v_i \text{ or } v_j \text{ is an atomic nucleus contained within} \\ &\quad \text{the molecular fragment but not both} \\ &= 0 \text{ otherwise} \end{aligned} \quad (7)$$

These local analogues can also be expressed in matrix form for each molecular vector $\bar{x} \in \mathfrak{R}^n$ and $\bar{y} \in \mathfrak{R}^n$. It should be remarked that the above scheme follows the spirit of a Mulliken population analysis [82]. It should be point out also that for every partitioning of a molecule into Z molecular fragments there will be Z local molecular fragment matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrix $\mathbf{M}^k [\mathbf{S}^k]$ can be partitioned into Z local matrices $\mathbf{M}_L^k [\mathbf{S}_L^k]$, $L = 1, \dots, Z$, and the k th power of matrix $\mathbf{M} [\mathbf{S}]$ is exactly the sum of the k th power of the local Z matrices:

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

$$\mathbf{S}^k = \sum_{L=1}^Z \mathbf{S}_L^k \quad (9)$$

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

$${}^k m_{ij} = \sum_{L=1}^Z {}^k m_{ijL} \quad (10)$$

$${}^k s_{ij} = \sum_{L=1}^Z {}^k s_{ijL} \quad (11)$$

and consequently, the total non-stochastic and stochastic quadratic [bilinear] indices are the sum of the non-stochastic and stochastic quadratic [bilinear] indices, respectively, of the Z molecular fragments:

$$q_k(\bar{x}) = \sum_{L=1}^Z q_{kL}(\bar{x}) \quad (12)$$

$$b_k(\bar{x}, \bar{y}) = \sum_{L=1}^Z b_{kL}(\bar{x}, \bar{y}) \quad (12a)$$

$${}^s q_k(\bar{x}) = \sum_{L=1}^Z {}^s q_{kL}(\bar{x}) \quad (13)$$

$${}^s b_k(\bar{x}, \bar{y}) = \sum_{L=1}^Z {}^s b_{kL}(\bar{x}, \bar{y}) \quad (13a)$$

Atomic, atom-type and group quadratic [bilinear] fingerprints are specific cases of local quadratic [bilinear] indices. First, notice that atomic quadratic [bilinear] indices, $q_k(\bar{x}_i)$ and ${}^s q_k(\bar{x}_i)$ [$b_{kL}(\bar{x}_i, \bar{y}_i)$ 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 quadratic and bilinear indices values for the common scaffold atoms can be directly used 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. As it can be seen, the k th total quadratic and bilinear indices (both non-stochastic and stochastic) are calculated by summing the atomic quadratic and bilinear indices, respectively, of all atoms in the molecule.

In addition, the atom-type quadratic and bilinear indices can also be calculated as local MDs. In the same way as atom-type E-state values [83], 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. That is, this approach provides the basis for application to a wider range of problems to which

the atomic quadratic indices formalism is applicable without the need for superposition. For this reason the present method represents a significant advantage over traditional QSAR methods. The atom-type quadratic and bilinear descriptors are calculated by adding the k th atomic quadratic and bilinear indices, correspondingly, for all atoms of the same type in the molecule. This atom type index lends itself to use in a group additive-type scheme in which an index appears for each atom type in the molecule.

In the atom-type quadratic and 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 [83]. That is to say, each atom in the molecule is categorized according to a valence-state classification scheme including the number of attached H-atoms [83]. The atom-type descriptors combine three important aspects of structural information: (1) Collective electron and topologic 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 (Hbonding) 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 quadratic [bilinear] indices are the sum of the individual atom-level quadratic [bilinear] indices for a particular group of atoms. For all data set structures, the k th group-based quadratic and bilinear indices provide important information for QSAR/QSPR studies.

2.3.2 Linear indices for atoms, group, atom-types and the whole molecule

If a molecule consists of n atoms (*vector of* \mathfrak{R}^n), then the k th atomic linear indices for atom i in a molecule, are calculated as linear maps on \mathfrak{R}^n (endomorphism on \mathfrak{R}^n) in canonical basis set. Specifically, the k th non-stochastic and stochastic atomic linear indices, $f_k(\bar{x}_i)$ and ${}^s f_k(\bar{x}_i)$, are computed from these k th non-stochastic and stochastic graph—theoretic electronic-density matrices, \mathbf{M}^k and \mathbf{S}^k , as shown in Eqs. 3 and 4, respectively:

$$f_k(\bar{x}_i) = \sum_{j=1}^n {}^k m_{ij} x^j = [X']^k = \mathbf{M}^k [X] \quad (14)$$

$${}^s f_k(\bar{x}_i) = \sum_{j=1}^n {}^k s_{ij} x^j = [XS']^k = \mathbf{S}^k [X] \quad (15)$$

where n is the number of atoms of the molecule and x^j are the coordinates of the atom-based molecular vector (\bar{x}) in the so-called canonical ('natural') basis. The coefficients ${}^k m_{ij}$ and ${}^k s_{ij}$ are the elements of the k th power of the matrix $\mathbf{M}(G)$ and $\mathbf{S}(G)$, correspondingly, of the molecular pseudograph. The defining equation (14) and (15) for

$f_k(\bar{x}_i)$ and ${}^s f_k(\bar{x}_i)$, respectively, may be also written as the single matrix equation, where $[X]$ is a column vector (an $n \times 1$ matrix) of the coordinates of \bar{x} in the canonical basis of \mathfrak{N}^n . Here, M^k and S^k denote the matrices of linear maps with respect to the natural basis set.

It should be remarked that both atom linear indices are defined as a linear transformation $f_k(\bar{x}_i)$ on molecular vector space \mathfrak{N}^n . This map is a correspondence that assigns a vector $f(x)$ to every vector \bar{x} in \mathfrak{N}^n in such a way that:

$$f(\lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2) = \lambda_1 f(\bar{x}_1) + \lambda_2 f(\bar{x}_2) \quad (16)$$

for any scalar λ_1, λ_2 and any vector \bar{x}_1, \bar{x}_2 in \mathfrak{N}^n .

Total (whole-molecule) atom-based non-stochastic and stochastic linear indices, $f_k(\bar{x})$ and ${}^s f_k(\bar{x})$, are calculated from local (atomic) linear indices as shown in Eqs. 17 and 18, correspondingly:

$$f_k(\bar{x}) = \sum_{i=1}^n f_k(\bar{x}_i) = [u]^t [X']^k = [u]^t M^k [X] \quad (17)$$

$${}^s f_k(\bar{x}) = \sum_{i=1}^n {}^s f_k(\bar{x}_i) = [u]^t [XS']^k = [u]^t S^k [X] \quad (18)$$

where n is the number of atoms, and $f_k(\bar{x}_i)$ and ${}^s f_k(\bar{x}_i)$ are the non-stochastic and stochastic atomic linear indices obtained by Eqs. 14 and 15, respectively. Then, both total linear form, $f_k(\bar{x})$ and ${}^s f_k(\bar{x})$, can also be written in matrix form for each molecular vector $\bar{x} \in \mathfrak{N}^n$, where $[u]^t$ is an n -dimensional unitary row vector. As it can be seen, the k th total linear indices (both non-stochastic and stochastic) are calculated by summing the local (atomic) linear indices of all atoms in the molecule.

Finally, in addition to total and atomic linear indices computed for each atom in the molecule, a local-fragment formalism can also be developed. In this sense, *group* and *atom-type* linear fingerprints are specific cases of local atom-based linear indices.

The atom-type linear descriptors (as well as quadratic and bilinear indices) are calculated by adding the k th atomic linear indices for all atoms of the same type in the molecule. The group-level linear indices are the sum of the individual atom-level linear indices for a particular group of atoms. Consequently, this atom- and group-type index lends itself to use in a group additive-type scheme in which an index appears for each atom type in the molecule. Here, if a molecule is partitioned into Z molecular fragments, the total non-stochastic [or stochastic] linear indices can be partitioned into Z local non-stochastic [or stochastic] linear indices $f_{kL}(\bar{x})$ [or ${}^s f_{kL}(\bar{x})$], $L = 1, \dots, Z$. That is to say, the total (both non-stochastic and stochastic) linear indices of order k can be expressed as the sum of the local linear indices of the Z fragments of the same order:

$$f_k(\bar{x}) = \sum_{L=1}^Z f_{kL}(\bar{x}) \quad (19)$$

$${}^s f_k(\bar{x}) = \sum_{L=1}^Z {}^s f_{kL}(\bar{x}) \quad (20)$$

2.4 3D-chiral extended (2.5) non-stochastic and stochastic atom-based TOMOCOMD-CARDD MDs for atoms, atom-types and group as well as for whole molecule

The total and local, non-stochastic and stochastic quadratic, bilinear, and 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 \bar{x} [66–69]. In these sense, a chirality molecular vector is obtained (${}^* \bar{x}$), where the components of \bar{x} (for instance, Pauling electronegativity (x_A) of the atom A) are substituted by the following term [$x_A + \sin((\omega_A + 4\Delta)\pi/2)$].

The *trigonometric 3D-chirality correction factor* use a dummy variable, ω_A and an integer parameter, Δ [66–69]:

$$\begin{aligned} \omega_A &= 1 \text{ and } \Delta \text{ is an odd number when } A \text{ has R (rectus), E (entgegen), or } a \text{ (axial)} \\ &\text{notation according to Cahn-Ingold-Prelog rules} \\ &= 0 \text{ and } \Delta \text{ is an even number, if } A \text{ does not have 3D specific enviroment} \\ &= -1 \text{ and } \Delta \text{ is an odd number when } A \text{ has S (sinister), Z (zusammen),} \\ &\text{or } e \text{ (ecuatorial) notation according to Cahn-Ingold-Prelog rules} \end{aligned} \quad (21)$$

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 [84]. This correction has principally a mathematical means and must not be source of any misunderstanding.

A severe limitation of the GBT [1] approach is the existence of different chirality corrections and we had great difficulty in selecting one of these. 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 TIs (GBT-like 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). In this Table is 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, 19, 25].

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	1	1	1	1	1	1	1
$\omega_{non-chiral} = 0$		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 (2.5) descriptor reduces to simple (2D) non-stochastic and stochastic quadratic, bilinear and linear indices ones for molecules without specific 3D characteristics because $\sin(0 + 4\Delta)\pi/2 = 0$, being Δ zero or any even number. That is, when all the atoms in the molecule are not chiral, the TOMOCOMD-CARDD MDs or any GBT-like chiral TI do not change upon the introduction of this factor. This means that, for example ${}^*\bar{x} = \bar{x}$ and thus, ${}^*q_k(\bar{x}) = q_k(\bar{x})$, ${}^*b_k(\bar{x}) = b_k(\bar{x})$ and ${}^*f_k(\bar{x}) = f_k(\bar{x})$.

3 Computational strategies

All computations were carried out on a PC Pentium-4 3.2 GHz. The TOMOCOMD program for Windows package developed in our laboratory was used for computing the molecular descriptors for the dataset of compounds. This software is an interactive program for molecular design and bioinformatic research [85]. 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). The modules are named CARDD (Computed-Aided ‘Rational’ Drug Design), CAMPS (Computed-Aided Modeling in Protein Science), CANAR (Computed-Aided Nucleic Acid Research) and CABPD (Computed-Aided Bio-Polymers Docking). In the present report, we outline salient features concerned with only one of these subprograms, CARDD and with the calculation of non-stochastic and stochastic 2D atom-based bilinear indices.

The main steps for the application of the present method in QSAR/QSPR and drug design can be summarized briefly in the following algorithm: (1) Draw the molecular structure for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol 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 [72, 73, 86], 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 1 [71–73, 86]; (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 \mathbf{P} and the bilinear fingerprints having, for instance, the following appearance,

$$\mathbf{P} = a_0 \mathbf{b}_0(\bar{x}, \bar{y}) + a_1 \mathbf{b}_1(\bar{x}, \bar{y}) + a_2 \mathbf{b}_2(\bar{x}, \bar{y}) + \cdots + a_k \mathbf{b}_k(\bar{x}, \bar{y}) + c \quad (22)$$

where \mathbf{P} is the measured property, $\mathbf{b}_k(\bar{x}, \bar{y})$ are the k th non-stochastic total bilinear indices, and the 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.

4 Qsar applications to central chirality codification: Comparison with other theoretical approaches

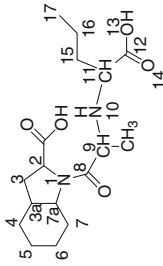
With the objective of assess the efficacy of 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors, we have tested their ability to predict pharmacological properties in several groups of compounds with a known stereochemical influence. We select these data sets because they have been used repeatedly in several QSAR studies in recent years. Now we are going to discuss the use of the 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors 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, an angiotensin-converting enzyme (ACE) inhibitors [87], was used to test the applicability of our method. This data set was previously used by Diaz et al. [19,87] and furthermore by us in some works [66,67]. ACE acts in plasma and blood vessels, removing the C-terminal dipeptide of undecapeptide Angiotensin I to produce the potent blood vessel constricting octapeptide Angiotensin II. In addition, ACE inactivates the hypotensive nonapeptide Bradykinin. For these reasons, ACE is the biological target of many important antihypertensive drugs called ACE inhibitors (ACEIs) [87]. In this study, active is taken to a mean a compound that has an IC_{50} value no higher than 110 nm.

In the reviewed works a Linear Discrimination Analysis (LDA) was used to develop a simple linear QSAR model to fit the classification functions in order to discriminate between two degrees of ACE inhibitory activity. To codify the biological activity a dummy variable (ACEiactv) was used. This variable indicates the presence of either a very active compound (ACEiactv = 1) or a non-active compound (ACEiactv = -1). In Table 4 we give the basic structure of perindoprilate stereoisomer and their classification in the training and prediction set when different approaches were used.

Table 4 Basic structure and chirality notation of active and non-active perindoprilate stereoisomers



No	Comp. ^a	Class ^b	IC ₅₀ ^c	MARCH-INSIDE MDs [19]	non-Stoch. Quadratic indices [67]	non-Stoch. Linear indices [66]	Stoch. Linear indices [66]	non-Stoch. Bilinear indices [68]	Stoch. Bilinear indices [68]
<i>Active compounds</i>									
1	SSRSS ^d	+	1.1	+	+	+	+	+	+
2	SRSSS ^d	+	1.2	+	+	+	+	+	+
3	SSSSS	+	1.5	+	+	+	+	+	+
4	SRRSS ^d	+	3.3	+	+	+	+	+	+
5	SSSSR	+	12.2	+	+	+	+	+	+
6	SSRSR	+	29.4	+	+	+	+	+	+
7	SRRSR	+	39.8	-	+	+	+	+	+
8	SRSSR	+	54	+	+	+	+	+	+
9	RRSSS	+	108	+	-	+	-	-	-
<i>Non-active compounds</i>									
10	SSSRS	-	1.1 × 10 ³	-	-	-	-	-	-
11	RSSSS	-	1.9 × 10 ³	-	-	-	-	-	-
12	SSRRR ^d	-	2.6 × 10 ³	+	-	-	-	-	-
13	RRSSR	-	5.5 × 10 ³	-	-	-	-	-	-
14	SSRRS	-	7.1 × 10 ³	+	-	-	-	-	-
15	RRSRS	-	7.8 × 10 ³	-	-	-	-	-	-
16	RSRRR ^d	-	23 × 10 ³	-	-	-	-	-	-
17	SRRRR	-	33 × 10 ³	-	-	-	-	-	-
18	RSSSR	-	36 × 10 ³	-	-	-	-	-	-

Table 4 continued

No	Comp. ^a	Class ^b	IC ₅₀ ^c	MARCH-INSIDE MDs [19]	non-Stoch. Quadratic indices [67]	non-Stoch. Linear indices [66]	Stoch. Linear indices [66]	non-Stoch. Bilinear indices [68]	Stoch. Bilinear indices [68]
19	<i>RSRSR</i>	–	47×10^3	–	–	–	–	–	–
20	<i>RSRSS^d</i>	–	60×10^3	–	–	–	–	–	–
21	<i>RRRRR</i>	–	10^5	–	–	–	–	–	–
22	<i>SRRRS</i>	–	10^5	–	–	–	–	–	–
23	<i>RRRSS</i>	–	10^5	–	–	–	–	–	–
24	<i>SRSRR^d</i>	–	10^5	–	–	–	–	–	–
25	<i>RRRRS</i>	–	10^5	–	–	–	–	–	–
26	<i>RRRRR</i>	–	10^5	–	–	–	–	–	–
27	<i>SSRRR</i>	–	10^5	–	–	–	–	–	–
28	<i>RSSRS^d</i>	–	10^5	–	–	–	–	–	–
29	<i>RRRSR</i>	–	10^5	–	–	–	–	–	–
30	<i>RSSRR</i>	–	10^5	–	–	–	–	–	–
31	<i>RSRRS</i>	–	10^5	–	–	–	–	–	–
32	<i>SRSRS^d</i>	–	10^5	–	–	+	–	–	–

^a Notation of the chiral centres in each perindoprilate derivative in the following order C₂, C_{3a}, C_{7a}, C₉, C₁₁

^b Classification according to the value of the IC₅₀

^c Values of the IC₅₀, of the compound, for ACE in nM taken from previous works [19,25]

^d Compounds used in the Test set

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

Index	n	λ	D^2	Accuracy (training) (%)	Accuracy (training) (%)	F
Stochastic bilinear indices [69]	2	0.393	7.30	95.65	100.00	15.42
Non-stochastic bilinear indices [69]	2	0.398	7.164	95.65	100.00	15.13
Stochastic linear indices [66]	2	0.399	7.789	95.65	100.00	15.02
Non-stochastic quadratic indices [67]	2	0.42	7.12	95.65	100.00	13.73
Non-stochastic linear indices [66]	2	0.398	7.82	100.00	88.88	15.08
MARCH-INSIDE molecular descriptors [19]	3	0.38	8.43	91.30	88.88	10.30

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

The statistical analyses were carried out with the STATISTICA software [88]. The quality of the reported discriminant functions was determined examining the statistics parameter of multivariable comparison (Wilk's λ statistic, the square of Mahalanobis distance, Fisher ratio F , the corresponding p -level [$p(F)$] as well as the percentage of good classification, the proportion between the cases and variables in the equation) and the use of an external validation set. Table 5 show the classification of 32 perindopirilate's stereoisomers and the statistical parameters of the QSAR models obtained using different molecular descriptors (MDs).

As can be seen, all the obtained models for the prediction between the two degrees of ACE inhibitory activity involved two MDs and one case (when MARCH-INSIDE MDs were used) needed three parameters. The accuracy values were over 91% in all the models for the training set and in one of them (QSAR model obtained with non-stochastic atom-based linear indices) it reaches 100%. An accuracy of the 95.65% (22/23) was obtained in four of the six models for the training set. In addition, the values of the Wilk's λ statistic were good for the six models. The Wilks' λ statistical helpful to value the total discrimination and can take values between zero (perfect discrimination) and one (no discrimination), the best values (0.38 and 0.393) were obtained when MARCH-INSIDE MDs and atom-based stochastic bilinear indices were used respectively. Notice that, the second one use one variable less than the other model previously mentioned.

Validation of the models is the other major bottleneck in QSAR [89,90]. One of the most popular validation criteria is internal cross-validation (leave-one-out, leave- n -out, leave-25%(group)-out and so on). Nevertheless, there can exist a lack of correlation between the good results in internal cross-validation and the high predictive ability of QSAR models [89,90]. Thus, the good high behavior in internal cross-val-

idation appears to be the necessary but not the sufficient condition for the models to have a high predictive power. In this sense, Golbraikh and Tropsha emphasize that the predictive ability of a QSAR model can only be estimated using an external test set (external validation) of compounds that was not used for building the model and formulated a set of criteria for evaluation of predictive ability of QSAR model [90]. In this sense, four models had an accuracy of the 100% for the test set. We can say that, the best of all is the model obtained with 3D-chiral (2.5) stochastic bilinear indices because it has a high value of accuracy for both, training (95.65%) and test (100%) set. This model involves only two parameters and has one of the best values of the Wilks' λ statistic. In Table 5 this comparison can be made easier.

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

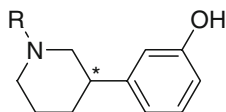
After that, a short data set of seven pairs of chiral *N*-alkylated 3-(3-hydroxyphenyl) piperidines that bind to σ -receptors, are also selected as illustrative example of the 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors application. This data set was introduced in QSAR studies by de Julian-Ortiz et al. [25] in 1998, and after that has been repeatedly used by some authors [19, 66–68] in recent years, to validate new CTIs. The σ -receptors mediate severe side effects induced by various dopamine antagonists [25].

3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors are non-symmetric and reduce to classical descriptors when symmetry is not codified. Besides, González-Díaz et al. conclude that σ receptor antagonist activities is not a pseudoscalar property [19] and we can expect at least a good correlation with 3D-chiral (2.5) *TOMOCOMD-CARDD* descriptors. Table 6 show the experimental and predicted values of Log IC₅₀ for these compounds, previously reported in the reviewed papers [19, 25].

A linear multiple regression (LMR) analysis was used to obtain quantitative models that related the CTIs and σ -receptors antagonist activities [19, 25, 66–68]. The search for the best model can be processed in terms of the highest regression coefficient (R) or F-test equations (Fisher-ratio's *p*-level [*p*(F)]), and the lowest standard deviation equations (*s*). The quality of the models was determined examining the regression's statistic parameters and of the cross-validation procedures [90, 91]. In this sense, the quality of models was determined by examining the determination coefficients (also know as squared regression coefficient; R²) and the leave-one-out (LOO) press statistics (*q*², *s*_{cv}) [89, 90]. The results of these works are summarized in the Table 7, where a comparison between all these approaches can be easily carried out.

All the obtained QSAR models use two variables except one of them who involve three descriptors. The best results for modeling σ -receptor antagonist activities were obtained with the non-stochastic and stochastic atom-based bilinear indices. These models explain more than the 95 % and 96% of the experimental values of log IC₅₀, respectively. They also showed the lowest values of standard deviation 0.238 and 0.219, correspondingly. These results are better than others reported previously in the literature, for the same data set, using different CTIs.

Predictability and stability (robustness) of the obtained models to data variation was carried out here by means of LOO cross-validation. In this sense, the models with

Table 6 Results of multivariate regression analysis of the log IC₅₀ of a group of *n*-alkylated 3-(3-hydroxyphenyl)piperidines for the σ -receptor

Compound (Alkyl group) ^a	Log IC ₅₀ (σ -receptor)											
	Obs. ^b	Cal. ^c	Res. ^d	Cal. ^e	Res. ^d	Cal. ^f	Res. ^d	Cal. ^g	Res. ^d	Cal. ^h	Res. ^d	
<i>(R)</i> -3-HPP												
H	-0.66	-0.43	-0.23	-0.54	-0.12	-0.48	-0.18	-0.50	-0.16	-0.71	0.05	
CH ₃	0.43	0.12	0.31	0.13	0.30	0.28	0.15	0.06	0.37	0.35	0.08	
C ₂ H ₅	0.95	0.72	0.23	0.72	0.23	0.70	0.25	0.67	0.28	0.75	0.20	
<i>n</i> -C ₃ H ₇	1.52	1.36	0.16	1.32	0.20	1.45	0.07	1.34	0.18	1.55	-0.03	
<i>i</i> -C ₃ H ₇	0.61	1.27	-0.66	1.30	-0.69	0.84	-0.23	1.10	-0.49	1.11	-0.50	
<i>n</i> -C ₄ H ₉	2.05	2.00	0.05	1.93	0.12	1.89	0.16	2.04	0.01	1.83	0.22	
2-Phenylethyl	2.10	2.22	-0.12	2.22	-0.12	2.41	-0.31	2.20	-0.10	2.21	-0.11	
<i>(S)</i> -3-HPP												
H	-1.19	-1.06	-0.13	-1.09	-0.10	-0.80	-0.39	-1.02	-0.17	-1.25	0.06	
CH ₃	-0.28	-0.48	0.20	-0.42	0.14	-0.56	0.28	-0.44	0.16	-0.19	-0.09	
C ₂ H ₅	-0.01	0.13	-0.14	0.17	-0.18	0.19	-0.20	0.17	-0.18	0.19	-0.20	
<i>n</i> -C ₃ H ₇	0.81	0.77	0.04	0.77	0.04	0.57	0.24	0.85	-0.04	0.98	-0.17	
<i>i</i> -C ₃ H ₇	0.68	0.68	0.00	0.75	-0.07	0.62	0.06	0.62	0.06	0.53	0.15	
<i>n</i> -C ₄ H ₉	1.51	1.40	0.11	1.37	0.14	1.18	0.33	1.54	-0.03	1.27	0.24	
2-Phenylethyl	1.80	1.62	0.18	1.67	0.13	2.03	-0.23	1.70	0.10	1.70	0.10	

^a Alkyl (R) group at nitrogen ring^b Observed values of the Log IC₅₀ for the σ -receptor taken from the literature [19,25,67]^c Values calculated using non-stochastic quadratic indices [67]^d Residual, defined as [Log IC₅₀(σ)Obs - Log IC₅₀(σ)Cal]^e Values calculated using non-stochastic linear indices [66]^f Values calculated using stochastic linear indices [66]^g Values calculated using non-stochastic bilinear indices [68]^h Values calculated using stochastic bilinear indices [68]Abbreviations: HPP, *N*-alkylated 3-Hydroxyphenyl piperidines**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	<i>N</i>	<i>n</i>	<i>R</i> ²	<i>s</i>	<i>q</i> ²	<i>s</i> _{cv}	<i>F</i>
Stochastic bilinear indices [68]	14	2	0.961	0.219	0.946	0.235	134.02
Non-stochastic bilinear indices [68]	14	2	0.953	0.238	0.935	0.259	111.93
Stochastic linear indices [66]	14	2	0.941	0.267	0.90	0.319	87.93
Non-stochastic quadratic indices [67]	14	2	0.940	0.270	0.912	0.289	85.82
Non-stochastic linear indices [66]	14	2	0.939	0.271	0.909	0.305	84.87
Chiral TIs [25]	14	3	0.931	0.301	*	*	45.70
MARCH-INSIDE molecular descriptors [19]	14	2	0.922	0.295	*	0.32	71.17

* Values are not reported in the literature

reported cross-validation regression coefficient (q^2) present high predictive power; these values of q^2 are between 0.90 and 0.946. The values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models [90–92]. Unfortunately, the authors of previous works, Diaz et al. [19] and Julian de Ortiz et al. [25]

do not report the result of the cross-validation. Considering all these statistical criteria we can conclude that, the obtained model with stochastic bilinear indices is the best QSAR model to describe the property studied in this section.

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

Finally, in order to validate even more the 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs in QSAR studies, we select a molecular set that are well-known to QSAR researchers, the so-called Cramer's steroid database. This data set was introduced by Cramer et al. [37] in 1988 using Comparative Molecular Field Analysis (CoMFA) methodology and since then has become a benchmark for the assessment of novel QSAR methods [93,94]. 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 MDs [95]. Even though this data set is not the ideal 3D benchmark data set [95] it was used for the sake of comparability [96]. We use this molecular set, because all compounds in this data set contain chiral atoms, and binding affinities of these compounds are available [37]. 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 Fig. 1 in reference [37] or Fig. 1 in reference [94].

Different methods were used to develop 3D-QSAR models for this data set, including CoMFA [37], Comparative Molecular Similarity Indices Analysis (CoMSIA) [97], Molecular Quantum Similarity Measures (MQSM) [98], Topological Quantum Similarity Indices (TQSI) [99], and Comparative Molecular Moment Analysis (CoMMA) [94], Mapping Property Distributions of Molecular Surfaces (MAP) [96], and so on [100–103].

The molecular set used in our study is made up of 31 steroids for which the binding affinity to the corticosteroid-binding globulin was measured. The names of the structures and the corresponding biological activities are listed in Table 8, the predicted values for this data set using 3D-Chiral (2.5) atom-based *TOMOCOMD-CARRD* MDs are also shown.

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) statistic has been used as a means of demonstrating predictive capability. The values of cross-validation square correlation coefficients for this data set fluctuate from 0.630 to 0.940, but most of them are concentrated between 0.705 and 0.799. These values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models [89–91]. 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 value of q^2 and the comparison can be more easily carried out. It is remarkably that, the results achieved with 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs show comparable results to other highly predictive QSAR models; even when they use more sophisti-

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

	Observed CBG affinity (pKa) ^a	Non-stoch. quadratic indices [69]	Stoch. qua- dratic indi- ces [69]	Non-stoch. linear indi- ces [66]	Stoch. lin- ear indices [66]	Non-stoch. bilinear indices [68]	Stoch. bilinear indices [68]
1	Aldosterone	-6.279	-6.259	-6.339	-6.149	-6.222	-6.289
2	Androstanediol	-5.000	-5.500	-4.641	-5.161	-4.984	-5.056
3	Androstenediol	-5.000	-5.026	-5.008	-4.965	-4.930	-4.999
4	Androstenedione	-5.763	-6.257	-6.615	-6.691	-6.583	-6.444
5	Androsterone	-5.613	-5.286	-5.526	-5.265	-5.342	-5.581
6	Corticosterone	-7.881	-7.242	-7.204	-7.283	-7.535	-7.354
7	Cortisol	-7.881	-7.226	-7.370	-7.380	-7.794	-7.603
8	Cortisone	-6.892	-7.212	-6.908	-6.892	-7.222	-7.385
9	Dehydroepiandrosterone	-5.000	-4.713	-5.094	-5.094	-5.033	-4.859
10	Deoxycorticosterone	-7.653	-7.081	-7.067	-7.307	-6.820	-7.351
11	Deoxycortisol	-7.881	-7.238	-7.393	-7.522	-7.202	-7.447
12	Dihydrotestosterone	-5.919	-5.570	-5.477	-5.700	-6.025	-5.704
13	Estradiol	-5.000	-5.096	-5.454	-4.803	-4.888	-5.181
14	Estrone	-5.000	-5.061	-4.937	-5.194	-5.071	-4.888
15	Ethiocholanolone	-5.000	-4.747	-5.175	-4.960	-4.954	-4.533
16	Pregnenolone	-5.255	-5.286	-5.526	-5.265	-5.342	-5.581
17	17-Hydroxyregnenolone	-5.255	-5.579	-5.671	-5.450	-5.529	-5.471
18	Progesterone	-5.000	-5.536	-5.598	-5.463	-5.405	-5.507
19	17-Hydroxyprogesterone	-7.380	-6.939	-7.110	-6.730	-6.889	-7.060
20	Testosterone	-7.740	-7.071	-7.116	-7.025	-6.954	-7.221
21	Prednisolone	-6.724	-6.572	-6.619	-6.535	-6.480	-6.866
22	Cortisolacetate	-7.512	-7.812	-7.830	-7.735	-7.687	-7.785
23	4-Pregnene-3,11,20-trione	-7.553	-7.789	-7.957	-7.700	-7.647	-7.754
24	Epicorticosterone	-6.779	-7.069	-6.587	-6.441	-7.007	-6.953
25		-7.200	-7.425	-7.671	-7.441	-7.695	-7.279

Table 8 continued

	Observed CBG affinity (pKa) ^a	Non-stoch. quadratic indices [69]	Stoch. qua- dratic indi- ces [69]	Non-stoch. linear indi- ces [66]	Stoch. lin- ear indices [66]	Non-stoch. bilinear indices [68]	Stoch. bilinear indices [68]
26	19-Nortestosterone	-6.549	-6.016	-6.858	-6.758	-6.651	-6.524
27	16a,17a-Dihydroxyprogesterone	-7.232	-5.715	-7.439	-6.118	-6.437	-6.246
28	16a-Methylprogesterone	-7.243	-7.319	-6.793	-7.239	-7.237	-7.187
29	19-Norprogesterone	-6.949	-6.936	-7.019	-7.927	-6.773	-6.386
30	2a-Methylcortisol	-7.756	-8.122	-7.773	-5.864	-8.016	-7.899
31	2a-Methyl-9a-fluorocortisol	-5.657	-5.970	-5.940	-6.824	-5.564	-5.579

^a Observed CBG affinity values taken from Robert et al. [98]

Table 9 Comparison between of prediction for the steroid data set 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs and other 3D QSAR approaches

QSAR method	<i>N</i>	<i>n</i>	Statistical method	q^2	Ref.
Similarity matrices-based molecular descriptors	31	6	genetic NN	0.940	[102]
TQSAR	31	6	MLR after PCA	0.842	[98]
3D-chiral bilinear indices (<i>stochastic</i>)	31	7	MLR	0.833	[68]
3D-chiral bilinear indices (<i>non-stochastic</i>)	31	6	MLR	0.799	[68]
3D-chiral linear indices (<i>stochastic</i>)	31	7	MLR	0.788	[66]
3D-chiral quadratic indices (<i>non-stochastic</i>)	31	6	MLR	0.781	[69]
MEDV	31	5	GA and RLM	0.777	[104]
TQSI	31	3	MLR	0.775	[99]
3D-chiral linear indices (<i>non-stochastic</i>)	31	6	MLR	0.767	[66]
MEDV	31	6	GA and RLM	0.765	[104]
3D-chiral quadratic indices (<i>stochastic</i>)	31	7	MLR	0.735	[69]
Similarity indices	31	1	PLS	0.734	[101]
E-State and kappa shape index*	31	4	MLR*	0.730	[105]
MQSM	31	4	MLR after PLS	0.727	[106]
E-State and kappa shape index	31	4	MLR	0.720	[105]
MQMS	31	3	MLR and PCA	0.705	[99]
CoMMA	31	6	PCR	0.689	[107]
MEDV	31	4	GA and RLM	0.648	[104]
Wagener's	31	–	k-NN and FNN	0.630	[100]

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

* one variable has a non-linear relationship

cated 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.

5 Concluding remarks

In these studies we demonstrated that atom-based 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs 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 summarized in the present work, the generalized atom-based 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs are 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. In this sense, we show that for three data sets chiral-QSAR models obtained with 3D-chiral (2.5) *TOMOCOMD-CARRD* MDs had better or similar predictive ability as compared to other previously reported chiral and/or 3D-QSAR.

6 Future outlooks

At present, most of researches working in drug discovery with the use of TIs concentrate their efforts in the development of more powerful MDs. In this sense, our research group is working in the definition of novel MDs based in group theory and geometric properties. We are also interested in apply our indices to codify planar and axial chirality as well as conformation alpha beta and other chirality data sets for example chromatographic retention. Also we have planed concentrate our efforts in the use of more sophisticated statistical techniques to be used with the *TOMOCOMD-CARRD* MDs.

Acknowledgements One of the authors (M-P. Y) thanks the program ‘Estades Temporals per a Investigadors Convidats’ for a fellowship to work at Valencia University (2008). M-P. Y also thanks support from Spanish MEC (Project Reference: SAF2006-04698). F. T. acknowledges financial support from the Spanish MEC DGI (Project No. CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173. Finally, but not less important, M-P. Y thanks are given to the projects entitle “Strengthening postgraduate education and research in Pharmaceutical Sciences”. This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium.

References

1. A. Golbraikh, D. Bonchev, A. Tropsha, *J. Chem. Inf. Comput. Sci.* **41**, 147–158 (2001)
2. J.V. de Julian-Ortiz, R. Garcia-Domenech, J. Galvez, R. Soler, F.J. García-March, G.M. Antón-Fos, *J. Chromatogr. A* **719**, 37–44 (1996)
3. V.M. Potapov, *Stereochemistry* (Khimia, Moscow, 1988)
4. K. Kislow, in *Fuzzy Logic in Chemistry*, ed. by D.H. Rouvray (Academic Press, San Diego, 1997), pp. 65–90
5. E. Eliel, S. Wilen, L. Mander, *Stereochemistry of Organic Compounds* (John Wiley & Sons Inc., 1994)
6. W.T. Kelvin, *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light* (Clay, C. J., London, 1904), p. 619
7. R.T. Morrison, R.N. Boyd, *Organic Chemistry* (Allyn and Bacon, New York, 1983)
8. J. Jacques, *Sur la Dissymétrie Moléculaire* (Christian Bourgeois, Paris, 1986)
9. L. Pasteur, in *Researches on the Molecular Asymmetry of Natural organic Products*, Alembic Club Reprint No. 14, (Alembic Club, Edinburgh, UK, 1905)
10. J. Solms, L. Vuataz, R.H. Egli, *Experientia* **21**, 692–694 (1965)
11. S.S. Schiffman, T.B. Clark, J. Gagnon 3rd, *Physiol. Behav.* **28**, 457–465 (1982)
12. M. Laska, P. Teubner, *Chem. Senses* **24**, 161–170 (1999)
13. E.H. Polak, A.M. Fombon, C. Tilquin, P.H. Punter, *Behav. Brain Res.* **31**, 199–206 (1989)
14. W.H. DeCamp, *Chirality* **1**, 2 (1989)
15. A.J. Hutt, S.C. Tan, *Drugs* **52**(Suppl 5), 1–12 (1996)
16. S. Wnendt, K. Zwingenberger, *Nature (London)* **385**, 303–304 (1997)
17. H. Schumacher, D.A. Blake, J.M. Gurian, J.R. Gillette, *J. Pharmacol. Exp. Ther.* **160**, 189–200 (1968)
18. S.C. Stinson, *Chem. Eng. News.* **78**, 43 (2000)
19. H.G. Diaz, I.H. Sanchez, E. Uriarte, L. Santana, *Comput. Biol. Chem.* **27**, 217–227 (2003)
20. P.-A. Guye, *Compt. Rend.* **714** (1890)
21. J. Aires-de-Sousa, J. Gasteiger, I. Gutman, D. Vidovic, *J. Chem. Inf. Comput. Sci.* **44**, 831–836 (2004)
22. J. Aires-de-Sousa, J. Gasteiger, *J. Mol. Graph. Model.* **20**, 373–388 (2002)
23. J. Aires-de-Sousa, J. Gasteiger, *J. Chem. Inf. Comp. Sci.* **41**, 369–375 (2001)
24. E. Ruch, *Acc. Chem. Res.* **5**, 49–56 (1972)
25. J.V. de Julian-Ortiz, C. de Gregorio Alapont, I. Rios-Santamarina, R. Garcia-Domenech, J. Galvez, *J. Mol. Graph. Model.* **16**, 14–18 (1998)
26. R. Benigni, M. Cotta-Ramusino, G. Gallo, F. Giorgi, A. Giuliani, M.R. Vari, *J. Med. Chem.* **43**, 3699–3703 (2000)

27. S.A. Wildman, G.M. Crippen, *J. Chem. Inf. Comput. Sci.* **43**, 629–636 (2003)
28. H.P. Schultz, E.B. Schultz, T.P. Schultz, *J. Chem. Inf. Comput. Sci.* **35**, 864–870 (1995)
29. J. Aires-de-Sousa, J. Gasteiger, *J. Comb. Chem.* **7**, 298–301 (2005)
30. E. Estrada, E. Uriarte, *Curr. Med. Chem.* **8**, 1573–1588 (2001)
31. A. Pyka, *J. Serb. Chem. Soc.* **62**, 251–269 (1997)
32. A. Pyka, *J. Planar Chromatogr. Mod. TLC* **6**, 282–288 (1993)
33. A. Pyka, *J. Liquid Chromatogr. Relat. Technol.* **22**, 41–50 (1999)
34. I. Gutman, A. Pyka, *J. Serb. Chem. Soc.* **62**, 261–265 (1997)
35. A.B. Buda, K. Mislow, *J. Mol. Struct. (Theochem)* **232**, 1–12 (1991)
36. G. Moreau, *J. Chem. Inf. Comput. Sci.* **37**, 929–938 (1997)
37. R.D. Cramer, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* **110**, 5959–5967 (1988)
38. P.J. Goodford, in *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*, ed. by F. Sanz, J. Giraldo, F. Manaut (Prous Science, Barcelona, 1995), pp. 199–205
39. Y. Marrero-Ponce, *Molecules* **8**, 687–726 (2003)
40. Y. Marrero-Ponce, *J. Chem. Inf. Comput. Sci.* **44**, 2010–2026 (2004)
41. Y. Marrero-Ponce, F. Torrens, Y.J. Alvarado, R. Rotondo, *J. Comput. Aided Mol. Des.* **20**, 685–701 (2006)
42. Y. Marrero-Ponce, A. Meneses-Marcel, O.M. Rivera-Borroto, R. García-Domenech, V. de Julián-Ortiz, A. Montero, J.A. Escario, A. Gómez Barrio, D. M. Pereira, J. J. Nogal, G. Grau, F. Torrens, C. Vogel, V. Arán, *J. Comput. Aided Mol. Des.* doi:10.1007/s10822-008-9171-1
43. Y. Marrero-Ponce, M.T.H. Khan, G.M. Casañola-Martin, A. Ather, K.M. Khan, M.K. Khan, F. Torrens, R. Rotondo, *J. Comput. Aided Mol. Des.* **21**, 167–188 (2007)
44. Y. Marrero-Ponce, M.T.H. Khan, G.M. Casañola-Martin, A. Ather, M.N. Sultankhodzhaev, F. Torrens, R. Rotondo, *Chem. Med. Chem.* **2**, 449–478 (2007)
45. Y. Marrero Ponce, M.A. Cabrera Perez, V. Romero Zaldivar, H. Gonzalez Diaz, F. Torrens, *J. Pharm. Pharmaceut. Sci.* **7**, 186–199 (2004)
46. Y. Marrero-Ponce, A. Montero-Torres, C.R. Zaldivar, M.I. Veitia, M.M. Perez, R.N. Sanchez, *Bioorg. Med. Chem.* **13**, 1293–1304 (2005)
47. Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E.A. Castro, *Bioorg. Med. Chem.* **13**, 2881–2899 (2005)
48. Y. Marrero-Ponce, R.M. Marrero, F. Torrens, Y. Martinez, M.G. Bernal, V.R. Zaldivar, E.A. Castro, R.G. Abalo, *J. Mol. Model.* **12**, 255–271 (2006)
49. Y. Marrero-Ponce, Y. Machado-Tugores, D.M. Pereira, J.A. Escario, A.G. Barrio, J.J. Nogal-Ruiz, C. Ochoa, V.J. Aran, A.R. Martinez-Fernandez, R.N. Sanchez, A. Montero-Torres, F. Torrens, A. Meneses-Marcel, *Curr. Drug Discov. Technol.* **2**, 245–265 (2005)
50. Y. Marrero-Ponce, M. Iyarreta-Veitia, A. Montero-Torres, C. Romero-Zaldivar, C.A. Brandt, P.E. Avila, K. Kirchgatter, Y. Machado, *J. Chem. Inf. Model.* **45**, 1082–1100 (2005)
51. Y. Marrero-Ponce, A. Huesca-Guillen, F. Ibarra-Velarde, *J. Mol. Struct. (Theochem)* **717**, 67–79 (2005)
52. Y. Marrero-Ponce, J.A. Castillo-Garit, F. Torrens, V. Romero-Zaldivar, E. Castro, *Molecules* **9**, 1100–1123 (2004)
53. Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A.M. Sanchez, F. Torrens, E.A. Castro, *Bioorg. Med. Chem.* **13**, 1005–1020 (2005)
54. 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, *J. Comput.-Aided Mol. Design* **18**, 615–634 (2004)
55. Y. Marrero-Ponce, M.A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio, F. Torrens, *Internet. Electron. J. Mol. Des.* **4**, 124–150 (2005)
56. Y. Marrero-Ponce, *Bioorg. Med. Chem.* **12**, 6351–6369 (2004)
57. Y. Marrero-Ponce, M.A. Cabrera, V. Romero, E. Ofori, L.A. Montero, *Int. J. Mol. Sci.* **4**, 512–536 (2003)
58. A. Montero-Torres, M.C. Vega, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J.A. Escario, V.J. Aran, A.R. Martinez-Fernandez, A. Meneses-Marcel, *Bioorg. Med. Chem.* **13**, 6264–6275 (2005)
59. A. Montero-Torres, R.N. Garcia-Sanchez, Y. Marrero-Ponce, Y. Machado-Tugores, J.J. Nogal-Ruiz, A.R. Martinez-Fernandez, V.J. Aran, C. Ochoa, A. Meneses-Marcel, F. Torrens, *Eur. J. Med. Chem.* (2006)

60. 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, Bioorg. Med. Chem. Lett. **15**, 3838–3843 (2005)
61. G.M. Casanola-Martin, M.T. Khan, Y. Marrero-Ponce, A. Ather, M.N. Sultankhodzhaev, F. Torrens, Bioorg. Med. Chem. Lett. **16**, 324–330 (2006)
62. Y. Marrero-Ponce, D. Nodarse, H.D. González, R. Ramos de Armas, V. Romero-Zaldivar, F. Torrens, E. Castro, Int. J. Mol. Sci. **5**, 276–293 (2004)
63. Y. Marrero Ponce, J.A. Castillo Garit, D. Nodarse, Bioorg. Med. Chem. **13**, 3397–3404 (2005)
64. Y. Marrero-Ponce, R. Medina, E.A. Castro, R. de Armas, H. González, V. Romero, F. Torrens, Molecules **9**, 1124–1147 (2004)
65. Y. Marrero-Ponce, R. Medina-Marrero, J.A. Castillo-Garit, V. Romero-Zaldivar, F. Torrens, E.A. Castro, Bioorg. Med. Chem. **13**, 3003–3015 (2005)
66. Y. Marrero-Ponce, J.A. Castillo-Garit, J. Comput.-Aided Mol. Design **19**, 369–383 (2005)
67. Y. Marrero-Ponce, H.G. Díaz, V. Romero, F. Torrens, E.A. Castro, Bioorg. Med. Chem. **12**, 5331–5342 (2004)
68. J.A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, R. Rotondo, J. Mol. Graph. Model. **26**, 32–47 (2007)
69. J.A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, Bioorg. Med. Chem. **14**, 2398–2408 (2006)
70. C.H. Edwards, D.E. Penney, *Elementary Linear Algebra* (Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1988)
71. L. Pauling, *The Nature of Chemical Bond* (Cornell University Press, Ithaca, NY, 1939), pp. 2–60
72. V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. **43**, 682–692 (2002)
73. R. Todeschini, P. Gramatica, Persp. Drug Disc. Des. **9–11**, 355–380 (1998)
74. D.H. Rouvray, in *Chemical Applications of Graph Theory*, ed. by A.T. Balaban (Academic Press, London, 1976), pp. 180–181
75. N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FL, 1983), pp. 32–33
76. I. Gutman, O.E. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer-Verlag, Berlin, 1986)
77. E. Estrada, G. Patlewicz, Croat. Chim. Acta. **77**, 203–211 (2004)
78. D.J. Klein, Internet Electron. J. Mol. Des. **2**, 814–834 (2003)
79. A. Browder, *Mathematical Analysis. An Introduction* (Springer-Verlag, New York, 1996), pp. 176–296
80. S. Axler, *Linear Algebra Done Right* (Springer-Verlag, New York, 1996), pp. 37–70
81. M. Randić, J. Math. Chem. **7**, 155–168 (1991)
82. P.D. Walker, P.G. Mezey, J. Am. Chem. Soc. **115**, 12423–12430 (1993)
83. L.B. Kier, L.H. Hall, *Molecular Structure Description. The Electrotopological State* (Academic Press, New York, 1999)
84. E.L. Eliel, S. Wilen, L. Mander, *Stereochemistry of Organic Compounds* (John Wiley & Sons Inc, New York, 1994)
85. 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)
86. B.L. Kier, L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis* (Research Studies Press, Letchworth, UK, 1986)
87. M. Vicent, B. Marchand, G. Rémond, S. Jaquelin-Guinamant, G. Damien, B. Portevin, J. Baupal, J. Volland, J. Bouchet, P. Lambert, B. Serkiz, W. Luitjen, M. Lauibie, P. Schiavi, Drug Des. Discov. **9**, 11 (1992)
88. STATISTICA version. 6.0 Statsoft, I
89. A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. **20**, 269–276 (2002)
90. S. Wold, L. Erikson, in *Chemometric Methods in Molecular Design* ed. by H. van de Waterbeemd (VCH Publishers, Weinheim, 1995), pp. 309–318
91. D.A. Belsey, E. Kuh, R.E. Welsch, *Regression Diagnostics* (Wiley, New York, 1980)
92. M.T. Cronin, T.W. Schultz, J. Mol. Struct. (Theochem) **622**, 39–51 (2003)
93. E.A. Coats, *In 3D QSAR in Drug Design*. (Kluwer/ESCOM:Dordrecht, 1998), pp. 119–213
94. B.D. Silverman, Quant. Struct.-Act. Relat. **19**, 237–246 (2000)
95. E.A. Coats, Persp. Drug Disc. Des. **12–14**, 199–213 (1998)
96. N. Stiefl, K. Baumann, J. Med. Chem. **46**, 1390–1407 (2003)
97. G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. **37**, 4130–4146 (1994)

98. D. Robert, L. Amat, R. Carbo-Dorca, *J. Chem. Inf. Comput. Sci.* **39**, 333–344 (1999)
99. M. Lobato, L. Amat, E. Besalu, R. Carbo-Dorca, *Quant. Struct.-Act. Relat.* **16**, 465–472 (1997)
100. M. Wagener, J. Sadowski, J. Gasteiger, *J. Am. Chem. Soc.* **117**, 7769–7775 (1995)
101. M.F. Parretti, R.T. Kroemer, J.H. Rothman, W.G. Richards, *J. Comput. Chem.* **18**, 1334–1353 (1997)
102. S.S. So, M. Karplus, *J. Med. Chem.* **40**, 4347–4359 (1997)
103. H. Chen, J. Zhou, G. Xie, *J. Chem. Inf. Comp. Sci.* **38**, 243–250 (1998)
104. S.S. Liu, C.S. Yin, L.S. Wang, *J. Chem. Inf. Comput. Sci.* **42**, 749–756 (2002)
105. H.H. Maw, L.H. Hall, *J. Chem. Inf. Comput. Sci.* **41**, 1248–1254 (2001)
106. E. Besalu, X. Girones, L. Amat, R. Carbo-Dorca, *Acc. Chem. Res.* **35**, 289–295 (2002)
107. B.D. Silverman, D.E. Platt, *J. Med. Chem.* **39**, 2129–2140 (1996)