

Novel 2D TOMOCOMD-CARDD molecular descriptors: atom-based stochastic and non-stochastic bilinear indices and their QSPR applications

Yovani Marrero-Ponce · Francisco Torrens ·
Ramón García-Domenech ·
Sadiel E. Ortega-Broche ·
Vicente Romero Zaldivar

Received: 10 February 2007 / Accepted: 23 November 2007 / Published online: 13 June 2008
© Springer Science+Business Media, LLC 2008

Abstract Novel atom-based molecular descriptors (MDs) based on a bilinear map similar to use defined in linear algebra are presented. These MDs, called “*local (atom, group and atom-type) and total (global) bilinear indices*”, are proposed here as a new molecular parametrization easily calculated from the 2D molecular information. The proposed non-stochastic and stochastic molecular fingerprints try to match molecular structure provided by the molecular topology by using the k th non-stochastic (Marrero-Ponce, J Chem Inf Comput Sci 44, 2010, 2004; Marrero-Ponce, Molecules 8, 687, 2003) and stochastic (Marrero-Ponce, et al. J Mol Struc (Theochem) 717, 67, 2005; Marrero-Ponce, Castillo-Garit, J Comput-Aided Mol Design 19, 369, 2005) graph-theoretic electronic-density matrices, \mathbf{M}^k and \mathbf{S}^k , respectively. That is to say, the k th non-stochastic and stochastic bilinear indices are calculated using \mathbf{M}^k and \mathbf{S}^k as matrix operators of bilinear transformations. Moreover, chemical information is codified by using different pair combinations of atomic weightings (atomic mass, polarizability, van der Waals volume, and electronegativity). The prediction ability in Quantitative

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

Y. Marrero-Ponce · F. Torrens
Institut Universitari de Ciència Molecular, Universitat de València, Edifici d’Instituts de Paterna, Poligon la Coma s/n (detras de Canal Nou), P. O. Box 22085, 46071 Valencia, Spain

Y. Marrero-Ponce · R. García-Domenech
Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Valencia, Spain

V. R. Zaldivar
Faculty of Informatics, University of Cienfuegos, Cienfuegos 59430, Cuba

Structure-Property Relationships (QSPR) of the new MDs was tested by analysing regressions of these MDs for six selected properties of octane isomers. It was clearly demonstrated that prediction ability was higher than those showed by other 2D/3D well-known sets of MDs. The obtained results suggest that with the present method it is possible to obtain a good estimation of these physicochemical properties for octanes. The approach described in the present report appears to be a prominent method to find quantitative models for description of physicochemical and biological properties.

Keywords *TOMOCOMD-CARDD* software · Non-stochastic and stochastic atom-based bilinear indices · QSPR model · Genetic algorithm · Physicochemical properties · Octane isomers

1 Introduction

Molecular descriptors (MDs) have deserved more and more attention from chemists along the later years [1,2]. In this connection, a large number of QSAR/QSPR studies have been reported in recent literature that uses MDs in prediction of the physicochemical and biological properties of molecules. MDs are numbers that characterize a specific aspect of molecular structure. The important common feature to all those MDs is the independence of their numerical values on renumbering atoms in a chemical structure [3,4].

At present, there are a great number of MDs that can be used in QSAR/QSPR and drug-design studies [3–5]. Among them, the so-called topological indices (TIs) are among most useful MDs known nowadays [6–8]. TIs are MDs derived from graph-theoretical invariants and codify structural information contained in ‘molecular connectivity’ [6–9].

The introduction of new TIs has focused the efforts of the scientists working on molecular topology [10–12]. In fact, our research group has proposed several atom- and bond-based topological and topographic MDs. In this sense, Estrada’s group has defined several bond-based MDs using the Randić-type graph-theoretical invariant, spectral moment (trace sum)-theoretical invariant and so on [13–19]. On the other hand, González and co-workers have developed a new method based on the Markov chain theory, which has been successfully employed in QSPR and QSAR studies [20–23]. More recently, one of the present authors (M-P. Y) has introduced new sets of total and atom-level MDs relevant to QSAR/QSPR studies and ‘rational’ drug design, atom-based quadratic [$q_k(x)$] and linear indices [$f_k(x)$] [24–27]. These MDs are based on the calculation of quadratic and linear maps similar to those defined in linear algebra [28]. Our research team has focused its efforts mainly on their application in drug design, so that many new drug(lead)-like compounds within different physicochemical and pharmacological skills have been found out [24,25,27,29–47]. In addition, these MDs have been extended to consider three-dimensional features of small/medium-sized molecules based on the trigonometric-3D-chirality-correction factor approach [26,48,49]. Finally, promising results have been found in the modeling of the interaction between drugs and HIV Ψ -RNA packaging-region in the field of bioinformatics using macromolecular indices [50,51]. An alternative formulation of our approach for structural characterization of proteins was also carried out recently [52,53].

The main purpose of the present paper is to present new sets of MDs, namely *non-stochastic and stochastic bilinear indices* and establish their abilities (both total and local) for the description of the molecular structure by correlating them with six selected physicochemical properties of octane isomers.

2 Methodology

In previous reports, we outlined outstanding features concerned with the theory of 2D atom-based *TOMOCOMD-CARDD* MDs. This method codifies the molecular structure by means of mathematical quadratic, linear and bilinear transformations. In order to calculate these algebraic maps for a molecule, the atom-based molecular vector, \bar{x} (vector representation) and k th “non-stochastic and stochastic graph–theoretic electronic-density matrices”, \mathbf{M}^k and \mathbf{S}^k correspondingly (matrix representations), are constructed [24, 25, 27, 29–47]. Such atom-adjacency relationships and chemical-information codification will be applied in the present report to generate a series of atom-based MDs, *atom, group and atom-type as well as total bilinear indices*, to be used in drug design and chemoinformatic studies.

Therefore the structure of this section will be as follows: (1) a background in atom-based molecular vector and non-stochastic and stochastic graph–theoretic electronic-density matrices will be described in the next Sections (2.1 and 2.2, respectively), and (2) an outline of the mathematical definition of bilinear maps and a definition of our procedures will be developed in subsections 2.3 and 2.4, correspondingly.

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 [24, 25, 27, 29–47]. The components (x) of \bar{x} are numeric values, which represent a certain standard atomic property (atom label). That is to say, these weights correspond to different atomic properties for organic molecules. Thus, a molecule having 5, 10, 15, \dots , n atomic nuclei can be represented by means of vectors, with 5, 10, 15, \dots , n components, belonging to the spaces \mathfrak{R}^5 , \mathfrak{R}^{10} , \mathfrak{R}^{15} , \dots , \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 1 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 Log P [54], surface contributions of polar atoms [55], atomic molar refractivity [56], atomic hybrid polarizabilities [57], Gasteiger-Marsilli atomic charge [58], atomic masses (M) [59], the van der Waals volumes (V) [59], the atomic polarizabilities [59], atomic electronegativity (E) in Pauling scale [60] and so on.

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

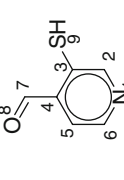
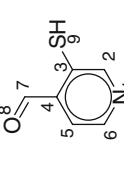
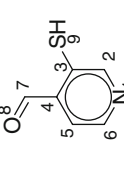
A)	molecular structure	molecular pseudograph (H-atoms-suppressed pseudograph) ^a					
							
B)	k^{th} non-stochastic graph–theoretic electronic-density matrices, \mathbf{M}^k ($k = 0-3$)						
zero order ($k = 0$)	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 &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

Table 1 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 ($^k s_{ij}$) have been approximated

Table 2 Values of the atomic weights used for bilinear indices calculation [59,60]

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

Now, 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. In the present report, we characterized each atomic nucleus with the following parameters: atomic masses (M) [59], the van der Waals volumes (V) [59], the atomic polarizabilities (P) [59], and atomic electronegativity (E) in Pauling scale [60]. The values of these atomic labels are shown in Table 2. 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 hyphen (-) expresses the combination (pair) of two selected atom-label chemical properties. In order to illustrate this, let us consider the same organic molecule as in the example above (3-mercapto-pyridine-4-carbaldehyde) and the following weighting scheme: M and V ($\bar{x}_M - \bar{y}_V = \bar{x}_V - \bar{y}_M$). The following molecular vectors, $\bar{x} = [14.01, 12.01, 12.01, 12.01, 12.01, 12.01, 12.01, 16.0, 32.07]$ and $\bar{y} = [15.599, 22.449, 22.449, 22.449, 22.449, 22.449, 22.449, 11.494, 24.429]$ are obtained when we use M and V as chemical weights for codifying each atom in the example molecule in \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*

Table 3 Values of atom-based bilinear indices for 3-mercapto-pyridine-4-carbaldehyde

		$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 8$	$k = 15$
<i>k</i> -th non-stochastic MDs, $b_{kL}(\bar{x}_i, \bar{y}_i)^a$							
Atom (<i>i</i>)	Atom-level bilinear indices						
N ₁	17,1589	69,30704	207,92112	677,6482	217716,1986	711876800,9	
C ₂	39,51024	105,09455	382,76839	1173,27708	380330,8902	1242200889	
C ₃	39,51024	172,57929	475,22504	1557,467468	498904,5116	1625489636	
C ₄	39,51024	158,04096	526,894748	1699,867938	545797,7789	1784733323	
C ₅	39,51024	118,53072	381,66623	1237,280798	406509,5989	1327324524	
C ₆	39,51024	105,09455	328,71982	1025,6697	337449,5327	1105114810	
C ₇	39,51024	77,743778	316,08192	837,86986	288840,4172	930027508,4	
O ₈	9,218188	38,233538	75,10629	305,868304	85899,64937	284074052,7	
S ₉	70,8441	54,04857	232,98981	647,3843	213952,7799	698815935,8	
Total (Sum)	334,282628	898,672996	2927,373368	9162,333648	2975401,357	9709657480	
Group (<i>i</i>)	Group bilinear indices						
Heteroatoms (Sum of local-index values for N, O and S atoms)	97,221188	161,589148	516,01722	1630,900804	517568,6278	1694766789	
Atom-type (<i>i</i>)	Atom-type bilinear indices						
CH-atom (Sum of local-index values for C-atoms 2, 5 and 6)	118,53072	328,71982	1093,15444	3436,227578	1124290,022	3674640223	
<i>k</i> -th stochastic MDs, ${}^s b_{kL}(\bar{x}_i, \bar{y}_i)^a$							
Atom (<i>i</i>)	Atom-level bilinear indices						
N ₁	17,1589	23,10234667	22,34781083	23,66748208	23,89914186	24,01863153	
C ₂	39,51024	33,38525667	40,80307944	39,00990209	38,9415613	38,95862823	
C ₃	39,51024	60,9142225	44,90684156	47,06658972	45,29450781	45,13795589	

Table 3 continued

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 8$	$k = 15$
C ₄	39,51024	42,80276	47,40669467	47,62823907	47,72612389	47,70941996
C ₅	39,51024	37,86398	37,79391122	39,18662953	40,21791136	40,32910683
C ₆	39,51024	35,03151667	35,08198711	34,64609483	36,44270378	36,7396616
C ₇	39,51024	27,639906	37,134922	32,10412486	34,42791207	33,95447149
O ₈	9,218188	16,116086	10,803039	14,98191176	13,79473544	14,04362415
S ₉	70,8441	29,6352825	42,1269325	34,8188677	34,3736933	34,24319914
Total (Sum)	334,282628	306,491357	318,4052183	313,1098416	315,1182908	315,1346988
Group bilinear indices						
Heteroatoms (Sum of local-index values for N, O and S atom)	97,221188	68,85371517	75,27778233	73,46826154	72,0675706	72,30545482
Atom-type bilinear indices						
CH-atom (Sum of local-index values for C-atoms 2, 5 and 6)	118,53072	106,2807533	113,6789778	112,8426264	115,6021765	116,0273967

^a Calculation development using VdW volume (V) and Polarizability (P) (see Table 2) as combination (pairs) of two atom-label chemical properties from our weighting schemes

is a collection of vertices (points) and edges (lines or bonds) connecting these vertices [61–63]. In more formal terms, a simple graph G is defined as an ordered pair $[V(G), E(G)]$ which consists of a non-empty set of vertices $V(G)$ and a set $E(G)$ of unordered pairs of elements of $V(G)$, called edges [61–63]. 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 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 $\{\{u, v\} | u, v \in V\}$ (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 [24, 25, 64].

In the earlier reports we have introduced new molecular matrices that describe changes along the time in the electronic distribution throughout the molecular backbone. 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 [24–41]. The coefficients ${}^k m_{ij}$ are the elements of the k th power of $\mathbf{M}(G)$ and are defined as follows [24–41]:

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

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 links 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 (a Kekule valence 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 (a Kekule valence structure, that is, structures without charges), such as furan, thiophene and pyrrol are represented by a multigraph. In order to illustrate the calculation of these matrices, let us consider the same molecule selected in the previous section. Table 1 depicts the molecular structure of this compound and its labeled molecular pseudograph. The zero ($k = 0$), first ($k = 1$), second ($k = 2$) and third ($k = 3$) powers of the non-stochastic graph–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 [65].

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 [26, 27, 38]:

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

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 non-negative entries having this property is called a “stochastic matrix” [28]. 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 [$p(\text{C}_{sp}) > p(\text{C}_{sp}^2) > p(\text{C}_{sp}^2 \text{ arom}) > p(\text{C}_{sp}^3)$] (see Table 1 for more details). This is a logical result if the electronegativity scale of these hybrid states is taken into account.

2.3 Mathematical bilinear forms: a theoretical framework

In mathematics, a **bilinear form** in a real vector space is a mapping $b : V \times V \rightarrow \mathfrak{R}$, which is linear in both arguments [66–71]. That is, this function satisfies the following axioms for any scalar α and any choice of vectors \bar{v} , \bar{w} , \bar{v}_1 , \bar{v}_2 , \bar{w}_1 and \bar{w}_2 .

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

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

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

$$\bar{x} = \sum_{i=1}^n x^i \bar{e}_i \quad (3)$$

and,

$$\bar{y} = \sum_{i=1}^n y^i \bar{e}_i \quad (4)$$

Subsequently,

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

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

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

Then,

$$b(\bar{x}, \bar{y}) = \sum_{i,j} a_{ij} x^i y^j = [X]^T A [Y] = [x^1 \dots x^n] \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix} \quad (7)$$

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

Finally, we introduce the formal definition of **symmetric bilinear form**. Let V be a real vector space and b be a bilinear function in $V \times V$. The bilinear function b is called symmetric if $b(\bar{x}, \bar{y}) = b(\bar{y}, \bar{x}), \forall \bar{x}, \bar{y} \in V$ [66–71]. Then,

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

2.4 Non-stochastic and stochastic atom-based bilinear indices: total (global) definition

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

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

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

where n is the number of atoms in 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 canonical basis set of \mathfrak{R}^n .

The defined Eqs. 9 and 10 for $b_k(\bar{x}, \bar{y})$ and ${}^s b_k(\bar{x}, \bar{y})$ may be also written as the single matrix equations:

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

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

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

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

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

In the last decade, Randić [72] 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 bilinear indices* computed for the whole molecule, a local-fragment (atomic, group or atom-type) formalism can be developed. These descriptors are termed *local non-stochastic and stochastic bilinear indices*, $b_{kL}(\bar{x}, \bar{y})$ and ${}^s b_{kL}(\bar{x}, \bar{y})$, respectively. The definition of these descriptors is as follows:

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

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

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:

$$\begin{aligned} {}^k m_{ijL} [{}^k s_{ijL}] &= {}^k m_{ij} [{}^k s_{ijL}] \quad \text{if both } v_i \text{ and } v_j \text{ are atomic nuclei contained within the} \\ &\quad \text{molecular fragment} \\ &= \frac{1}{2} {}^k m_{ij} [{}^k s_{ijL}] \quad \text{if } v_i \text{ or } v_j \text{ is an atomic nucleus contained within the} \\ &\quad \text{molecular fragment but not both} \\ &= 0 \quad \text{otherwise} \end{aligned} \quad (15)$$

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

$$b_L(\bar{x}, \bar{y}) = [X]^T \mathbf{M}_L^k [Y] \quad (16)$$

$${}^s b_L(\bar{x}, \bar{y}) = [X]^T \mathbf{S}_L^k [Y] \quad (17)$$

It should be remarked that the scheme above follows the spirit of a Mulliken population analysis [73]. It should be also pointed out that for every partitioning of a molecule into Z molecular fragments there will be Z local molecular fragment matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrix \mathbf{M}^k [\mathbf{S}^k] can be correspondingly partitioned into Z local matrices \mathbf{M}_L^k [\mathbf{S}_L^k], $L = 1, \dots, Z$, 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. In this way, the total non-stochastic and stochastic bilinear indices are the sum of the non-stochastic and stochastic bilinear indices, respectively, of the Z molecular fragments:

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

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

Atomic, group and atom-type bilinear fingerprints are specific cases of local bilinear indices. Atomic bilinear indices, $b_{kL}(\bar{x}_i, \bar{y}_i)$ and ${}^s b_{kL}(\bar{x}_i, \bar{y}_i)$, can be computed for each atom i in the molecule and contain electronic and topological structural information from all other atoms within the structure. The atom-level bilinear indices values for the common scaffold atoms can be 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.

In addition, the atom-type bilinear indices can also be calculated. In the same way as atom-type E-state values [11], 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 bilinear indices formalism is applicable without the need for superposition [74, 75]. For this reason the present method represents a significant advantage over traditional QSAR methods. The atom-type bilinear descriptors are calculated by adding the k th atomic bilinear indices 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 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 [11, 74, 75]. 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 [11]. 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 (H-bonding) to heteroatoms (O, N and S in all valence states), halogen atoms (F, Cl, Br and I), all aliphatic carbon chains (several atom types), all aromatic atoms (aromatic rings), and so on. The group-level bilinear indices are the sum of the individual atom-level bilinear indices for a particular group of atoms. For all data set structures, the k th group-based bilinear indices provide important information for QSAR/QSPR studies.

2.6 Sample calculation

It is useful to perform a calculation on a molecule to illustrate the effect of structure on atomic and global bilinear indices values. For this we use the 3-mercapto-pyridine-4-carbaldehyde molecule. The labeled (atom numbering) molecular structure of this chemical and the non-stochastic and stochastic (atom-level, group and atom-type as well as total) atom-based bilinear indices are shown at the Table 3.

3 Computational strategies

All computations were carried out on a PC Pentium-4 2.0GHz. 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 [76]. 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, i.e., atomic mass (M), atomic polarizability (P), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (E). The values of these atomic labels are shown in Table 2; (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. That is to say, we can find a quantitative relation between a property \mathbf{P} and the bilinear fingerprints having, for instance, the following appearance,

$$\mathbf{P} = a_0 \mathbf{b}_0(x, y) + a_1 \mathbf{b}_1(x, y) + a_2 \mathbf{b}_2(x, y) + \dots + a_k \mathbf{b}_k(x, y) + c \quad (20)$$

where \mathbf{P} is the measured property, $\mathbf{b}_k(x, y)$ are the k th non-stochastic total bilinear indices, and the a'_{ks} are the coefficients obtained by the MRA; (5) test the robustness and predictive power of the QSPR/QSAR equation by using internal cross-validation techniques.

The atom-based *TOMOCOMD-CARDD* MDs computed in this study were the following:

- (i) k th ($k = 15$) total (global) non-stochastic bilinear indices not considering and considering H-atoms in the molecule [$\mathbf{b}_k(x, y)$ and $\mathbf{b}_k^{\text{H}}(x, y)$, respectively].
- (ii) k th ($k = 15$) total (global) stochastic bilinear indices considering H-atoms in the molecule [$^{\text{s}}\mathbf{b}_k^{\text{H}}(x, y)$].
- (iii) k th ($k = 15$) group (methyl group, $-\text{CH}_3$) non-stochastic and stochastic bilinear indices considering H-atoms in the molecule. [$\mathbf{b}_{kL(\text{CH}_3)}^{\text{H}}(x, y)$ and $^{\text{s}}\mathbf{b}_{kL(\text{CH}_3)}^{\text{H}}(x, y)$, respectively]. These local descriptors are calculated taken into account

only one of the three bond types for carbon-hydrogen bonds ($C_{\text{primary}}\text{-H}$) that there are for octanes data.

4 Data sets

The data sets for this study were taken from Consonni et al. [77], because the data of physicochemical properties of octane isomers have been carefully selected for testing MDs. Octanes constituted a good set of chemicals for comparative study, since many experimental data among their physicochemical properties and QSPR studies are available [77–84]. In this sense, we analyzed the quality of the obtained QSPR models to describe the boiling point (BP), motor octane number (MON), heat of vaporization (HV), molar volume (MV), entropy (S), and heat of formation ($\Delta_f H$) of the octane isomers.

The use of octanes as a very suitable data set for testing TIs have been advocated by Randić and Trinajstić [85,86]. This selection is recommended due to the most of the fact that physicochemical properties commonly studied in QSPR analyses with TIs are interrelated for data sets of compounds with different molecular weights, for instance for alkanes with two to nine carbon atoms. These correlations are not necessarily observed when the same indices are used in isomeric data sets of compounds, such as the octane data set. In addition, these properties are hardly interrelated when octanes are used as a data set [82]. On the other hand, all the TIs are designed to have (gradual) increments with the increase in the molecular weight. By this way, if we do the present study by using a series of compounds having different molecular weights, we will find “false” interrelations between the indices by an overestimation of the size effects inherent to these MDs [79]. The same is also valid when the QSPR model is to be obtained. It is not difficult to find “good” linear correlations between TIs and physicochemical properties of alkanes in data sets with great size variability [79]. In fact, the simple use of the number of vertices in the molecular graph produced regression coefficients greater than 0.97 for most of the physicochemical properties of C2–C9 alkanes studied by Needham et al. [87] However, when data sets of isomeric compounds are considered, correlations that typically have high correlation coefficients when molecules of different sizes are considered will no longer show such good linear correlation. In conclusion, if a new proposed MD is not able to model the variation of at least one property of octane isomers, then it probably does not contain any useful molecular information.

5 Statistical method

The k th total and local atom-based bilinear indices were used as molecular descriptors for derived QSPRs. One of the difficulties with the large number of descriptors is deciding which ones will provide the best regressions. Furthermore, as testing a large number of all possible combinations of variables would be a tedious task and time-consuming procedure, we have used a genetic algorithm (GA) input selection [88–93].

GAs are a class of algorithms inspired by the process of natural evolution in which species having a high fitness under some conditions can prevail and survive to the next generation; the best species can be adapted by crossover and/or mutation in the search for better individuals. Genetic function approximation (GFA), a combination of GA and the linear polynomials, higher-order polynomials, splines (multivariate adaptive regression splines algorithm), or other non-linear functions, provides multiple models with high predictive ability [88–95].

The BuildQSAR [96] software was employed to perform variable selection and QSPR modeling. The mutation probability was specified as 35%. The length of the equations was set for three terms and a constant. The population size was established as 100. The GA with an initial population size of 100 rapidly converged (200 generations) and reached an optimal QSAR model in a reasonable number of GA generations. The search for the best model can be processed in terms of the highest correlation coefficient (R) or F -test equations (Fisher-ratio's p -level [$p(F)$]), and the lowest standard deviation equations (s) [96]. The quality of models was also determined by examining the Leave-One-Out (LOO) cross-validation (CV) (q^2 , s_{cv}) [97].

6 QSPR Application of physicochemical properties of octane isomers

The decisive criterion of quality for any MD is its ability to describe structure-related properties of molecules. In order to illustrate the applicability of the novel *TOMOCOMD-CARDD* MDs, we performed the QSPR models to describe six physicochemical properties of octane isomers. In this sense, regressions of octane properties based on the non-stochastic and stochastic bilinear indices will be compared to some regressions based on 2D (topologic) and 3D (geometric) descriptors taken from the literature [77–84]. Precisely, to evaluate the quality of the models based on our novel atom-based MDs we have taken as references: (1) the models published by Randić [80–82] based on diverse TIs such as the Wiener matrix invariants, (2) the equation published by Diudea [84] based on the SP indices, and (3) the best models obtained from a set constituted by the topological (69), WHIM (99), and GETAWAY descriptors (197) [77].

The best linear models found using non-stochastic and stochastic total and atom-type bilinear indices are presented in Table 4, together with the statistical information for the best regressions with 1, 2, and 3 MDs published so far [77–84] for each selected property of octane isomers.

As it can be appreciated from the statistical parameters of regression equations in Table 4, all of the physicochemical properties were well described by atom-based bilinear indices. In this table we can observe that the statistical parameters for the models obtained with atom-level bilinear indices to describe heat of vaporization (HV) (Eqs. 25 and 26), and molar volume (MV) (Eqs. 27 and 28) as well as heat of formation ($\Delta_f H$) (Eqs. 31 and 32) of octanes are better than those taken from the literature using 2D and 3D MDs. The second physicochemical property, that is MV, is well described exclusively by the atom-based Bilinear indices (both non-stochastic and stochastic ones). In addition, the obtained QSPR model (Eq. 32) for description of $\Delta_f H$ using only one stochastic linear index [$V-Es b_5^H(x, y)$] showed better statistical results and

Table 4 Statistical information for best multiple regression models of selected physicochemical properties of octane isomers

Property	Method	Size	Q^2_{LOO}	R^2	s	F	Model descriptors	Eq/Ref	
Boiling point (BP)	Non-stochastic atom-based bilinear indices	3	96.1	97.7	1.021	198.64	$BP = -1703.67 - 0.744E - P b_{0L(CH3)} H(x, y) + 0.056M - V b_{10} H(x, y) - 4.01 \times 10^{-6} M V b_{10} H(x, y)$	(21)	
		3	92.7	96.23	1.314	118.18	$BP = -3353.14 - 1.54V - E s b_9 H(x, y) + 16.83M - E s b_3 H(x, y) M - E s b_{0L(CH3)} H(x, y)$	(22)	
		3	98.12	98.78	0.744			$2 \chi^2 \bar{\chi} HATS_6(p)$	[77]
		3	97.10	98.32	0.897			$HATS_2(v) R_4(u) R_6(v)$	[77]
		2	96.62	97.58	1.013			$2 \chi HATS_6(p)$	[77]
		3	95.84	95.84	1.394			$S^3W S^4W SJ$	[84]
		2	94.78	94.78	1.508			$S^3W S^4W$	[84]
		2	84.86	89.62	2.098			$HATS_2(m) R_4^+(u)$	[77]
		2		81.36	2.810			$WW x_1$	[82]
		1		78.85	2.90			Z	[81]
Motor octane number (MON)	Non-stochastic atom-based bilinear indices	1	66.47	74.64	3.175		$HATS_2(m)$	[77]	
		1		67.77	3.630		$2 \chi W$	[84]	
		3	98.4	99.2	3.327	511.07		$MON = -557.56 + 0.23 \times 10^{-3} M - E b_9 H(x, y) - 2.17 \times 10^{-5} M - P b_{14}(x, y) + 0.42 \times 10^{-3} M - V b_7 H(x, y)$	(23)
		3	97.6	98.6	4.319	301.58		$MON = -20.03 + 8.13M - E s b_{05L(CH3)} H(x, y) - 20.24M - E s b_{5L(CH3)} H(x, y) + 12.11M - E s b_{15L(CH3)} H(x, y)$	(24)
		3	98.58	99.23	2.439			$v I_D^M Ts HATS_1(m)$	[77]
		3	97.42	98.62	3.259			$HATS_4(u) HATS_7(v) R_7(p)$	[77]
		2	98.05	98.05	3.855			$SX^1W \chi^7W \chi^3W$	[84]
		3	96.77	97.68	4.053			Ts H ₄ (e)	[77]
		2	91.28	95.78	5.466			$HATS_7(m) R_4(u)$	[77]
		2	95.64	95.64	5.533			$SX^1W SX^3W$	[84]
1		95.22	5.589			X^1W	[84]		

Table 4 continued

Property	Method	size	Q_{LOO}^2	R^2	s	F	Model descriptors	Eq/Ref
Heat of vaporization (HV)	getaway + whim + top. topological	1	90.83	92.40	7.069		Ts	[77]
	getaway	1	85.64	91.97	7.270		I_{wD}	[81]
	Non-stochastic atom-based bilinear indices	3	98.4	98.9	8.515	422.46	REIG	[77]
Heat of vaporization (HV)	Stochastic atom-based bilinear indices	3	97.0	98.4	0.233	294.01	$HV = 249.75 + 9.89 \times 10^{-8}M - E_{\text{b}}b_{14L}(\text{CH}_3)H(x, y)$ $- 0.196 \times 10^{-2}M - V_{\text{b}}b_4(x, y)$ $+ 0.822 \times 10^{-2}P - V_{\text{b}}b_4(x, y)$ $- 0.45V - P_{\text{b}}b_{14}H(x, y)$ $- 0.02M - E_{\text{b}}b_{14L}(\text{CH}_3)H(x, y)$	(25)
							$HV = 339.98 + 0.78E - P_{\text{b}}b_4(x, y)$ $- 0.196 \times 10^{-2}M - V_{\text{b}}b_4(x, y)$ $+ 0.822 \times 10^{-2}P - V_{\text{b}}b_4(x, y)$ $- 0.45V - P_{\text{b}}b_{14}H(x, y)$ $- 0.02M - E_{\text{b}}b_{14L}(\text{CH}_3)H(x, y)$	(26)
Molar volume (MV)	getaway + whim + top. getaway	3	97.57	98.42	0.281		$0\bar{\chi}^3, \kappa R_6^+(u)$	[77]
	getaway	3	95.46	97.18	0.375		$HATS_6(u)R_4(u)R_1^+(m)$	[77]
	getaway + whim + top. topological	2	95.18	96.53	0.402		$2\chi R_6^+(u)$	[77]
	getaway	3	93.15	94.87	0.459		$\chi^1W\chi^2W\chi^3W$	[84]
	topological	2	92.62	92.62	0.488		$HATS_4(u)R_6(e)$	[77]
	topological	2	91.78	91.78	0.429		$4W^5W$	[84]
	getaway + whim + top. getaway	1	80.80	88.61	0.705		Z	[81]
	topological	1	79.74	85.70	0.790		2χ	[77]
	Non-stochastic atom-based bilinear indices	2	98.2	99.1	0.820		$R_2(m)$	[77]
	Non-stochastic atom-based bilinear indices	3	98.7	99.1	0.273	461.48	WW x1	[82]
Stochastic atom-based bilinear indices	Stochastic atom-based bilinear indices	3	98.7	99.1	0.264	491.33	$MV = 198.26 + 0.0144E - V_{\text{b}}b_4(x, y)$ $- 0.0138M - V_{\text{b}}b_3H(x, y)$ $- 8.15 \times 10^{-7}M - V_{\text{b}}b_{10}(x, y)$ $MV = 686.67 + 0.22P - M_{\text{b}}b_{12}H(x, y)$ $- 0.28M - V_{\text{b}}b_3H(x, y)$ $- 0.04M - V_{\text{b}}b_3H(x, y)$	(27)
							$MV = 198.26 + 0.0144E - V_{\text{b}}b_4(x, y)$ $- 0.0138M - V_{\text{b}}b_3H(x, y)$ $- 8.15 \times 10^{-7}M - V_{\text{b}}b_{10}(x, y)$ $MV = 686.67 + 0.22P - M_{\text{b}}b_{12}H(x, y)$ $- 0.28M - V_{\text{b}}b_3H(x, y)$ $- 0.04M - V_{\text{b}}b_3H(x, y)$	(28)
getaway + whim + top. getaway	getaway + whim + top. getaway	3	75.96	92.01	1.825		$K_s R_7^+(u) RT^+(m)$	[77]
							$HATS_6(p) RT^+(m) R_1(v)$	[77]

Table 4 continued

Property	Method	size	Q_{LOO}^2	R^2	s	F	Model descriptors	Eq/Ref
Entropy (S)	topological	3		88.29	2.210		$5W^6W^7W$	[84]
	getaway + whim + top.	2	54.49	84.96	2.419		$vI_D^M R_6^+(u)$	[77]
	getaway	2	45.49	81.79	2.662		$R_6^+(u) R_4(v)$	[77]
	getaway + whim + top.	1	32.66	67.61	3.437		$R_6(v)$	[77]
	topological	2		62.76	3.807		$3W^4W$	[84]
	topological	1		60.85	3.780		$7W$	[84]
	Non-stochastic atom-based	1	88.5	91.1	1.32	180.91	$S = 235.31 - 0.019M - E_3b_3^H(x, y)$	(29)
	bilinear indices							
	Stochastic atom-based	3	90.9	95.3	1.08	94.12	$S = 1070.60 + 8.71E - P_5b_2^H(x, y)$ $- 2.58V - E_5b_4^H(x, y) + 0.72V - P_5b_5^H(x, y)$	(30)
	bilinear indices							
Heat of formation ($\Delta_f H$)	getaway + whim + top.	3	97.17	97.96	0.711		$vI_{D,deg} TWC R_2^+(p)$	[77]
	getaway + whim + top.	2	96.42	97.14	0.814		$vI_{D,deg} TWC$	[77]
	getaway	3	93.45	95.84	1.016		$I_{SH} HATS_8(m) R_3(v)$	[77]
	getaway	2	92.19	94.76	1.101		$I_{SH} R_3(v)$	[77]
	getaway + whim + top.	1	89.86	92.51	1.274		$R_3(v)$	[77]
	topological	1		91.10	1.400		$\chi^{[1/2]}$	[81]
	topological	2		81.72	2.060		$x_1 x_2$	[82]
	Non-stochastic atom-based	3	97.3	98.4	0.053	292.84	$\Delta_f H = 36.211.46 \times 10^{-8} M - E_{14L} b_{14L}(CH_3)^H(x, y)$ $- 0.29 \times 10^{-3} M - V b_4^H(x, y)$ $+ 1.12 \times 10^{-3} P - V b_4^H(x, y)$	(31)
	bilinear indices							
	Stochastic atom-based	bilinear indices	1	96.6	97.61	0.062	632.63	$\Delta_f H = 117.78 - 0.15V - E_5b_5^H(x, y)$
getaway + whim + top.		3	95.06	96.60	0.254		$HATS_5(m) HATS_7(m) R_4(e)$	[77]
getaway + whim + top.		2	90.96	93.24	0.346		$^2\chi HATS_2(e)$	[77]
getaway		2	90.18	92.87	0.356		$HATS_7(u) R_2(m)$	[77]
getaway + whim + top.		1	87.18	89.34	0.421		$HATS_2(m)$	[77]
topological		3		87.05	0.492		$\Omega_1 \Omega_2 \Omega_3$	[80]
topological		2		86.86	0.478		$\Omega_1 \Omega_2$	[80]
topological		1		86.68	0.471		$1/2^2 \chi$	[81]
topological		2		78.70	0.570		$WW x_1$	[82]

predictive power that the equations developed from 2D and 3D (and their combination) MDs.

It should be pointed out that in the models based on the atom-level chemical bilinear indices, both regressions for the motor octane number (MON) (Eqs. 23 and 24) are better-to-similar than the models published so far. Only the models found by us to describe boiling point (BP) (Eqs. 21 and 22) and entropy (S) (Eqs. 29 and 30) have significant differences with the preceding models obtained by applying the selection procedure to the set given by GETAWAY descriptors plus WHIM and topological indices. However, these properties were described with our approach better than using several TIs.

According to the obtained QSPR results, it is possible to conclude that the novel MDs encode some useful molecular information different from that of previously proposed descriptors. Moreover, they are quite diverse among themselves being able to describe well the variation of different properties of octanes.

7 Conclusions

The approach described in this report appears to be a prominent method to find quantitative models for description of physical, thermodynamic, or biological properties. The novel MDs proposed here have been shown to have some interesting features, such as:

- i. their functional definitions are based on well-known and accepted algorithms and formulas in mathematics. These novel atom-based molecular descriptors are based on a bilinear map similar to those defined in linear algebra. The atom- and group-level as well as atom-type formalism will permit to expedite investigation of molecular mechanisms and rational design of molecules at the local level. These local bilinear indices together with global ones are now added as a new set of MDs to the significant arsenal of whole-molecule indices.
- ii. they can be easily and quickly calculated. That is, the calculation is simple and straightforward, requiring only the 2D information.
- iii. they show good prediction power in physicochemical-properties modeling. The obtained QSPR models presented here yield a good statistical account of the octanes data, giving both sound direct statistical as well as the press statistics. Further, it was clearly demonstrated that prediction ability was higher than those showed by other 2D/3D well-known sets of MDs.

Despite these positive features of non-stochastic and stochastic atom-based bilinear indices, additional study has to be done to further investigate their meaning and behaviour with respect to the structural features of the molecules. The applications of the present method to QSPR/QSAR and drug-design studies as well as similarity/diversity analysis of several classes of organic compounds are now in progress and will be the subject of a future publication. In fact, several promissory results have been achieved with the use of these novel total and atom-level MDs in *computational* drug discovery of new trichomonacids [98].

Acknowledgements Yovani Marrero-Ponce (M-P. Y) acknowledges the Valencia University (Spain) for kind hospitality during the second semester of 2007. M-P. Y thanks are given to the international relationships of the Valencia University for partial financial support as well as the program ‘Estades Temporals per an Investigadors Convitats’ for a fellowship to work at Valencia University (2007-2008). 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. R.G.D. acknowledges financial support from Spanish Ministry of Health (Project FIS No. SAF2005/PI052128). Finally, but not less, 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. J. Gálvez, *J. Chem. Inf. Comput. Sci.* **43**, 1231 (2003)
2. P.G. Seybold, M. May, U.A. Bagal, *J. Chem. Educ.* **64**, 575 (1987)
3. M. Karelson, *Molecular Descriptors in QSAR/QSPR* (Wiley, New York, 2000)
4. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors* (Wiley-VCH, Germany, 2000)
5. M. Randić, *J. Chem. Inf. Comput. Sci.* **37**, 672 (1997)
6. E. Estrada, E. Uriarte, *Curr. Med. Chem.* **8**, 1573 (2001)
7. J. Devillers, A.T. Balaban, *Topological Indices and Related Descriptors in QSAR and QSPR* (Gordon and Breach Amsterdam, The Netherlands, 1999)
8. F. Torrens, *Comb. Chem. High Throughput Screen* **6**, 801 (2003)
9. J. Gálvez, R. García, M.T. Salabert, R. Soler, J. Chem. Inf. Comput. Sci. **34**, 520 (1994)
10. D.H. Rouvray, *Chemical Applications of Graph Theory* (Academic Press, London, 1976), p. 180
11. L.B. Kier, L.H. Hall, *Molecular Structure Description. The Electrotopological State* (Academic Press, New York, 1999)
12. E. Estrada, *Chem. Phys. Lett.* **336**, 248 (2001)
13. E. Estrada, *J. Chem. Inf. Comput. Sci.* **35**, 31 (1995)
14. E. Estrada, *J. Chem. Inf. Comput. Sci.* **36**, 844 (1996)
15. E. Estrada, *J. Chem. Inf. Comput. Sci.* **39**, 1042 (1999)
16. E. Estrada, N. Guevara, I. Gutman, *J. Chem. Inf. Comput. Sci.* **38**, 428 (1998)
17. E. Estrada, E. Molina, *J. Mol. Graph. Model.* **20**, 54 (2001)
18. E. Estrada, E. Molina, *J. Chem. Inf. Comput. Sci.* **41**, 791 (2001)
19. E. Estrada, A. Ramirez, *J. Chem. Inf. Comput. Sci.* **36**, 837 (1996)
20. H. Gonzalez-Diaz, E. Tenorio, N. Castanedo, L. Santana, E. Uriarte, *Bioorg. Med. Chem.* **13**, 1523 (2005)
21. H. Gonzalez-Diaz, L.A. Torres-Gomez, Y. Guevara, M.S. Almeida, R. Molina, N. Castanedo, L. Santana, E. Uriarte, *J. Mol. Model.* **11**, 116 (2005)
22. A.R. Ramos, H. González-Díaz, R. Molina, M.P. González, E. Uriarte, *Bioorg. Med. Chem.* **12**, 4815 (2004)
23. H. González-Díaz, O. Gia, E. Uriarte, I. Hernández, R. Ramos, M. Chaviano, S. Seijo, J.A. Castillo, L. Morales, L. Santana, D. Akpaloo, E. Molina, M. Cruz, L. Torres, M.A. Cabrera, *J. Mol. Mod.* **9**, 395 (2003)
24. Y. Marrero Ponce, *J. Chem. Inf. Comput. Sci.* **44**, 2010 (2004)
25. Y. Marrero-Ponce, *Molecules* **8**, 687 (2003)
26. Y. Marrero-Ponce, J.A. Castillo-Garit, *J. Comput.-Aided Mol. Des.* **19**, 369 (2005)
27. Y. Marrero-Ponce, A. Huesca-Guillen, F. Ibarra-Velarde, *J. Mol. Struct. (Theochem)* **717**, 67 (2005)
28. C.H. Edwards, D.E. Penney, *Elementary Linear Algebra* (Prentice-Hall, Englewood Cliffs, NJ, USA 1988)
29. Y. Marrero-Ponce, Y.J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, E. Jorge, A. del Valle, F. Torrens, E.A. Castro, *J. Comput.-Aided Mol. Des.* **18**, 615 (2004)
30. 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 (2005)
31. Y. Marrero-Ponce, M.A. Cabrera, V. Romero, E. Ofori, L.A. Montero, *Int. J. Mol. Sci.* **4**, 512 (2003)

32. Y. Marrero-Ponce, *Bioorg. Med. Chem.* **12**, 6351 (2004)
33. Y. Marrero-Ponce, M.A. Cabrera, V. Romero, D.H. González, F. Torrens, *J. Pharm. Pharmaceut. Sci.* **7**, 186 (2004)
34. 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. Des.* **18**, 615 (2004)
35. Y. Marrero-Ponce, J.A. Castillo-Garit, F. Torrens, V. Romero-Zaldivar, E.A. Castro, *Molecules* **9**, 1100 (2004)
36. Y. Marrero-Ponce, M.A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio, F. Torrens, *Int. Electron. J. Mol. Des.* **4**, 124 (2005)
37. Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A. Sanchez, F. Torrens, E.A. Castro, *Bioorg. Med. Chem.* **13**, 1005 (2005)
38. 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. Comput. Sci.* **45**, 1082 (2005)
39. Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E.A. Castro, *Bioorg. Med. Chem.* **13**, 2881 (2005)
40. Y. Marrero-Ponce, A. Meneses-Marcel, Y. Machado-Tugores, D. Montero Pereira, J.A. Escario, J.J. Nogal-Ruiz, C. Ochoa, V.J.A.R. Martínez-Fernández, R.N. García Sánchez, A. Montero-Torres, F. Torrens, *Curr. Drug Disc. Tech.* **2**, 245 (2005)
41. Y. Marrero-Ponce, A. Montero-Torres, C.R. Zaldivar, M. Veitia, M. M. Perez, R.N. Sanchez, *Bioorg. Med. Chem.* **13**, 1293 (2005)
42. A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado-Tugores, A. Montero-Torres, D. Montero Pereira, J.A. Escario, J.J. Nogal-Ruiz, C. Ochoa, V.J. Arán, A.R. Martínez-Fernández, R.N. García, *Bioorg. Med. Chem. Lett.* **17**, 3838 (2005)
43. A. Montero-Torres, M. Celeste Vega, Y. Marrero-Ponce, M. Rolón, A. Gómez-Barrio, J.A. Escario, V.J. Arán, A.R. Martínez-Fernández, A. Meneses-Marcel, *Bioorg. Med. Chem.* **13**, 6264 (2005)
44. G.M. Casañola-Martin, M.T.H. Khan, Y. Marrero-Ponce, A. Ather, M.N. Sultankhodzhaev, F. Torrens, *Bioorg. Med. Chem. Lett.* **16**, 324 (2006)
45. M. Celeste-Vega, A. Montero-Torres, Y. Marrero-Ponce, M. Rolón, A. Gómez-Barrio, J.A. Escario, V.J. Arán, A.R. Martínez-Fernández, A. Meneses-Marcel, *Bioorg. Med. Chem. Lett.* **16**, 1898 (2006)
46. Y. Marrero-Ponce, R.M. Marrero, F. Torrens, Y. Martinez, M.G. Bernal, V.R. Zaldivar, E.A. Castro, R.G. Abalo, *J. Mol. Mod.* **12**, 255 (2006)
47. A. Montero-Torres, R.N. García-Sánchez, Y. Marrero-Ponce, Y. Machado-Tugores, J.J. Nogal-Ruiz, A.R. Martínez-Fernández, V.J. Arán, C. Ochoa, A. Meneses-Marcel, F. Torrens, *Eur. J. Med. Chem.* **41**, 483 (2006)
48. Y. Marrero-Ponce, H.G. Díaz, V. Romero, F. Torrens, E.A. Castro, *Bioorg. Med. Chem.* **12**, 5331 (2004)
49. J.A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, *Bioorg. Med. Chem.* **14**, 2398 (2006)
50. Y. Marrero Ponce, J.A. Castillo-Garit, D. Nodarse, *Bioorg. Med. Chem.* **13**, 3397 (2005)
51. Y. Marrero-Ponce, D. Nodarse, H.D. González, R.R. de Armas, V. Romero-Zaldivar, F. Torrens, E.A. Castro, *Int. J. Mol. Sci.* **5**, 276 (2004)
52. Y. Marrero-Ponce, R. Medina, E.A. Castro, R.R. de Armas, H. González, V. Romero, F. Torrens, *Molecules* **9**, 1124 (2004)
53. Y. Marrero-Ponce, R. Medina-Marrero, J.A. Castillo-Garit, V. Romero-Zaldivar, F. Torrens, E.A. Castro, *Bioorg. Med. Chem.* **13**, 3003 (2005)
54. R. Wang, Y. Gao, L. Lai, *Perspect. Drug Discov. Des.* **19**, 47 (2000)
55. P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **43**, 3714 (2000)
56. A.K. Ghose, G.M. Crippen, *J. Chem. Inf. Comput. Sci.* **27**, 21 (1987)
57. K.J. Millar, *J. Am. Chem. Soc.* **112**, 8533 (1990)
58. J. Gasteiger, M. Marsilli, *Tetrahedron Lett.* **34**, 3181 (1978)
59. L. Kier, L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis* (Research Studies Press, Letchworth, UK, 1986)
60. L. Pauling, *The Nature of Chemical Bond* (Cornell University Press, Ithaca, New York, 1939), pp. 2–60.
61. D.H. Rouvray, in *Chemical Applications of Graph Theory*, ed. by A.T. Balaban (Academic Press, London, 1976), pp. 180–181
62. N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FL, 1983; 2nd ed.; 1992), pp. 32–33

63. I. Gutman, O.E. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer-Verlag, Berlin, 1986)
64. E. Estrada, G. Patlewicz, *Croat. Chim. Acta.* **77**, 203 (2004)
65. D.J. Klein, *Int. Electron. J. Mol. Des.* **2**, 814 (2003)
66. N. Jacobson, *Basic Algebra I*, 2nd edn. (W.H. Freeman and Company, New York, 1985), pp. 343–361
67. K.F. Riley, M.P. Hobson, S.J. Vence, *Mathematical Methods for Physics and Engineering* (Ed. Cambridge University Press, 1998), pp. 228–236
68. E. Hernández, *Álgebra y Geometría* (Ed. Universidad Autonoma de Madrid, Madrid, Spain, 1987), pp. 521–544
69. J. de Burgos-Román, *Álgebra y Geometría Cartesiana* (Ed. McGraw-Hill Interamericana de España, 2^{da} Edición, 2000), pp. 208–246
70. J. de Burgos-Román, *Curso de Álgebra y Geometría* (Ed. Alambra Longman, Madrid, Spain, 1994), pp. 638–684
71. G. Werner, *Linear Algebra*, 4th edn. (Ed. Springer-Verlag, New York, 1981), pp. 261–288
72. M. Randić, *J. Math. Chem.* **7**, 155 (1991)
73. P.D. Walker, P.G. Mezey, *J. Am. Chem. Soc.* **115**, 12423 (1993)
74. L. Hall, C.T. Story, *J. Chem. Inf. Comput. Sci.* **36**, 1004 (1996)
75. J.D. Gough, L.H. Hall, *J. Chem. Inf. Comput. Sci.* **39**, 356 (1999)
76. Y. Marrero-Ponce, V. Romero. *TOMOCOMD software*. Central University of Las Villas (2002). TOMOCOMD (TOPOlogical MOlecular COmputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero-Ponce: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es
77. V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Comput. Sci.* **42**, 682 (2002)
78. R. Todeschini, P. Gramatica, *Perspect. Drug Dis. Des.* **9–11**, 355 (1998)
79. E. Estrada, L. Rodríguez, *J. Chem. Inf. Comput. Sci.* **39**, 1037 (1999)
80. M. Randić, *J. Mol. Struct. (Theochem)*. **233**, 45 (1991)
81. M. Randić, *Croat. Chim. Acta* **66**, 289 (1993)
82. M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, L. Naylor, *J. Chem. Inf. Comput. Sci.* **34**, 361 (1994)
83. M.V. Diudea, *J. Chem. Inf. Comput. Sci.* **36**, 535 (1996)
84. M.V. Diudea, O.M. Minailiuc, G. Katona, *Rev. Roum. Chim.* **42**, 239 (1997)
85. M. Randić, N. Trinajstić, *J. Mol. Struct. (Theochem)* **284**, 209 (1993)
86. M. Randić, N. Trinajstić, *J. Mol. Struct. (Theochem)* **300**, 551 (1993)
87. D.E. Needham, I.-C. Wei, P.G. Seybold, *J. Am. Chem. Soc.* **110**, 4186 (1988)
88. D.E. Goldberg, *Genetic Algorithms* (Addison Wesley, Reading, MA, 1989)
89. P. Willet, *Trends Biotechnol.* **13**, 516 (1995)
90. S.S. So, M. Karplus, *J. Med. Chem.* **39**, 1521 (1996)
91. S.S. So, M. Karplus, *J. Med. Chem.* **40**, 4347 (1997)
92. D. Rogers, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **34**, 854 (1994)
93. A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* **119**, 10509 (1997)
94. C.L. Senese, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **43**, 1297 (2003)
95. C.L. Senese, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **43**, 2180 (2003)
96. D.B. de Oliveira, A.C. Gaudio, *Quant. Struct.-Act. Relat.* **19**, 599 (2000)
97. S. Wold, L. Erikson, in *Statistical Validation of QSAR Results. Validation Tools; In Chemometric Methods in Molecular Design*, ed. by H. van de Waterbeemd (VCH Publishers, New York, 1995), pp. 309–318
98. Y. Marrero-Ponce, A. Meneses-Marcel, J.A. Catillo-Garit, Y. Machado-Tugores, J.A. Escario, A. Gómez-Barrio, D. Montero-Pereira, J.J. Nogal-Ruiz, V.J. Arán, A.R. Martínez-Fernández, F. Torrens, R. Rotondo, *Bioorg. Med. Chem.* **14**, 6502 (2006)