

## Bond-based 2D TOMOCOMD-CARDD approach for drug discovery: aiding decision-making in ‘in silico’ selection of new lead tyrosinase inhibitors

Yovani Marrero-Ponce · Mahmud Tareq Hassan Khan · Gerardo M. Casañola-Martín · Arjumand Ather · Mukhlis N. Sultankhodzhaev · Ramón García-Domenech · Francisco Torrens · Richard Rotondo

Received: 23 August 2006 / Accepted: 2 December 2006 / Published online: 28 February 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** In this paper, we present a new set of bond-level TOMOCOMD-CARDD molecular descriptors (MDs), the bond-based bilinear indices, based on a bilinear map similar to those defined in linear algebra. These novel MDs are used here in Quantitative Structure–Activity Relationship (QSAR) studies of tyrosinase inhibitors, for finding functions that discriminate between the tyrosinase inhibitor compounds and inactive ones. In total 14 models were obtained and the best two discriminant functions (Eqs. 32 and 33) shown globally good classification of 91.00% and 90.17%, respectively, in the training set. The test set had accuracies of 93.33% and 88.89% for the models

32 and 33, correspondingly. A simulated virtual screening was also carried out to prove the quality of the determined models. In a final step, the fitted models were used in the *biosilico* identification of new synthesized tetraketones, where a good agreement could be observed between the theoretical and experimental results. Four compounds of the novel bioactive chemicals discovered as tyrosinase inhibitors: **TK10** ( $IC_{50} = 2.09 \mu\text{M}$ ), **TK11** ( $IC_{50} = 2.61 \mu\text{M}$ ), **TK21** ( $IC_{50} = 2.06 \mu\text{M}$ ), **TK23** ( $IC_{50} = 3.19 \mu\text{M}$ ), showed more potent activity than L-mimose ( $IC_{50} = 3.68 \mu\text{M}$ ). Besides, for this study a heterogeneous database of tyrosinase inhibitors was collected,

Y. Marrero-Ponce address web: URL: [www.uv.es/yoma](http://www.uv.es/yoma)

**Electronic Supplementary Material** The online version of this article (doi:10.1007/s10822-006-9094-7) contains supplementary material, which is available to authorized users.

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

Y. Marrero-Ponce · G. M. Casañola-Martín  
Unit of Computer-Aided Molecular “Biosilico” Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara 54830, Cuba

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  
e-mail: [ymarrero77@yahoo.es](mailto:ymarrero77@yahoo.es)

M. T. H. Khan  
Pharmacology Research Lab., Faculty of Pharmaceutical Sciences, University of Science and Technology, Chittagong, Bangladesh

M. T. H. Khan  
Department of Pharmacology, Institute of Medical Biology, University of Tromsø, Tromsø 9037, Norway

G. M. Casañola-Martín  
Department of Biological Sciences, Faculty of Agricultural Sciences, University of Ciego de Avila, 69450 Ciego de Avila, Cuba

A. Ather  
The Norwegian Structural Biology Centre (NorStruct), University of Tromsø, Tromsø 9037, Norway

M. N. Sultankhodzhaev  
S. Yunusov Institute of Chemistry of Plant Substances, Academy of Sciences, Uzbekistan, Tashkent

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

and could be a useful tool for the scientist in the domain of tyrosinase enzyme researches. The current report could help to shed some clues in the identification of new chemicals that inhibits enzyme tyrosinase, for entering in the pipeline of drug discovery development.

**Keywords** *TOMOCOMD-CARDD* software · Non-stochastic and stochastic bond-based bilinear indices · LDA-based QSAR model · Tyrosinase inhibitor · TetraKetones · Virtual screening · *Biosilico* identification · Experimental results

## Introduction

Melanogenesis is a physiological process resulting in the synthesis of melanin pigments, which play a crucial protective role against skin photocarcinogenesis. In humans and other mammals, the biosynthesis of melanin takes place in a lineage of cells known as melanocytes, which contain the enzyme tyrosinase [1]. Tyrosinase (phenoloxidase) is known to be a key enzyme for melanin biosynthesis. This enzyme is mainly involved in the initial steps of the pathway which consist of the hydroxylation of the L-tyrosine (monophenolase activity) and the oxidation of the product of this reaction, the L-DOPA (diphenolase activity), to give rise to *o*-dopaquinone [2]. This *o*-quinone is transformed into melanins, followed by a series of divergent steps that give rise to a predominantly indolic pigment (eumelanin) and a closely related pigment containing benzothiazine subunits (phaeomelanin). The current view is that most human pigmentation involves a combination of these pathways giving rise to mixtures of varying composition [3, 4].

Many approaches are based on the use of analogue substrates for tyrosinase which are designed to maximize the generation of reactive orthoquinone oxidation products and increasing their diffusion range by preventing the spontaneous self-extinguishing cyclization reaction [5]. These, if released into the cytosol through the defective melanosomal membranes of malignant melanocytes, have the potential to react with vital cellular components and cause irreversible damage [6]. Therefore, inhibitors of tyrosinase should be useful as therapeutic agents for the treatment of melanin hyperpigmentation and cosmetic materials for whitening after sunburn [7, 8].

On other hand, the computational methods have become in a suitable alternative to the drug design,

and have recently applied to QSAR studies of tyrosinase inhibitors [9–11], using congeneric or heterogeneous dataset of compounds. In this sense QSAR methods can reduce the costly failures of drug candidates in clinical trials by filtering virtual libraries of chemicals.

One of our research group has carried out QSAR/QSPR studies related to chemical, physicochemical and biological properties of different chemicals and drugs [12–16], including studies in nucleic acid–drug interactions [17, 18] and discovery of antimalarial compounds [19]. The ‘in house’ TOpologicalMOlecular COMputer Design-Computer Aided‘Rational’ Drug Design (*TOMOCOMD-CARDD*) software [20] a novel computer-aided molecular design scheme, based in the graph theory and linear algebra; has been used to develop this entire works and many others.

Here we propose a new set of molecular descriptors (MDs) namely *non-stochastic and stochastic bond-based bilinear indices*, its application to discriminate tyrosinase inhibitor compounds (actives) from inactive ones using QSAR models, is shown. Furthermore a virtual screening is carried out with a small library of chemicals and as a final point we present the *in silico* identification, synthesis and *in vitro* assays of a new set of tetraketones, a procedure that can arise the potentialities of these new MDs into a real world application, that could help to speed up the discovery of new lead compounds to treat the hyperpigmentation and skin disorders.

## Theoretical framework

The basis of the extension of bilinear indices that will be given here is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature [21, 22], and rediscovered by Estrada as an important source of new MDs [23–28]. In this section, we first will define the nomenclature to be used in this work, then the atom-based molecular vector ( $\bar{x}$ ) will be redefined for bond characterization using the same approach as previously reported, and finally some new definition of bond-based non-stochastic and stochastic bilinear indices will be given.

Background in edge-adjacency matrix and new edge-relations: stochastic edge-adjacency matrix

Let  $G = (V, E)$  be a simple graph, with  $V = \{v_1, v_2, \dots, v_n\}$  and  $E = \{e_1, e_2, \dots, e_m\}$  being the

vertex- and edge-sets of  $G$ , respectively. Then  $G$  represents a molecular graph having  $n$  vertices and  $m$  edge (bonds). The edge-adjacency matrix  $\mathbf{E}$  of  $G$  (likewise called bond-adjacency matrix,  $\mathbf{B}$ ) is a square and symmetric matrix whose elements  $e_{ij}$  are 1 if and only if edge  $i$  is adjacent to edge  $j$  [25, 28–30]. Two edges are adjacent if they are incidental to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the  $i$ th row (or column) of  $\mathbf{E}$  is named the edge-degree of bond  $i$ ,  $\delta(e_i)$  [23, 26, 27, 29, 30].

By using the edge (bond)-adjacency relationships we can find other new relation for a molecular graph that will be introduced here. The  $k$ th stochastic edge-adjacency matrix,  $\mathbf{ES}^k$  can be obtained directly from  $\mathbf{E}^k$ . Here,  $\mathbf{ES}^k = [{}^k es_{ij}]$  is a square table of order  $m$  ( $m$  = number of bonds) and the elements  ${}^k es_{ij}$  are defined as follows:

$${}^k es_{ij} = \frac{{}^k e_{ij}}{{}^k \text{SUM}(E^k)_i} = \frac{{}^k e_{ij}}{{}^k \delta(e)_i} \quad (1)$$

where,  ${}^k e_{ij}$  are the elements of the  $k$ th power of  $\mathbf{E}$  and the SUM of the  $i$ th row of  $\mathbf{E}^k$  are named the  $k$ -order edge degree of bond  $i$ ,  ${}^k \delta(e)_i$ . Note that the matrix  $\mathbf{ES}^k$  in Eq. 1 has the property that the sum of the elements in each row is 1. An  $m \times m$  matrix with nonnegative entries having this property is called a *stochastic matrix* [31].

#### Chemical information and bond-based molecular vector

The atom-based molecular vector ( $\bar{x}$ ) used to represent small-to-medium size organic chemicals has been explained in some detail elsewhere [12–14, 16, 17, 32–44]. In a manner parallel to the development of  $\bar{x}$ , we present the expansion of the bond-based molecular vector ( $\bar{w}$ ). The components ( $w$ ) of  $\bar{w}$  are numeric values, which represent a certain standard bond property (bond-label). That is to say, these weights correspond to different bond properties for organic molecules. Thus, a molecule having 5, 10, 15, ...,  $m$  bonds can be represented by means of vectors, with 5, 10, 15, ...,  $m$  components, belonging to the spaces  $\mathfrak{R}^5$ ,  $\mathfrak{R}^{10}$ ,  $\mathfrak{R}^{15}$ , ...,  $\mathfrak{R}^m$ , respectively; where  $m$  is the dimension of the real sets ( $\mathfrak{R}^m$ ). This approach allows us encoding organic molecules such as 3-hydroxy-2-butenitrile through the molecular vector  $\bar{w} = [w_{\text{Csp3-Csp2}}, w_{\text{Csp2=Csp2}}, w_{\text{Csp2-Osp3}}, w_{\text{H-Osp3}}, w_{\text{Csp2-Csp}}, w_{\text{Csp=Nsp}}]$ . This vector belongs to the product space  $\mathfrak{R}^6$ .

These properties characterize each kind of bond (and bond-types) within the molecule. Diverse kinds of

bond weights ( $w$ ) can be used in order to codify information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distance [45–48], standard bond dipole [45–48] or even mathematical expressions involving atomic weights such as atomic log  $P$  [49], surface contributions of polar atoms [50], atomic molar refractivity [51], atomic hybrid polarizabilities [52], and Gasteiger–Marsilli atomic charge [53], atomic electronegativity in Pauling scale [54] and so on. Here, we characterized each bond with the following parameter:

$$w = x_i/\delta_i + x_j/\delta_j \quad (2)$$

which characterizes each bond. In this expression  $x_i$  can be any standard weight of the atom  $i$  bonded with atom  $j$ .  $\delta_i$  is the vertex (atom) degree of atom  $i$ . The use of each scale (bond property) defines alternative molecular vectors,  $\bar{w}$ .

The chemical information can also be codify by means of two different molecular vectors, for instance,  $\bar{w} = [w_1, \dots, w_n]$  and  $\bar{u} = [u_1, \dots, u_n]$ ; then different combinations of molecular vectors ( $\bar{w} \neq \bar{u}$ ) are possible when a weighting scheme is used. In the present report, we characterized each bond with mathematical expressions involving the following parameters: atomic masses ( $M$ ) [55], the van der Waals volumes ( $V$ ) [55], the atomic polarizabilities ( $P$ ) [55], and atomic electronegativity ( $E$ ) in Mulliken scale [55]. The values of these atomic labels are shown in Table 1. From this weighting scheme, six (or 12 if  $\bar{w}_M-\bar{u}_V \neq \bar{w}_V-\bar{u}_M$ ) combinations (pairs) of molecular vectors ( $\bar{w}, \bar{u}; \bar{w} \neq \bar{u}$ ) can be computed,  $\bar{w}_M-\bar{u}_V$ ,  $\bar{w}_M-\bar{u}_P$ ,  $\bar{w}_M-\bar{u}_K$ ,  $\bar{w}_V-\bar{u}_P$ ,  $\bar{w}_V-\bar{u}_K$ , and  $\bar{w}_P-\bar{u}_K$ . Here, we used the symbols  $\bar{w}_X-\bar{u}_Z$ , where the subscripts  $X$  and  $Z$  mean two mathematical expressions involving atomic properties from our weighting scheme and a hyphen (-) expresses the combination (pair) of two selected bond-label chemical properties. In order to illustrate this we will consider this in an example describe in other section of this work.

#### Definition of mathematical bilinear forms

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 [58–60]. That is, this function satisfies the following axioms for any scalar  $\alpha$  and any choice of vectors  $\bar{v}, \bar{w}, \bar{v}_1, \bar{v}_2, \bar{w}_1$  and  $\bar{w}_2$ .

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

**Table 1** Values of the atom weights used for linear indices calculation [54–57]

ID	Atomic mass	VdW <sup>a</sup> Volume (Å <sup>3</sup> )	Mulliken electronegativity	Polarizability (Å <sup>3</sup> )
H	1.01	6.709	2.592	0.667
B	10.81	17.875	2.275	3.030
C	12.01	22.449	2.746	1.760
N	14.01	15.599	3.194	1.100
O	16.00	11.494	3.654	0.802
F	19.00	9.203	4.000	0.557
Al	26.98	36.511	1.714	6.800
Si	28.09	31.976	2.138	5.380
P	30.97	26.522	2.515	3.630
S	32.07	24.429	2.957	2.900
Cl	35.45	23.228	3.475	2.180
Fe	55.85	41.052	2.000	8.400
Co	58.93	35.041	2.000	7.500
Ni	58.69	17.157	2.000	6.800
Cu	63.55	11.494	2.033	6.100
Zn	65.39	38.351	2.223	7.100
Br	79.90	31.059	3.219	3.050
Sn	118.71	45.830	2.298	7.700
I	126.90	38.792	2.778	5.350

<sup>a</sup> VdW: van der Waals

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  $\bar{w}$  and  $\bar{u}$  of  $V$ , where  $(w^1, w^2, \dots, w^n) \in \mathfrak{R}^n$  and  $(u^1, u^2, \dots, u^n) \in \mathfrak{R}^n$  are the coordinates of the vectors  $\bar{x}$  and  $\bar{u}$ , respectively. That is to say,

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

and,

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

Subsequently,

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

if we take the  $a_{ij}$  as the  $n \times n$  scalars  $b(\bar{e}_i, \bar{e}_j)$ , That is,  $a_{ij} = b(\bar{e}_i, \bar{e}_j)$ , to  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, n$  (6)

Then,

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

As it can be seen, the defined equation for  $b$  may be written as the single matrix equation (see Eq. 7), where  $[U]$  is a column vector (an  $n \times 1$  matrix) of the coordinates of  $\bar{u}$  in a basis set of  $\mathfrak{R}^n$ , and  $[W]^T$  (a  $1 \times n$  matrix) is the transpose of  $[W]$ , where  $[W]$  is a column vector (an  $n \times 1$  matrix) of the coordinates of  $\bar{w}$  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  $b(\bar{w}, \bar{u}) = b(\bar{u}, \bar{w}), \forall \bar{w}, \bar{u} \in V$  [58–60] Then,

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

The total non-stochastic and stochastic bond-based bilinear indices

If a molecule consists of  $m$  bonds (*vector of  $\mathfrak{R}^m$* ), then the  $k$ th total bilinear indices are calculated as bilinear maps (bilinear form) in  $\mathfrak{R}^m$  in canonical basis set. Specifically, the  $k$ th total non-stochastic and stochastic bond bilinear indices,  $b_k(\bar{w}, \bar{u})$  and  ${}^s b_k(\bar{w}, \bar{u})$ , are computed from these  $k$ th non-stochastic and stochastic edge adjacency matrices,  $\mathbf{E}^k$  and  $\mathbf{ES}^k$ , as shown in Eqs. 9 and 10, correspondingly:

$$b_k(\bar{w}, \bar{u}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{ij} w^i u^j = [W]^t \mathbf{E}^k [U] \quad (9)$$

$${}^s b_k(\bar{w}, \bar{u}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e s_{ij} w^i u^j = [W]^t \mathbf{ES}^k [U] \quad (10)$$

where,  $m$  is the number of bonds of the molecule, and  $w^1, \dots, w^m$  and  $u^1, \dots, u^m$  are the coordinates of the bond-based molecular vectors  $\bar{w}$  and  $\bar{u}$  in a canonical basis set of  $\mathfrak{R}^m$ . Therefore, if we used the canonical basis set, the coordinates  $[(w^1, \dots, w^m)$  and  $(u^1, \dots, u^m)]$  of any molecular vectors ( $\bar{w}$  and  $\bar{u}$ ) coincide with the components of those vectors  $[(w_1, \dots, w_n)$  and  $(u_1, \dots, u_n)]$  [28, 45, 46]. For that reason, those coordinates can be considered as weights (bond-labels) of the edge of the molecular graph. The coefficients  ${}^k e_{ij}$  and  ${}^k e s_{ij}$  are the elements of the  $k$ th power of the matrix  $\mathbf{E}(G)$  and  $\mathbf{ES}(G)$ , correspondingly, of the molecular pseudograph. The defining Eqs. 9 and 10 for  $b_k(\bar{w}, \bar{u})$  and  ${}^s b_k(\bar{w}, \bar{u})$ , respectively, may be also written as the single matrix equation (see Eqs. 9 and 10), where  $[U]$  is a column vector (an  $n \times 1$  matrix) of the coordinates of  $\bar{u}$  in the canonical basis set of  $\mathfrak{R}^n$ , and  $[W]^t$  is the transpose of  $[W]$ , where  $[W]$  is a column vector (an  $n \times 1$  matrix) of the

coordinates of  $\bar{w}$  in the canonical basis of  $\mathfrak{R}^n$ . Here,  $\mathbf{E}^k$  and  $\mathbf{ES}^k$  denote the matrices of bilinear 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,  $M$  and  $V$  are used as weights to compute these MDs, two different sets of stochastic bilinear indices,  ${}^{M-Vs}\mathbf{b}_k^H(\bar{w}, \bar{u})$  and  ${}^{V-Ms}\mathbf{b}_k^H(\bar{w}, \bar{u})$  [because  $\bar{w}_M-\bar{u}_V \neq \bar{w}_V-\bar{u}_M$ ] can be obtained and only one group of non-stochastic bilinear indices ( ${}^{M-Vs}\mathbf{b}_k^H(\bar{w}, \bar{u}) = {}^{V-Ms}\mathbf{b}_k^H(\bar{w}, \bar{u})$ ) because in this case  $\bar{w}_M-\bar{u}_V = \bar{w}_V-\bar{u}_M$ ) can be calculated.

The local non-stochastic and stochastic bond-based bilinear indices

Finally, in addition to total bond-based quadratic indices, computed for the whole molecule, a local-fragment (bond and bond-type) formalism can be developed. These descriptors are termed local non-stochastic and stochastic bilinear indices,  $b_{kL}(\bar{w}, \bar{u})$  and  ${}^s b_{kL}(\bar{w}, \bar{u})$ , respectively. The definition of these descriptors is as follows:

$$b_{kL}(\bar{w}, \bar{u}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{ijL} w^i u^j = [W]^t \mathbf{E}_L^k [U] \quad (11)$$

$${}^s b_{kL}(\bar{w}, \bar{u}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{sijL} w^i u^j = [W]^t \mathbf{ES}_L^k [U] \quad (12)$$

where,  $m$  is the number of bonds and  ${}^k e_{ijL}$  [ ${}^k e_{sijL}$ ] is the  $k$ th element of the row “ $i$ ” and column “ $j$ ” of the local matrix  $\mathbf{E}_L^k$  [ $\mathbf{ES}_L^k$ ]. This matrix is extracted from the  $k$ th  $[\mathbf{ES}^k]$  matrix and contains information referred to the edges (bonds) of the specific molecular fragments and also of the molecular environment in  $k$  steps. The matrix  $\mathbf{E}_L^k$  [ $\mathbf{ES}_L^k$ ] with elements  ${}^k e_{ijL}$  [ ${}^k e_{sijL}$ ] is defined as follows:

$$\begin{aligned} &{}^k e_{ijL} [{}^k e_{sijL}] = {}^k e_{ij} [{}^k e_{sij}] \text{ if both} \\ &e_i \text{ and } e_j \text{ are edges} \\ &(\text{bonds}) \text{ contained within the molecular fragment} \\ &= 1/2^k e_{ij} [{}^k e_{sij}] \text{ if } e_i \text{ and } e_j \text{ are edges (bonds)} \\ &\text{contained within the molecular fragment but not both} \\ &= 0 \text{ otherwise} \end{aligned} \quad (13)$$

It is important to highlight that the scheme above follows the spirit of a Mulliken population analysis [61]. It should be remarked 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 matrices  $\mathbf{E}^k$  [ $\mathbf{ES}^k$ ] can be correspondingly partitioned into  $Z$  local matrices  $\mathbf{E}_L^k$  [ $\mathbf{ES}_L^k$ ],  $L = 1, \dots, Z$ , and the  $k$ th power of matrix  $\mathbf{E}$  [ $\mathbf{ES}$ ] is exactly the sum of the  $k$ th power of the local  $Z$  matrices. In this way, the total (both non-stochastic and stochastic) bond-based bilinear indices are the sum of the non-stochastic and stochastic bond-based bilinear indices, respectively, of the  $Z$  molecular fragments:

$$b_k(\bar{w}, \bar{u}) = \sum_{L=1}^Z b_{kL}(\bar{w}, \bar{u}) \quad (14)$$

$${}^s b_k(\bar{w}, \bar{u}) = \sum_{L=1}^Z {}^s b_{kL}(\bar{w}, \bar{u}) \quad (15)$$

Bond and bond-type bilinear fingerprints are specific cases of local bond-based bilinear indices. The  $k$ th bond-type bilinear indices of the edge-adjacency matrix are calculated by summing up the  $k$ th bond bilinear indices for all bonds of the same type in the molecule. That is to say, this extension of the bond bilinear index is similar to group additive schemes, in which an index appears for each bond type in the molecule together with its contribution based of the bond bilinear index.

In the bond-type bilinear indices formalism, each bond in the molecule is classified into a bond-type (fragment). In this sense, bonds may be classified into bond types in terms of the characteristics of the two atoms that define the bond. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the  $k$ th fragment (bond-type) quadratic indices provide much useful information. Thus, the development of the bond-type bilinear indices description provides the basis for application to a wider range of biological problems in which the local formalism is applicable without the need for superposition or a closely related set of structures.

It is useful to perform a calculation on a molecule to illustrate the steps in the procedure. For this, in the next section we depict a pictorial representation of the calculus of the non-stochastic and stochastic bilinear indices of the bond matrix (both total and local) using a simple chemical example.

### Sample calculation

The bilinear indices of the bond matrix are calculated in the following way. Considering the molecule of 3-hydroxy-2-butenitrile as a simple example, we

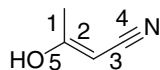
have the following labeled molecular graph and bond-based adjacency matrices (**E** and **ES**). The second ( $k = 2$ ) and third ( $k = 3$ ) power of these matrices and bond-based molecular vector,  $\bar{w}$ , are also given:

The molecule contains five localized bonds (cor-

$$u_4 = y_C/4 + y_N/3 = 12.01/4 + 14.01/3 = 7.6725$$

$$u_5 = y_C/4 + y_O/1 = 12.01/4 + 16.00/1 = 19.0025$$

and therefore,  $\bar{u} = [15.0125, 7.005833, 7.005833, 7.6725, 19.0025]$ .



$$E^0 = ES^0 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \quad E^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 10 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad E^2 = \begin{bmatrix} 2 & 1 & 1 & 0 & 1 \\ 1 & 3 & 0 & 1 & 1 \\ 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix} \quad E^3 = \begin{bmatrix} 2 & 4 & 1 & 1 & 3 \\ 4 & 2 & 4 & 0 & 4 \\ 1 & 4 & 0 & 2 & 1 \\ 1 & 0 & 2 & 0 & 1 \\ 3 & 4 & 1 & 1 & 2 \end{bmatrix}$$

$$ES^1 = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix} \quad ES^2 = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0 & 0.2 \\ 0.16 & 0.5 & 0 & 0.16 & 0.16 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0.2 & 0.2 & 0.2 & 0 & 0.4 \end{bmatrix} \quad ES^3 = \begin{bmatrix} 0.18 & 0.36 & 0.090 & 0.090 & 0.27 \\ 0.28 & 0.14 & 0.28 & 0 & 0.28 \\ 0.12 & 0.5 & 0 & 0.25 & 0.12 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0.27 & 0.36 & 0.090 & 0.090 & 0.18 \end{bmatrix}$$

responding to five edges in the H-suppressed molecular graph). To these we will associate the five “bond orbitals”  $w_1, w_2, w_3, w_4$ , and  $w_5$ . Thus,  $\bar{w} = [w_1, w_2, w_3, w_4, w_5] = [w_{(C-C)}, w_{(C=C)}, w_{(C-C)}, w_{(C=N)}, w_{(C-O)}]$  and each “bond orbital” can be computed by Eq. 2 using, for instance, the atomic electronegativity in Pauling scale ( $x$ ) [54] as atomic weight (atom-label):

$$w_1 = x_C/1 + x_C/4 = 2.55/1 + 2.55/4 = 3.1875$$

$$w_2 = x_C/4 + x_C/3 = 2.55/4 + 2.55/3 = 1.4875$$

$$w_3 = x_C/3 + x_C/4 = 2.55/3 + 2.55/4 = 1.4875$$

$$w_4 = x_C/4 + x_N/3 = 2.55/4 + 3.04/3 = 1.650833$$

$$w_5 = x_C/4 + x_O/1 = 2.55/4 + 3.44/1 = 4.0775$$

and therefore,  $\bar{w} = [3.1875, 1.4875, 1.4875, 1.650833, 4.0775]$ .

Besides other vector,  $\bar{u}$  must be calculated in the same way that  $\bar{w}$ , but using other property, for example the atomic masses [55] as atomic weight (atom-label):

$$u_1 = y_C/1 + y_C/4 = 12.01/1 + 12.01/4 = 15.0125$$

$$u_2 = y_C/4 + y_C/3 = 12.01/4 + 12.01/3 = 7.005833$$

$$u_3 = y_C/3 + y_C/4 = 12.01/3 + 12.01/4 = 7.005833$$

Each non-stochastic and stochastic total bilinear index will have the form:

$$b_k(\bar{w}, \bar{u}) = {}^k e_{11} w^1 u^1 + {}^k e_{21} w^1 u^2 + {}^k e_{31} w^1 u^3 + {}^k e_{41} w^1 u^4 + {}^k e_{51} w^1 u^5 + {}^k e_{12} w^1 u^2 + {}^k e_{22} w^2 u^2 + {}^k e_{32} w^2 u^3 + {}^k e_{42} w^2 u^4 + {}^k e_{52} w^2 u^5 + {}^k e_{13} w^1 u^3 + {}^k e_{23} w^2 u^3 + {}^k e_{33} w^3 u^3 + {}^k e_{43} w^3 u^4 + {}^k e_{53} w^3 u^5 + {}^k e_{14} w^1 u^4 + {}^k e_{24} w^2 u^4 + {}^k e_{34} w^3 u^4 + {}^k e_{44} w^4 u^4 + {}^k e_{54} w^4 u^5 + {}^k e_{15} w^1 u^5 + {}^k e_{25} w^2 u^5 + {}^k e_{35} w^3 u^5 + {}^k e_{45} w^4 u^5 + {}^k e_{55} w^5 u^5 = \sum_{(i)} {}^k e_{ii} w^i u^i + 2 \sum_{(i,j)} {}^k e_{ij} w^i u^j \quad (16)$$

$$s b_k(\bar{w}, \bar{u}) = {}^k e_{s11} w^1 u^1 + {}^k e_{s21} w^1 u^2 + {}^k e_{s31} w^1 u^3 + {}^k e_{s41} w^1 u^4 + {}^k e_{s51} w^1 u^5 + {}^k e_{s12} w^1 u^2 + {}^k e_{s22} w^2 u^2 + {}^k e_{s32} w^2 u^3 + {}^k e_{s42} w^2 u^4 + {}^k e_{s52} w^2 u^5 + {}^k e_{s13} w^1 u^3 + {}^k e_{s23} w^2 u^3 + {}^k e_{s33} w^3 u^3 + {}^k e_{s43} w^3 u^4 + {}^k e_{s53} w^3 u^5 + {}^k e_{s14} w^1 u^4 + {}^k e_{s24} w^2 u^4 + {}^k e_{s34} w^3 u^4 + {}^k e_{s44} w^4 u^4 + {}^k e_{s54} w^4 u^5 + {}^k e_{s15} w^1 u^5 + {}^k e_{s25} w^2 u^5 + {}^k e_{s35} w^3 u^5 + {}^k e_{s45} w^4 u^5 + {}^k e_{s55} w^5 u^5 = \sum_{(i)} {}^k e_{sii} w^i u^i + 2 \sum_{(i,j)} {}^k e_{sij} w^i u^j \quad (17)$$

The  ${}^k e_{ii}$ 's and  ${}^k e_{ij}$ 's can be considered a measure of the attraction of an electron for a bond in the  $k$  step. The  ${}^k e_{ij}$ 's and  ${}^k e_{ji}$ 's are the terms of interaction between two bonds in the  $k$  step. The  ${}^k e_{ij}$ 's =  ${}^k e_{ji}$ 's are equal by symmetry (non-oriented molecular graph). However,  ${}^k e_{ij} \neq {}^k e_{ji}$ . This is a logical result because the  $k$ th  $e_{ij}$  elements are the transition probabilities with the 'electrons' moving from bond  $i$  to  $j$  at the discrete time periods  $t_k$  and it should be different in both senses. This result is in total agreement if the electronegativity of the two atom types in the bonds are taken into account.

In this way,  $\mathbf{E}^k$  and  $\mathbf{ES}^k$  can be seen as graph-theoretic electronic-structure models [62]. In fact, quantum chemistry starts from the fact a molecule is made up of electrons and nuclei. The distinction here between bonded and non-bonded atoms is difficult to justify. Any two nuclei of a molecule interact directly and indirectly through the electrons present in the molecule. Only the intensity of this interaction varies in going from one pair of nuclei to another. In this sense, the electron in an arbitrary bond  $i$  can move (step-by-step) to other bonds at different discrete time periods  $t_k$  ( $k = 0, 1, 2, 3, \dots$ ) through the chemical-bonding network. That is to say, the  $\mathbf{E}^1$  and  $\mathbf{ES}^1$  matrices consider the valence-bond electrons in one step and their power ( $k = 0, 1, 2, 3, \dots$ ) can be considering as an interacting-electron chemical-network model in  $k$  step. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas [62].

On the other hand, the  $k$ th ( $k = 0-3$ ) non-stochastic total quadratic indices can be expressed as the sum of the local (bond) quadratic indices for this molecule as follows:

$$\begin{aligned} q_0(\bar{w}, \bar{u}) &= q_{0L}(\bar{w}, \bar{u}_1) + q_{0L}(\bar{w}, \bar{u}_2) + q_{0L}(\bar{w}, \bar{u}_3) \\ &+ q_{0L}(\bar{w}, \bar{u}_4) + q_{0L}(\bar{w}, \bar{u}_5) = 47,85234 \\ &+ 10,42118 + 10,42118 + 12,66602 \\ &+ 77,48269 = 158,8434 \end{aligned}$$

$$\begin{aligned} q_1(\bar{w}, \bar{u}) &= q_{1L}(\bar{w}, \bar{u}_1) + q_{1L}(\bar{w}, \bar{u}_2) + q_{1L}(\bar{w}, \bar{u}_3) \\ &+ q_{1L}(\bar{w}, \bar{u}_4) + q_{1L}(\bar{w}, \bar{u}_5) = 83,22306 \\ &+ 61,16852 + 21,91033 + 11,48915 \\ &+ 89,30822 = 267,09929 \end{aligned}$$

$$\begin{aligned} q_2(\bar{w}, \bar{u}) &= q_{2L}(\bar{w}, \bar{u}_1) + q_{2L}(\bar{w}, \bar{u}_2) + q_{2L}(\bar{w}, \bar{u}_3) \\ &+ q_{2L}(\bar{w}, \bar{u}_4) + q_{2L}(\bar{w}, \bar{u}_5) = 201,2588 \\ &+ 93,50003 + 71,5897 + 24,15517 \\ &+ 272,6899 = 663,1936 \end{aligned}$$

$$\begin{aligned} q_3(\bar{w}, \bar{u}) &= q_{3L}(\bar{w}, \bar{u}_1) + q_{3L}(\bar{w}, \bar{u}_2) + q_{3L}(\bar{w}, \bar{u}_3) \\ &+ q_{3L}(\bar{w}, \bar{u}_4) + q_{3L}(\bar{w}, \bar{u}_5) = 414,6557 \\ &+ 265,5164 + 115,4104 + 78,92521 \\ &+ 511,0498 = 1385,5575 \end{aligned}$$

The terms in the summations for calculating the total quadratic indices are the so-called local (bond) quadratic indices. We have written these terms in the consecutive order of the bond labels in the graph. For instance, the non-stochastic bond quadratic indices of order 0, 1, 2 and 3 for the bond labeled as 1 are 47.85234, 83.22306, 201.2588 and 414.6557, respectively.

The  $k$ th total stochastic quadratic indices values are also the sum of the  $k$ th local (bond) stochastic quadratic indices values for all bonds in the molecule:

$$\begin{aligned} {}^s q_0(\bar{w}, \bar{u}) &= {}^s q_{0L}(\bar{w}, \bar{u}_1) + {}^s q_{0L}(\bar{w}, \bar{u}_2) + {}^s q_{0L}(\bar{w}, \bar{u}_3) \\ &+ {}^s q_{0L}(\bar{w}, \bar{u}_4) + {}^s q_{0L}(\bar{w}, \bar{u}_5) \\ &= 47,85234 + 10,42118 + 10,42118 \\ &+ 12,66602 + 77,48269 = 158,8434 \end{aligned}$$

$$\begin{aligned} {}^s q_1(\bar{w}, \bar{u}) &= {}^s q_{1L}(\bar{w}, \bar{u}_1) + {}^s q_{1L}(\bar{w}, \bar{u}_2) + {}^s q_{1L}(\bar{w}, \bar{u}_3) \\ &+ {}^s q_{1L}(\bar{w}, \bar{u}_4) + {}^s q_{1L}(\bar{w}, \bar{u}_5) \\ &= 39,75061 + 25,47438 + 12,93994 \\ &+ 8,597788 + 42,27359 = 129,0363 \end{aligned}$$

$$\begin{aligned} {}^s q_2(\bar{w}, \bar{u}) &= {}^s q_{2L}(\bar{w}, \bar{u}_1) + {}^s q_{2L}(\bar{w}, \bar{u}_2) + {}^s q_{2L}(\bar{w}, \bar{u}_3) \\ &+ {}^s q_{2L}(\bar{w}, \bar{u}_4) + {}^s q_{2L}(\bar{w}, \bar{u}_5) \\ &= 40,43786 + 18,32877 + 16,63249 \\ &+ 10,15001 + 54,77602 = 140.3252 \end{aligned}$$

$$\begin{aligned} {}^s q_3(\bar{w}, \bar{u}) &= {}^s q_{3L}(\bar{w}, \bar{u}_1) + {}^s q_{3L}(\bar{w}, \bar{u}_2) + {}^s q_{3L}(\bar{w}, \bar{u}_3) \\ &+ {}^s q_{3L}(\bar{w}, \bar{u}_4) + {}^s q_{3L}(\bar{w}, \bar{u}_5) \\ &= 39,15194 + 22,05334 + 13,87389 \\ &+ 13,8189 + 48,32158 = 137,2196 \end{aligned}$$

## Material and methods

### TOMOCOMD-CARDD approach

The total and local (bond-type) bond-based bilinear indices were calculate by the interactive program for molecular design and bioinformatic research *TOMOCOMD-CARDD* [20]. The software was developed based on a user-friendly philosophy. That is to say, this computer graphics software shows a great efficiency of interaction with the user, without *prior* knowledge of programming skills (e.g. practicing pharmaceutical and organic chemist, teacher, university student, and so on). CARDD subprogram allows drawing the structures (drawing mode) and calculating 2D (topologic), 3D-chiral (2.5D) and 3D (geo-

metric and topographic) non-stochastic and stochastic MDs (calculation mode).

The main steps for the application of this method in QSAR/QSPR and for drug design can be briefly summarized as follows:

1. Drawing of the molecular pseudographs for each molecule in the data set, using the drawing mode.
2. Use appropriate weights in order to differentiate the molecular atoms. The weights used in this work are those previously proposed for the calculation of the DRAGON descriptors [55–57], i.e., atomic mass ( $M$ ), atomic polarizability ( $P$ ), atomic Mulliken electronegativity ( $K$ ) plus the van der Waals atomic volume ( $V$ ). The values of these atomic labels are shown in Table 1 [54–57].
3. Computation of the total and local (bond and bond-type) bond bilinear indices of the bond adjacency matrix 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 the columns correspond to the bond-based (both total and local) bilinear maps or other MD family implemented in this program.
4. Development of a QSPR/QSAR equation by using several multivariate analytical techniques, for instance, linear discrimination analysis. That is to say, one can find a quantitative relationship between an activity  $\mathbf{A}$  and the bond-based bilinear fingerprints having, for instance, the following appearance:

$$\mathbf{A} = a_0 \mathbf{b}_0(\bar{w}, \bar{u}) + a_{11}(\bar{w}, \bar{u}) + a_2 \mathbf{b}_2(\bar{w}, \bar{u}) + \dots + a_k \mathbf{b}_k(\bar{w}, \bar{u}) + c \quad (18)$$

where  $\mathbf{A}$  is the measured activity,  $\mathbf{b}_k(\bar{w}, \bar{u})$  are the  $k$ th non-stochastic total bond-based bilinear indices, and the  $a_k$ 's are the coefficients obtained by the linear regression analysis.

5. Test of the robustness and predictive power of the QSPR/QSAR equation by using internal [leave-one-out (LOO)] and external (using a test set and an external predicting set) validation techniques.

The bond-based *TOMOCOMD-CARDD* descriptors computed in this study were the following:

- (1)  $k$ th ( $k = 15$ ) total non-stochastic bond-based bilinear indices not considering and considering H-atoms in the molecular graph ( $G$ ) [ $\mathbf{b}_b(\bar{w}, \bar{u})$  and  $\mathbf{b}_b^H(\bar{w}, \bar{u})$ , respectively].

- (2)  $k$ th ( $k = 15$ ) total stochastic bond-based bilinear indices not considering and considering H-atoms in the molecular graph ( $G$ ) [ $\mathbf{b}_b(\bar{w}, \bar{u})$  and  $\mathbf{b}_b^H(\bar{w}, \bar{u})$ , respectively].
- (3)  $k$ th ( $k = 15$ ) bond-type local (group = heteroatoms: S, N, O) non-stochastic bilinear indices not considering and considering H-atoms in the molecular graph ( $G$ ) [ $\mathbf{b}_{kL}(\bar{w}_E, \bar{u}_E)$  and  $\mathbf{b}_{kL}^H(\bar{w}_E, \bar{u}_E)$ , correspondingly]. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.
- (4)  $k$ th ( $k = 15$ ) bond-type local (group = heteroatoms: S, N, O) stochastic bilinear indices not considering and considering H-atoms in the molecular graph ( $G$ ) [ $\mathbf{b}_{kL}(\bar{w}_E, \bar{u}_E)$ , and  $\mathbf{b}_{kL}^H(\bar{w}_E, \bar{u}_E)$ , correspondingly]. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.

#### Database construction

The database collected to our study of tyrosinase inhibitory activity consists of 685 compounds in total. The active compounds inside this set were of 246, having reported activity against the enzyme tyrosinase. The rest, 412 organic-chemicals were chosen as inactive compounds. In both cases (active and inactive ones) we consider the structural molecular variability as important goal to assure the quality of our QSAR study.

In the case of tyrosinase inhibitor compounds (actives) many different subsystems were included. An example of the most representative tyrosinase reference drugs is illustrates in Fig. 1, together with some tyrosinase inhibitors of different families.

The names of compounds in the active database together with their experimental data taken from the literature are shown in Table 1 of Supporting Information. In the same way, we depict in Table 2 (Supporting Information) the molecular structures of these 246 tyrosinase inhibitors. This dataset provides a helpful tool for scientific research in many chemistry fields related with the tyrosinase enzyme and its inhibitors.

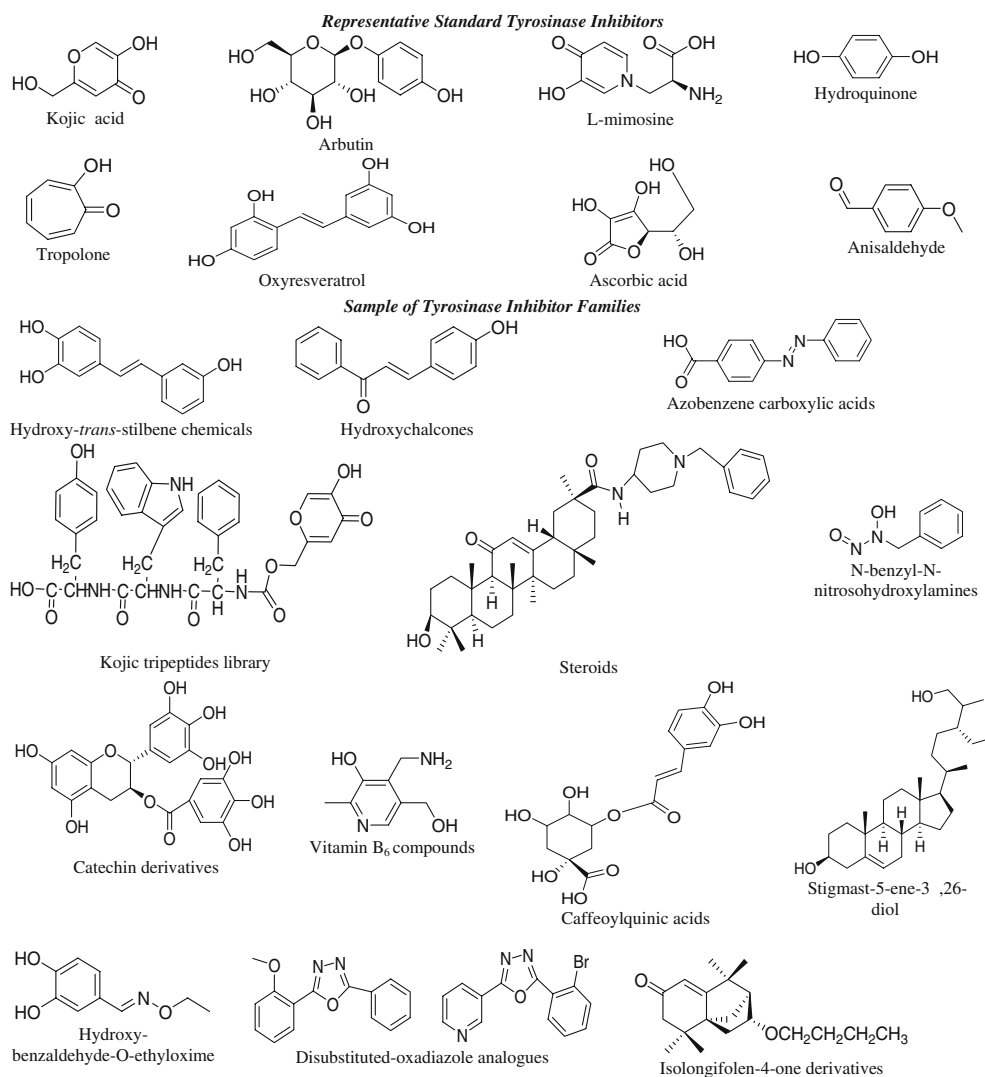
By other way, the rest 412 compounds having different pharmacological uses were selected for the inactive set. All these chemicals were taken from the Negwer Handbook [63], where their names, synonyms and structural formulas can be found.

#### Statistical techniques

The STATISTICA software [64] was used to develop the different statistical methods used in this report. In



**Fig. 1** Random, but not exhaustive, sample of the molecular families of tyrosinase inhibitors studied here and some reference drugs



first place we employed the cluster analysis as a method that recognizes similarities among cases and it contains them according to these criteria [65]. In our case *k*-MCA (*k*-means cluster analysis) and *k*-NNCA (*k*-nearest neighbors cluster analysis) algorithms were used to design the training and prediction series [64–67]. The dendrograms were obtained using the Euclidean distance (*X*-axis) and the complete linkage (*Y*-axis), and show the distance between the compounds inside the clusters, that are grouped according to its chemical similarity encoded by the MDs used as variables. Linear Discriminant Analysis (LDA) a simple and very useful technique in drug design was carried out to find the QSAR models [13, 16, 17, 19, 34, 35, 37, 38, 42–47, 68–73]. Here, the forward stepwise procedure was fixed as the strategy for variable selection and taken into account the principle of parsimony (Occam's razor) for model selection.

The classification of cases was carried out by mean of posterior classification probabilities. Tyrosinase inhibitory activity was codified by a dummy variable “**Class**”. This variable indicates the presence of either an active compound (**Class** = 1) or an inactive compound (**Class** = -1). By using the models, one compound can then be classified as active, if  $\Delta P\% > 0$ , being  $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ , or as inactive otherwise. *P* (Active) and *P* (Inactive) are the probabilities with which the equations classify a compound as active or inactive, respectively.

The Randić's method of orthogonalization was used in this study as a way to avoid the interrelation among the molecular fingerprints [45, 74–79]. This may possible a better statistical interpretation of the correlation coefficient and to evaluate the role of individual MDs in the QSAR model.

The data set was standardized before the orthogonalization process, because the different MDs included here used entirely “different types of scales”. This process to proportionate each variable has a mean of 0 and a standard deviation of 1.

### Experimental methods

The synthesis and characterization of the 24 tetraketones, their biological studies and cross references have been reported by other of our research team [80].

Tyrosinase inhibition assay was performed with kojic acid and L-mimosine as standard inhibitors for the tyrosinase in a 96-well microplate format using a SpectraMax 340 micro-plate reader (Molecular Devices, CA, USA) according to the method developed by Hearing [81]. Briefly, the compounds were first screened for the *o*-diphenolase inhibitory activity of tyrosinase using L-DOPA as substrate. All the active inhibitors from the preliminary screening were subjected to IC<sub>50</sub> studies. Compounds were dissolved in methanol to a concentration of 2.5%. Thirty units of mushroom tyrosinase (28 nM from Sigma Chemical Co., USA) were first preincubated with the test compounds in 50 nM Na-phosphate buffer (pH 6.8) for 10 min at 25 °C. Then the L-DOPA (0.5 mM) was added to the reaction mixture and the enzymatic reaction was monitored by measuring the change in absorbance at 475 nm (at 37 °C) due to the formation of the DOPochrome for 10 min. The percentage of inhibition of the enzyme was calculated as follows, by using MS Excel<sup>®</sup>™ 2000 (Microsoft Corp., USA) based program developed for this purpose:

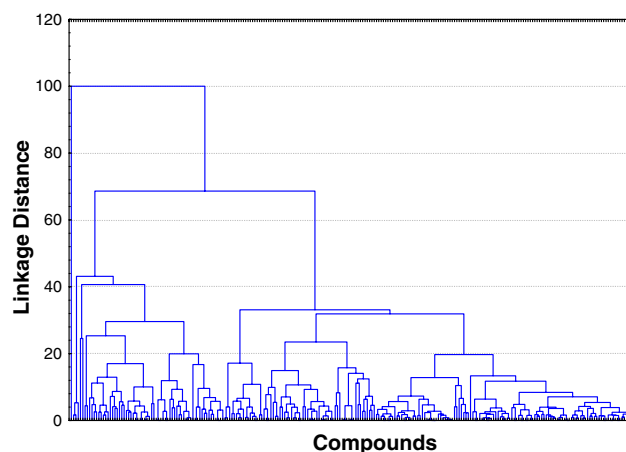
$$\text{Percent inhibition} = [(B - S)/B] \times 100 \quad (19)$$

Here, *B* and *S* are the absorbances for the blank and samples, respectively. After the screening of the compounds, 50% inhibitory concentrations (IC<sub>50</sub>) were also calculated. Kojic acid and L-mimosine were used as standard inhibitors for the tyrosinase and both of them were purchased from Sigma Chem. Co., USA.

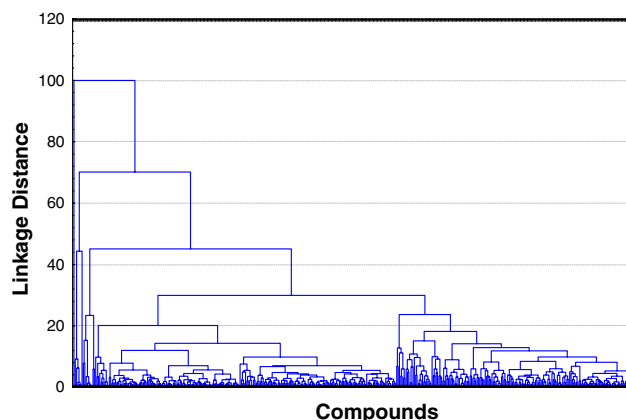
### Results and discussion

Dividing the training and prediction series through cluster analysis

In above section we describe the database selection process, now the structural variability of such set must be proved. This is a crucial aspect in any QSAR study



**Fig. 2** A dendrogram illustrating the results of the hierarchical *k*-NNCA of the set of tyrosinase inhibitors used in the training and prediction set of the present work



**Fig. 3** A dendrogram illustrating the results of the hierarchical *k*-NNCA of the set of inactive compounds (non-tyrosinase inhibitors) used in the training and prediction set of the present work

in order to explain its reliability. Following this main reason, different cluster analysis techniques were carried out. In first place was used a *k*-NNCA to prove the structural diversity in the families presented in the data. Two dendrograms, one for the active compounds series and other for the inactive ones, were obtained through hierarchical cluster analysis (Figs. 2, 3) were can be observed different structural patterns which demonstrate the chemical variability of the database.

Now the dataset should be partitioned in training and prediction sets, to find the discriminant functions, but due to the difficulty of evaluating the output dendrograms other kind of CA must be do it, for the selection of compounds in a ‘rational’ way.

Therefore we chose the *k*-MCA to solve this problem, and were applied to active and inactive subsets. The first *k*-MCA for tyrosinase inhibitors divide this

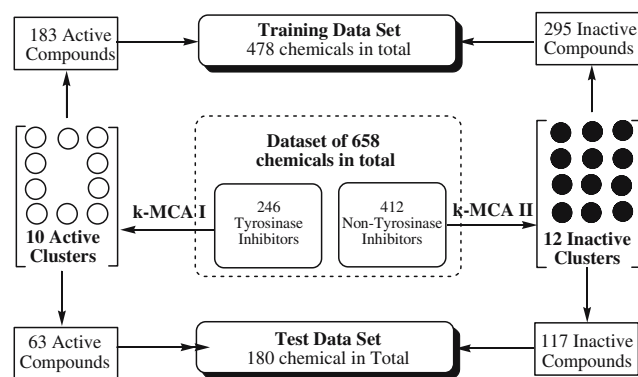
**Table 2** Main results of the *k*-MCAs, for tyrosinase inhibitors and inactives drug-like compounds

Analysis of variance				
Variables	Between SS <sup>a</sup>	Within SS <sup>b</sup>	Fisher ratio ( <i>F</i> )	<i>p</i> -level <sup>c</sup>
<i>Tyrosinase inhibitors clusters (k-MCA I)</i>				
<i>VP</i> <b>b</b> <sub>1L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	187.01	20.29	241.65	0.00
<i>VK</i> <b>b</b> <sub>1</sub> <sup>H</sup> ( <i>x</i> )	301.99	34.58	228.98	0.00
<i>VK</i> <b>b</b> <sub>0L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	298.30	56.31	138.91	0.00
<i>VK</i> <b>b</b> <sub>0</sub> <sup>H</sup> ( <i>x</i> )	309.23	33.40	242.77	0.00
<i>VP</i> <b>b</b> <sub>5</sub> <sup>H</sup> ( <i>x</i> )	213.04	25.90	215.68	0.00
<i>VK</i> <b>b</b> <sub>2</sub> <sup>H</sup> ( <i>x</i> )	265.14	35.70	194.73	0.00
<i>VP</i> <b>b</b> <sub>0L</sub> ( <i>x</i> <sub>E</sub> )	157.80	25.41	162.86	0.00
<i>MK</i> <b>b</b> <sub>1L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	164.30	29.24	147.37	0.00
<i>MP</i> <b>b</b> <sub>0L</sub> ( <i>x</i> <sub>E</sub> )	50.63	9.54	139.12	0.00
<i>MK</i> <b>b</b> <sub>1L</sub> ( <i>x</i> <sub>E</sub> )	192.77	25.84	195.60	0.00
<i>MP</i> <b>b</b> <sub>0L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	26.04	8.80	77.64	0.00
<i>VP</i> <b>b</b> <sub>1L</sub> ( <i>x</i> <sub>E</sub> )	187.01	20.29	241.65	0.00
<i>Inactives clusters (k-MCA II)</i>				
<i>VP</i> <b>b</b> <sub>1L</sub> ( <i>x</i> <sub>E</sub> )	455.59	126.84	130.62	0.00
<i>VK</i> <b>b</b> <sub>1</sub> <sup>H</sup> ( <i>x</i> )	312.80	60.56	187.83	0.00
<i>VK</i> <b>b</b> <sub>0L</sub> ( <i>x</i> <sub>E</sub> )	406.45	74.23	199.12	0.00
<i>VK</i> <b>b</b> <sub>0</sub> <sup>H</sup> ( <i>x</i> )	365.14	75.46	175.96	0.00
<i>VP</i> <b>b</b> <sub>5</sub> <sup>H</sup> ( <i>x</i> )	259.67	208.61	45.26	0.00
<i>VK</i> <b>b</b> <sub>2</sub> <sup>H</sup> ( <i>x</i> )	298.04	50.65	213.97	0.00
<i>VP</i> <b>b</b> <sub>0L</sub> ( <i>x</i> <sub>E</sub> )	679.51	85.82	287.93	0.00
<i>MK</i> <b>b</b> <sub>1L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	412.22	125.66	119.29	0.00
<i>MP</i> <b>b</b> <sub>0L</sub> ( <i>x</i> <sub>E</sub> )	1171.79	88.34	482.36	0.00
<i>MK</i> <b>b</b> <sub>1L</sub> ( <i>x</i> <sub>E</sub> )	451.92	131.25	125.21	0.00
<i>MP</i> <b>b</b> <sub>0L</sub> <sup>H</sup> ( <i>x</i> <sub>E</sub> )	1215.08	106.85	413.53	0.00

<sup>a</sup> Variability between groups

<sup>b</sup> Variability within groups

<sup>c</sup> Level of significance

**Fig. 4** General algorithm used to design training and test sets throughout *k*-MCA

set into 10 clusters. On other hand the *k*-MCA II split the inactive set into 12 clusters. The variables used were the *k*th non-stochastic bond-bilinear indices, and the analyses of variance for these *k*-MCAs are depicted in Table 2.

The following process using the cluster analysis techniques to divide entire database in training and prediction series is shown in shown in Fig. 4. How can

be observed in the same diagram there are 183 active compounds and 295 inactive ones belonging to training set (478 organic-chemicals). The prediction series of 180 compounds have 63 tyrosinase inhibitors and 117 non-inhibitors of tyrosinase.

### Developing the discriminant functions

The representative selection of training set permit continues to the next step, the finding of the classification functions to discriminate between active and inactive. For this we select the LDA as statistical technique due to it's broadly use and simplicity.

In total were obtained fourteen models, the first six models developed with the non-stochastic bond-based bilinear indices and the other first six perform with the stochastic molecular descriptors, these equations are depicted Table 3. Besides, below we shown the Eqs. 32 and 33 of the last seven models in both cases (non-stochastic and stochastic molecular fingerprints) resulting in a combination of all pairs of atom weights (atomic labels):

$$\begin{aligned}
 \text{Class} = & -0.636 - 8.422 \times 10^{-2MP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) \\
 & + 0.107^{MP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E) \\
 & + 1.792 \times 10^{-2MK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) \\
 & - 2.373 \times 10^{-2MK} \mathbf{b}_{1L}(\bar{w}_E, \bar{u}_E) \\
 & + 3.287 \times 10^{-5VP} \mathbf{b}_5^H(\bar{w}, \bar{u}) \\
 & - 9.590 \times 10^{-2VP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E) \\
 & + 1.166 \times 10^{-2VP} \mathbf{b}_{1L}(\bar{w}_E, \bar{u}_E) \\
 & + 2.277 \times 10^{-2VK} \mathbf{b}_0^H(\bar{w}, \bar{u}) \\
 & + 5.4 \times 10^{-3VK} \mathbf{b}_1^H(\bar{w}, \bar{u}) \\
 & - 4.04 \times 10^{-3VK} \mathbf{b}_2^H(\bar{w}, \bar{u}) \\
 & + 2.34 \times 10^{-2VK} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)
 \end{aligned} \quad (32)$$

$$\begin{aligned}
 N = 478 \quad \lambda = 0.45 \quad D^2 = 5.13 \quad F = 51.6 \quad \text{Canonical} \\
 R = 0.74 \quad \chi^2 = 374.8 \quad Q_{Total} = 91.00\% \quad C = 0.81
 \end{aligned}$$

$$\begin{aligned}
 \text{Class} = & -0.302 + 5.290 \times 10^{-3MV} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) \\
 & + 6.267 \times 10^{-3MP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E) \\
 & + 1.262 \times 10^{-2MK} \mathbf{b}_0(\bar{w}, \bar{u}) - 3.458 \\
 & \times 10^{-2MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 1.734 \times 10^{-2VP} \mathbf{b}_0(\bar{w}, \bar{u}) \\
 & + 1.286 \times 10^{-2VP} \mathbf{b}_{14L}(\bar{w}, \bar{u}) - 4.840 \\
 & \times 10^{-2VP} \mathbf{b}_{4L}(\bar{w}_E, \bar{u}_E) + 0.129^{VK} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) \\
 & - 0.133^{VK} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) + 1.067 \times 10^{-2VK} \mathbf{b}_{0L} \\
 & (\bar{w}_E, \bar{u}_E)
 \end{aligned} \quad (33)$$

**Table 3** Discriminant models obtained with total and local non-stochastic and stochastic bond-based bilinear indices used in this study*LDA-based QSAR models obtained using non-stochastic bilinear indices*

$$\text{Class} = -0.333 + 6.296 \times 10^{-3MV} \mathbf{b}_0^H(\bar{w}, \bar{u}) + 2.519 \times 10^{-3MV} \mathbf{b}_1^H(\bar{w}, \bar{u}) - 1.062 \times 10^{-3MV} \mathbf{b}_2^H(\bar{w}, \bar{u}) + 5.213 \times 10^{-13MV} \mathbf{b}_{15}^H(\bar{w}, \bar{u}) - 1.018 \times 10^{-2MV} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 1.592 \times 10^{-3MV} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 1.877 \times 10^{-4MV} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) + 4.316 \times 10^{-3MV} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 4.201 \times 10^{-3MV} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 4.939 \times 10^{-4MV} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 1.303 \times 10^{-6MV} \mathbf{b}_{6L}^H(\bar{w}_E, \bar{u}_E) - 2.632 \times 10^{-2MV} \mathbf{b}_{14L}^H(\bar{w}_E, \bar{u}_E) \quad (20)$$

$$\text{Class} = -0.351 + 6.742 \times 10^{-2MP} \mathbf{b}_0^H(\bar{w}, \bar{u}) + 3.085 \times 10^{-3MP} \mathbf{b}_1^H(\bar{w}, \bar{u}) - 1.727 \times 10^{-2MP} \mathbf{b}_2^H(\bar{w}, \bar{u}) + 2.900 \times 10^{-2MP} \mathbf{b}_1(\bar{w}, \bar{u}) - 1.381 \times 10^{-2MP} \mathbf{b}_2(\bar{w}, \bar{u}) + 3.085 \times 10^{-3MP} \mathbf{b}_3(\bar{w}, \bar{u}) - 0.118 \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 1.584 \times 10^{-2MP} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) + 5.058 \times 10^{-2MP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 6.603 \times 10^{-2MP} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 9.529 \times 10^{-3MP} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 3.077 \times 10^{-4MP} \mathbf{b}_{4L}^H(\bar{w}_E, \bar{u}_E) \quad (21)$$

$$\text{Class} = -1.577 + 6.753 \times 10^{-2MK} \mathbf{b}_0^H(\bar{w}, \bar{u}) - 5.903 \times 10^{-4MK} \mathbf{b}_4^H(\bar{w}, \bar{u}) + 3.015 \times 10^{-8MK} \mathbf{b}_{10}^H(\bar{w}, \bar{u}) + 1.585 \times 10^{-3MK} \mathbf{b}_2^H(\bar{w}, \bar{u}) - 9.832 \times 10^{-2MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 1.899 \times 10^{-2MK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 2.051 \times 10^{-3MK} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) - 2.572 \times 10^{-12MK} \mathbf{b}_{15L}^H(\bar{w}_E, \bar{u}_E) + 3.148 \times 10^{-2MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 2.375 \times 10^{-2MK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) - 2.385 \times 10^{-5MK} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) \quad (22)$$

$$\text{Class} = -0.630 + 3.534 \times 10^{-2VP} \mathbf{b}_0^H(\bar{w}, \bar{u}) + 3.696 \times 10^{-2VP} \mathbf{b}_1^H(\bar{w}, \bar{u}) - 5.218 \times 10^{-3VP} \mathbf{b}_3^H(\bar{w}, \bar{u}) + 1.494 \times 10^{-4VP} \mathbf{b}_5^H(\bar{w}, \bar{u}) - 6.702 \times 10^{-2VP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 1.834 \times 10^{-3VP} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) + 2.303 \times 10^{-2VP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 3.110 \times 10^{-2VP} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 4.899 \times 10^{-3VP} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 1.810 \times 10^{-7VP} \mathbf{b}_{10L}^H(\bar{w}_E, \bar{u}_E) + 1.188 \times 10^{-9VP} \mathbf{b}_{14L}^H(\bar{w}_E, \bar{u}_E) - 1.813 \times 10^{-10VP} \mathbf{b}_{15L}^H(\bar{w}_E, \bar{u}_E) \quad (23)$$

$$\text{Class} = -1.058 \times 10^{-2} - 2.314 \times 10^{-2VK} \mathbf{b}_0^H(\bar{w}, \bar{u}) + 2.108 \times 10^{-2VK} \mathbf{b}_1^H(\bar{w}, \bar{u}) - 3.513 \times 10^{-3VK} \mathbf{b}_2^H(\bar{w}, \bar{u}) + 2.053 \times 10^{-12VK} \mathbf{b}_{15}^H(\bar{w}, \bar{u}) + 5.867 \times 10^{-3VK} \mathbf{b}_1(\bar{w}, \bar{u}) - 1.494 \times 10^{-3VK} \mathbf{b}_2(\bar{w}, \bar{u}) + 1.075 \times 10^{-2VK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 2.066 \times 10^{-2VK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 4.864 \times 10^{-3VK} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 1.212 \times 10^{-6VK} \mathbf{b}_{8L}^H(\bar{w}_E, \bar{u}_E) + 4.842 \times 10^{-10VK} \mathbf{b}_{14L}^H(\bar{w}_E, \bar{u}_E) - 8.174 \times 10^{-11VK} \mathbf{b}_{15L}^H(\bar{w}_E, \bar{u}_E) \quad (24)$$

$$\text{Class} = 0.182 - 0.271 \mathbf{b}_0^H(\bar{w}, \bar{u}) + 0.311 \mathbf{b}_1^H(\bar{w}, \bar{u}) - 6.625 \times 10^{-2PK} \mathbf{b}_2^H(\bar{w}, \bar{u}) + 2.712 \times 10^{-11PK} \mathbf{b}_{15}^H(\bar{w}, \bar{u}) + 9.137 \times 10^{-2PK} \mathbf{b}_1(\bar{w}, \bar{u}) - 1.794 \times 10^{-2PK} \mathbf{b}_2(\bar{w}, \bar{u}) + 1.452 \times 10^{-3PK} \mathbf{b}_{4L}^H(\bar{w}_E, \bar{u}_E) + 0.162 \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 0.289 \times 10^{-2PK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 5.107 \times 10^{-2PK} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 3.957 \times 10^{-4PK} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) \quad (25)$$

*LDA-based QSAR models obtained using stochastic bilinear indices*

$$\text{Class} = -0.303 - 4.647 \times 10^{-3MV} \mathbf{b}_0^H(\bar{w}, \bar{u}) - 2.557 \times 10^{-3MV} \mathbf{b}_1^H(\bar{w}, \bar{u}) + 8.831 \times 10^{-3MV} \mathbf{b}_{15}^H(\bar{w}, \bar{u}) + 5.245 \times 10^{-3MV} \mathbf{b}_1(\bar{w}, \bar{u}) - 7.103 \times 10^{-3MV} \mathbf{b}_2(\bar{w}, \bar{u}) + 2.955 \times 10^{-3MV} \mathbf{b}_{15}(\bar{w}, \bar{u}) + 3.191 \times 10^{-2MV} \mathbf{b}_{4L}^H(\bar{w}_E, \bar{u}_E) - 2.634 \times 10^{-2MV} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) - 8.967 \times 10^{-3MV} \mathbf{b}_{15L}^H(\bar{w}_E, \bar{u}_E) + 3.855 \times 10^{-3MV} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 4.915 \times 10^{-3MV} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) \quad (26)$$

$$\text{Class} = -0.696 - 0.148 \mathbf{b}_2^H(\bar{w}, \bar{u}) + 6.719 \times 10^{-2MP} \mathbf{b}_1^H(\bar{w}, \bar{u}) + 0.102 \mathbf{b}_{15}^H(\bar{w}, \bar{u}) + 4.644 \times 10^{-2MP} \mathbf{b}_1(\bar{w}, \bar{u}) - 6.597 \times 10^{-2MP} \mathbf{b}_2(\bar{w}, \bar{u}) + 2.518 \times 10^{-2MP} \mathbf{b}_{14}(\bar{w}, \bar{u}) - 5.617 \times 10^{-2MP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 0.174 \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 0.154 \mathbf{b}_{12L}^H(\bar{w}_E, \bar{u}_E) + 5.6291 \times 10^{-2MP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 6.346 \times 10^{-2MP} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) \quad (27)$$

$$\text{Class} = -0.269 - 1.435 \times 10^{-2MK} \mathbf{b}_0(\bar{w}, \bar{u}) + 1.412 \times 10^{-2MK} \mathbf{b}_1(\bar{w}, \bar{u}) - 3.842 \times 10^{-2MK} \mathbf{b}_9(\bar{w}, \bar{u}) + 4.586 \times 10^{-2MK} \mathbf{b}_{14}(\bar{w}, \bar{u}) - 3.746 \times 10^{-2MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) + 4.210 \times 10^{-2MK} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) + 3.622 \times 10^{-2MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 2.421 \times 10^{-2MK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) - 2.423 \times 10^{-2MK} \mathbf{b}_{8L}^H(\bar{w}_E, \bar{u}_E) \quad (28)$$

$$\text{Class} = -0.422 - 3.200 \times 10^{-2VP} \mathbf{b}_0^H(\bar{w}, \bar{u}) - 5.377 \times 10^{-2VP} \mathbf{b}_2^H(\bar{w}, \bar{u}) + 0.112 \mathbf{b}_{13}^H(\bar{w}, \bar{u}) - 1.711 \times 10^{-2VP} \mathbf{b}_0(\bar{w}, \bar{u}) + 2.413 \times 10^{-2VP} \mathbf{b}_{14}(\bar{w}, \bar{u}) - 5.030 \times 10^{-2VP} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 0.543 \mathbf{b}_{4L}^H(\bar{w}_E, \bar{u}_E) - 0.438 \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) - 0.103 \mathbf{b}_{15L}^H(\bar{w}_E, \bar{u}_E) + 2.156 \times 10^{-2VP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 4.654 \times 10^{-2VP} \mathbf{b}_{4L}^H(\bar{w}_E, \bar{u}_E) + 2.222 \times 10^{-2VP} \mathbf{b}_{14L}^H(\bar{w}_E, \bar{u}_E) \quad (29)$$

$$\text{Class} = -0.736 - 5.040 \times 10^{-2VK} \mathbf{b}_0^H(\bar{w}, \bar{u}) + 5.701 \times 10^{-2VK} \mathbf{b}_{15}^H(\bar{w}, \bar{u}) - 1.007 \times 10^{-2VK} \mathbf{b}_0(\bar{w}, \bar{u}) + 1.079 \times 10^{-2VK} \mathbf{b}_1(\bar{w}, \bar{u}) + 1.004 \times 10^{-2VK} \mathbf{b}_{14}(\bar{w}, \bar{u}) - 2.952 \times 10^{-2VK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 0.153 \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 7.459 \times 10^{-2VK} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) - 6.737 \times 10^{-2VK} \mathbf{b}_1^H(\bar{w}_E, \bar{u}_E) + 1.670 \times 10^{-2VK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 2.333 \times 10^{-2VK} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E) \quad (30)$$

$$\text{Class} = -0.401 - 0.558 \mathbf{b}_0^H(\bar{w}, \bar{u}) + 0.605 \mathbf{b}_1^H(\bar{w}, \bar{u}) - 0.502 \mathbf{b}_2^H(\bar{w}, \bar{u}) + 0.560 \mathbf{b}_{15}^H(\bar{w}, \bar{u}) - 0.161 \mathbf{b}_0(\bar{w}, \bar{u}) + 0.180 \mathbf{b}_1(\bar{w}, \bar{u}) + 0.118 \mathbf{b}_{14}(\bar{w}, \bar{u}) - 0.972 \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E) + 1.747 \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E) - 1.016 \mathbf{b}_{7L}^H(\bar{w}_E, \bar{u}_E) + 0.205 \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E) - 0.287 \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E) \quad (31)$$

$N = 478$     $\lambda = 0.46$     $D^2 = 5.00$     $F = 55.4$  Canonical  
 $R = 0.74$     $\chi^2 = 368.5$     $Q_{\text{Total}} = 90.17\%$     $C = 0.79$

Prediction performances of all the obtained models including these last two equations are given in Table 4, together with the Wilks' statistics ( $\lambda$ ), the square of the Mahalanobis distances ( $D^2$ ), and the Fisher ratio (F). The models selected showed to be statistically significant at  $p$ -level  $< 0.0001$ .

The fitted models 32 and 33, resulting of the combination of weighting schemes for the non-stochastic and stochastic bond-level bilinear indices, respectively, exhibit the best results, how can be observed in Table 4. These best two equations correctly classified the 91.00% and 90.17% of the training set, and showed values of the Matthews

correlation coefficients ( $C$ ) of 0.81 and 0.79, respectively. The most common parameters in medical statistics for all the models are depicted in the same Table 4.

Although these two best models exhibited good results, the interpretation of the individual role of every index in the model can become in a difficulty due to the interrelation among them (data not shown). This impelled us to use the Randić's orthogonalization process to avoid this problem, and eliminate the colinearity between the variables [74–78].

In Eqs. 34 and 35 are depicted the results of the orthogonalization process for the best two models of the non-stochastic and stochastic bilinear indices, correspondingly.

**Table 4** Prediction performances and statistical parameters for LDA-based QSAR models in the training set

Models <sup>a</sup>	Matthews corr. coefficient (C)	Accuracy 'Q <sub>Total</sub> ' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False positive rate (%)	Wilks' λ	D <sup>2</sup>	F	Chi-sqr (χ <sup>2</sup> )	Canonical R(R <sub>can</sub> ) <sup>b</sup>
LDA-based QSAR models obtained using non-stochastic bilinear indices										
Equation 20 (12)	0.75	87.87	82.1	88.0	11.9	0.52	3.83	35.2	279.3	0.67
Equation 21 (12)	0.70	85.77	79.8	84.1	13.2	0.52	3.89	35.7	307.1	0.69
Equation 22 (11)	0.80	90.59	86.3	89.6	8.8	0.47	4.81	48.4	358.3	0.73
Equation 23 (12)	0.70	85.77	79.2	85.2	13.9	0.58	3.11	28.6	243.4	0.64
Equation 24 (12)	0.74	87.45	81.5	86.9	12.2	0.51	4.03	37.0	288.9	0.68
Equation 25 (12)	0.73	87.24	81.4	86.3	12.2	0.51	4.03	37.0	315.9	0.70
Equation 32 (11)	0.81	91.00	86.8	90.2	8.5	0.45	5.13	51.6	374.8	0.74
LDA-based QSAR models obtained using stochastic bilinear indices										
Equation 26 (11)	0.71	86.19	79.7	85.8	13.6	0.54	3.52	35.4	286.0	0.67
Equation 27 (11)	0.70	85.98	82.6	80.3	10.5	0.57	3.23	32.5	267.8	0.66
Equation 28 (9)	0.71	85.98	79.6	85.2	13.6	0.52	3.85	47.5	306.2	0.69
Equation 29 (12)	0.67	84.31	77.6	83.1	14.9	0.57	3.13	28.8	260.9	0.65
Equation 30 (11)	0.73	87.24	81.8	85.8	11.9	0.52	3.94	39.6	310.6	0.70
Equation 31 (12)	0.76	88.28	82.6	88.0	11.5	0.53	3.75	34.5	299.2	0.69
Equation 33 (10)	0.79	90.17	85.8	89.1	9.2	0.46	5.00	55.4	368.5	0.74

<sup>a</sup> Between brackets the quantity of variables of the models

<sup>b</sup> Canonical correlation coefficient obtained from the linear discriminant canonical analysis

$$\begin{aligned}
 \text{Class} = & -0.331 - 1.515^1 O(^{VP} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E)) \\
 & + 2.037^2 O(^{VK} \mathbf{b}_1^H(\bar{w}, \bar{u})) - 2.406^3 O(^{VK} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)) \\
 & - 3.176^4 O(^{VK} \mathbf{b}_0^H(\bar{w}, \bar{u})) + 0.805^5 O(^{VP} \mathbf{b}_5^H(\bar{w}, \bar{u})) \\
 & - 6.540^6 O(^{VK} \mathbf{b}_2^H(\bar{w}, \bar{u})) \\
 & - 1.959^7 O(^{VP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)) \\
 & - 1.015^8 O(^{MK} \mathbf{b}_{1L}^H(\bar{w}_E, \bar{u}_E)) \\
 & + 1.913^9 O(^{MP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)) \\
 & - 5.997^{10} O(^{MK} \mathbf{b}_{1L}(\bar{w}_E, \bar{u}_E)) \\
 & - 16.337^{11} O(^{MP} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E)) \quad (34)
 \end{aligned}$$

$$N = 478 \quad \lambda = 0.45 \quad D^2 = 5.13 \quad F = 51.6 \quad \text{Canonical} \\
 R = 0.74 \quad \chi^2 = 374.8 \quad Q_{\text{Total}} = 91.00\% \quad C = 0.81.$$

$$\begin{aligned}
 \text{Class} = & -2.414 \times 10^{-2} - 1.197^1 O(^{VP} \mathbf{b}_{4L}(\bar{w}_E, \bar{u}_E)) \\
 & + 4.346^2 O(^{VK} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)) \\
 & + 1.075^3 O(^{VP} \mathbf{b}_{14}^H(\bar{w}, \bar{u})) - 3.196^4 O(^{VP} \mathbf{b}_0(\bar{w}, \bar{u})) \\
 & + 0.899^5 O(^{MP} \mathbf{b}_{0L}(\bar{w}_E, \bar{u}_E)) \\
 & - 1.197^6 O(^{MK} \mathbf{b}_{0L}^H(\bar{w}_E, \bar{u}_E)) \\
 & + 5.474^7 O(^{MK} \mathbf{b}_0(\bar{w}, \bar{u})) \\
 & + 2.093^8 O(^{VK} \mathbf{b}_{2L}^H(\bar{w}_E, \bar{u}_E)) \\
 & - 7.371^9 O(^{VK} \mathbf{b}_{3L}^H(\bar{w}_E, \bar{u}_E)) \\
 & + 3.461^{10} O(^{MV} \mathbf{b}_{5L}^H(\bar{w}_E, \bar{u}_E)) \quad (35)
 \end{aligned}$$

$N = 478 \quad \lambda = 0.46 \quad D^2 = 5.00 \quad F = 55.4 \quad \text{Canonical}$   
 $R = 0.74 \quad \chi^2 = 368.5 \quad Q_{\text{Total}} = 90.17\% \quad C = 0.79.$   
 Here, we used the symbols  ${}^m O[\mathbf{b}_k(\bar{w}, \bar{u})]$ , where the superscript  $m$  expresses the order of importance of the variable  $\mathbf{b}_k(\bar{w}, \bar{u})$  after a preliminary forward-stepwise analysis and  $O$  means orthogonal. If we take a look to the statistical parameter to every model before and after of the orthogonalization process, can be observed that they keep be the same for the non-orthogonal and orthogonal descriptors.

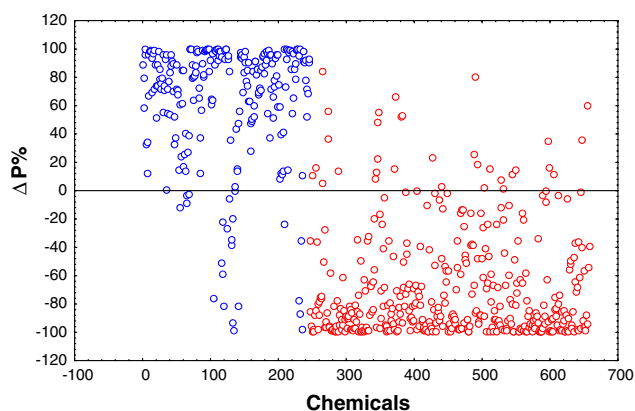
Assessing the predictive power of the models

Validation external process or most commonly namely test set is necessary to ensure the quality and extrapolation power of the QSAR models found in this report [82, 83]. Following this aim all the equations were evaluated and results are shown in Table 5. In the case of the best two discriminant functions Eqs. 32 and 33, presented overall accuracies of 93.33% ( $C = 0.85$ ) and 88.89% ( $C = 0.77$ ), respectively. Likewise a plot of the  $\Delta P\%$  for the entire dataset using models (32) and (33), is illustrates in Figs. 5 and 6.

The results of the classification using the total fourteen models, for all the active and inactive organic-chemicals in the training and external series are shown in Tables 3–10 of Supporting Information.

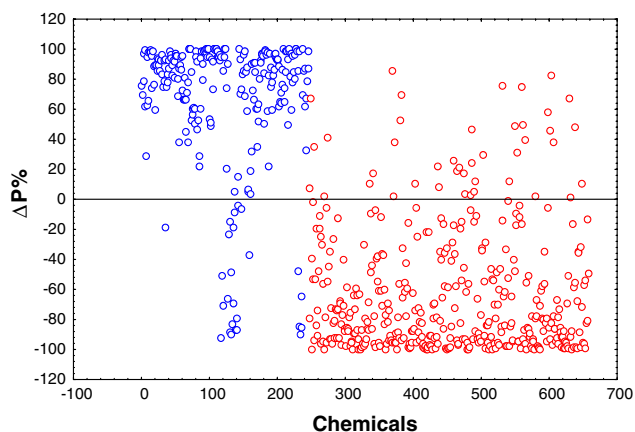
**Table 5** Prediction performances for LDA-based QSAR models in the test set

Models	Matthews corr. coefficient ( <i>C</i> )	Accuracy ' $Q_{Total}$ ' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False positive Rate (%)
<i>LDA-based QSAR models obtained using non-stochastic bilinear indices</i>					
Equation 20	0.69	85.56	76.8	84.1	13.7
Equation 21	0.64	83.33	73.9	81.0	15.4
Equation 22	0.73	87.78	82.5	82.5	9.4
Equation 23	0.65	83.89	75.0	81.0	14.5
Equation 24	0.67	83.89	73.0	85.7	17.1
Equation 25	0.67	84.44	74.0	85.7	16.2
Equation 32	0.85	93.33	89.2	92.1	6.0
<i>LDA-based QSAR models obtained using stochastic bilinear indices</i>					
Equation 26	0.68	85.00	75.0	85.7	15.4
Equation 27	0.69	86.11	81.7	77.8	9.4
Equation 28	0.74	86.67	74.7	93.7	17.1
Equation 29	0.65	83.33	72.6	84.1	17.1
Equation 30	0.71	86.11	75.7	88.9	15.4
Equation 31	0.68	85.00	75.7	84.1	14.5
Equation 33	0.77	88.89	79.5	92.1	12.8

**Fig. 5** Plot of the  $\Delta P\%$  from Eq. 32 (using non-stochastic bond-based bilinear indices) for each compound in the training and test sets. Compounds 1–183 and 184–246 are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247–541 and 542–658 are inactive (non-tyrosinase inhibitors) in both training and test sets, correspondingly

#### Simulated virtual screening of new tyrosinase inhibitors

The good behavior of the results obtained above, encouraged us, to expand moreover the possibilities of this novel approach for the *in silico* discovery of novel tyrosinase inhibitor compounds. Virtual High Throughput Screening (HTS) can become an important tool capable to resolve the largely query of database of thousands of compounds, and has the potential to transform early-stage drug discovery. To prove the ability of our models a simulated virtual screening to a data of 104 organic-chemicals (Table 6) reported from the literature as inactive/inactive (see the last column of the same Table 6: Ref) was carried out. The molecular structures of these compounds are shown in Table 11 of Supporting Information.

**Fig. 6** Plot of the  $\Delta P\%$  from Eq. 33 (using stochastic bond-based bilinear indices) for each compound in the training and test sets. Compounds 1–183 and 184–246 are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247–541 and 542–658 are inactive (non-tyrosinase inhibitors) in both training and test sets, correspondingly

Besides to assure the great possibility of our models to identify several classes of compounds a *k*-NNCA to this data was carried and the dendrogram obtained can be observed in Fig. 7, where a great molecular diversity can be visualized. In Table 6 are depicted the results of the classifications of the 104 compounds. Additionally the posterior classification probabilities (including canonical scores) for all the equations are summarized in Table 12 (Supporting Information). The percent of globally good classification were of 85.57% and 84.61% for the non-stochastic and stochastic molecular descriptors, correspondingly.

This method could very useful, due that making use of this many great databases of drug-like compounds could be make it, and some compounds identified

**Table 6** Results of the virtual screening

Compound <sup>a</sup>	Class <sup>b</sup>	Ref. <sup>c</sup>	Compound <sup>a</sup>	Class <sup>b</sup>	References <sup>c</sup>
<i>Active compounds (tyrosinase inhibitors)</i>					
<b>1</b> <i>p</i> -Nitrophenol	+++++++	A	<b>27</b> Dithiothreitol	+++++++	O
	+++++++	B		+++++++	
<b>2</b> 3-(3,4-Dihydroxyphenyl)-l-alanine	+++++++	C	<b>28</b> Azelaic acid	-+-----	P
	+++++++			++++-----	
<b>3</b> 3-Amino-4-hydroxybenzoic acid	+++++++	C	<b>29</b> Undecandioic acid	-+-----	P
	+++++++			++++-----	
<b>4</b> 4-Amino-3-hydroxybenzoic acid	+++++++	C	<b>30</b> Suberic acid	-+-----	P
	+++++++			++++-----	
<b>5</b> 3,4-Diaminobenzoic acid	+++++++	C	<b>31</b> Sebacic acid	-+-----	P
	+++++++			++++-----	
<b>6</b> 3-Aminobenzoic acid	+++++++	C	<b>32</b> Dodecandioic acid	-+-----	P
	+++++++			++++-----	
<b>7</b> 4-aminobenzoic acid	+++++++	C	<b>33</b> Tridecandioic acid	-+-----	P
	+++++++			++++-----	
<b>8</b> 4,6- <i>O</i> -hexahydroxy	-+-----	D	<b>34</b> Traumatic acid	-+-----	P
	+++++++			++++-----	
<b>9</b> Tunicamycin	+++-----	E	<b>35</b> Pantothenic acid	+++-----	K
	+++++++			++++-----	
<b>10</b> Methyl <i>p</i> -coumarate	+++++++	F	<b>36</b> 5-(hydroxymethyl)-2-furfural	+++++++	Q
	+++++++			+++-----	R
<b>11</b> <i>o</i> -Phenylphenol	+++++++	F	<b>37</b> Hinokitiol	+++++++	S
	+++++++			++++-----	
<b>12</b> Phenylhydroquinone	+++++++	F	<b>38</b> Penicillamine	+++-----	T
	+++++++			++++-----	
<b>13</b> Chamaecin	+++++++	F	<b>39</b> Toluic acid	+++++++	A
	+++++++	G		+-----	
<b>14</b> Stearyl glycyrrhetinate	+++++++	H	<b>40</b>	+++++++	U
	+++++++			++++-----	
<b>15</b> 2-(4-Methylphenyl)-1,3-selenazol-4-one	+++-----	I	<b>41</b>	+++++++	U
	-----	J		++++-----	
<b>16</b>	+++-----	I	<b>42</b> 3,5-dihydroxy-4'- <i>O</i> -methoxystilbene	+++++++	V
	-----			++++-----	
<b>17</b>	++++-----	I	<b>43</b> <i>p</i> -Hydroxybenzoic acid	+++++++	W
	-----			++++-----	
<b>18</b>	+++-----	I	<b>44</b> <i>o</i> -Hydroxybenzoic acid	+++++++	W
	-----			++++-----	
<b>19</b> 3-Fluorotyrosine	+++++++	K	<b>45</b> Cysteine	+++++++	X
	+++++++			++++-----	
<b>20</b> <i>N</i> -acetyltyrosine	-----	K	<b>46</b> Methimazole	+++++++	X
	-----			+-----	
<b>21</b> <i>N</i> -formyltyrosine	+++++++	K	<b>47</b> BMY-28438	+++++++	X
	+++++++			++++-----	
<b>22</b> Gentisic acid	+++++++	L	<b>48</b> Captopril	+++++++	Y
	+++++++			+++-----	
<b>23</b> 6-BH <sub>4</sub>	-----	M	<b>49</b> Yohimbine	+-----	Z
	+++-----			++++-----	
<b>24</b> 7-BH <sub>4</sub>	-----	M	<b>50</b> 4-(phenylazo)phenol	+++++++	<i>a</i>
	+++-----			++++-----	
<b>25</b> Propylparaben	+++++++	N	<b>51</b> SACat	+++++++	<i>a</i>
	+++-----			++++-----	
<b>26</b> Phenylalanine	+++++++	K	<b>52</b> NPACat	+++++++	<i>a</i>
	+++++++			++++-----	
<b>53</b> DNPCat	+++-----	<i>a</i>	<b>70</b> 27	+++++++	<i>d</i>
	+++++++			++++-----	

**Table 6** Continued

Compound <sup>a</sup>	Class <sup>b</sup>	Ref. <sup>c</sup>	Compound <sup>a</sup>	Class <sup>b</sup>	References <sup>c</sup>
<b>54</b> EDTA	—	<i>b</i>	<b>71</b> 28	+++++++	<i>d</i>
<b>55</b> Dodecyl gallate	+++++++	<i>c</i>	<b>72</b> 29	+++++++	<i>d</i>
<b>56</b> Gallic acid	+++++++	<i>c</i>	<b>73</b> 30	+++++++	<i>d</i>
<b>57</b> (±)-flavanone	+++++++	<i>d</i>	<b>74</b> 31	+++++++	<i>d</i>
<b>58</b> (-)-Pinoembrin	+++++++	<i>d</i>	<b>75</b> 32	+++++++	<i>d</i>
<b>59</b> (±)-Naringenin	+++++++	<i>d</i>	<b>76</b> 34	+++++++	<i>d</i>
<b>60</b> (+)-Dihydromorin	+++++++	<i>d</i>	<b>77</b> 35	+++++++	<i>d</i>
<b>61</b> Flavone	+++++++	<i>d</i>	<b>78</b> 36	+++++++	<i>d</i>
<b>62</b> Myricetin	+++++++	<i>d</i>	<b>79</b> 37	+++++++	<i>d</i>
<b>63</b> Artocarpin	+++++++	<i>d</i>	<b>80</b> 38	+++++++	<i>d</i>
<b>64</b> Artocarpesin	+++++++	<i>d</i>	<b>81</b> 39	+++++++	<i>d</i>
<b>65</b> Isoartocarpesin	+++++++	<i>d</i>	<b>82</b> 40	+++++++	<i>d</i>
<b>66</b> (-)-Angolensin	+++++++	<i>d</i>	<b>83</b> 41	+++++++	<i>d</i>
<b>67</b> Pinosylvin	+++++++	<i>d</i>	<b>84</b> 2'- <i>O</i> -feruloylaloetin	+++++++	<i>e</i>
<b>68</b> 4-Prenyloxyresveratrol	+++++++	<i>d</i>	<b>85</b> Barbaloin	+++++++	<i>e</i>
<b>69</b> 26	+++++++	<i>d</i>		+++++++	
<i>Inactive compounds</i>					
<b>86</b>	+++++ -+++++	<i>f</i>	<b>95</b> 2-Formyl-5-methoxyfuran	+++++++	<i>I</i>
<b>87</b>	+++ -+++	<i>f</i>	<b>96</b> 5-Methyluracil	—	<i>k</i>
<b>88</b>	— -+	<i>f</i>	<b>97</b> Uracil	—	<i>k</i>
<b>89</b> Caffeine	-+ —	<i>g</i>	<b>98</b> Thiourea	—	<i>l</i>
<b>90</b> Trimethylresveratrol	+— +—+	<i>h</i>	<b>99</b> Veratric acid methyl ester	+++++++	<i>m</i>
<b>91</b> 4-Aminoazobenzene-4'-sulfonic acid	+++++++ +++++++	<i>j</i>	<b>100</b> 6-Nitroquipazine	—	<i>n</i>
<b>92</b> 2-Methoxy-4-isopropyl benzaldehyde	+++++++ +++ +—+	<i>G</i>	<b>101</b> 4-Methoxybenzaldehyde- <i>O</i> -ethylxime	+++ —	<i>o</i>
<b>93</b> Petroselinic acid	+++ +++++++	<i>P</i>	<b>102</b>	+++++++	<i>o</i>
<b>94</b> Crocusatins F	+— +++ +—+	<i>k</i>	<b>103</b>	+++++++	<i>o</i>



**Table 6** Continued

Compound <sup>a</sup>	Class <sup>b</sup>	Ref. <sup>c</sup>	Compound <sup>a</sup>	Class <sup>b</sup>	References <sup>c</sup>
<b>104</b> Phetyl-1-hexanoate	+++++	<i>p</i>			
	+++++				

<sup>a</sup> The molecular structures of these tyrosinase inhibitors is given as Supporting Information (see Table 11)

<sup>b</sup> Results of the classification of compounds in this set: (i) *Above*, classification of each compounds using the obtained models with non-stochastic bond-based linear indices in the following order: Eqs. 20, 21, 22, 23, 24, 25 and 32; and (ii) *Below*; classification of each compounds using the obtained models with stochastic bond-based linear indices in the following order Eqs. 26, 27, 28, 29, 30, 31 and 33

<sup>c</sup> References taken from the literature: <sup>A</sup>Bubacco, L.; van Gastel, M.; Groenen, E. J. J.; Vijgenboom, E.; Canters, G.W. *J. Biol. Chem.* **2003**, *278*, 7381–7389. <sup>B</sup>van Gastela, M.; Bubaccob, L.; Groenena, E. J. J.; Vijgenboomc, E.; Cantersc, G. W. *FEBS Lett.* **2000**, *474*, 228–232. <sup>C</sup>Gasowska, B.; Kafarskia, P.; Wojtasek, H. *Biochim. Biophys. Acta.* **2004**, *1673*, 170–177. <sup>D</sup><http://open.cacb.org.tw/index.php> (2005-03-03 09:09:51). <sup>E</sup>Takahashi, H.; Parsons, P. G. *J. Invest. Dermatol.* **1992**, *98*, 481–487. <sup>F</sup>Kubo, I.; Niheia, K.; Tsujimoto, K. *Bioorg. Med. Chem.* **2004**, *12*, 5349–5354. <sup>G</sup>Nihei, K-I.; Yamagiwa, Y.; Kamikawab, T.; Kubo, I. *Bioorg. Med. Chem Lett.* **2004**, *14*, 681–683. <sup>H</sup>Um, S-J.; Park, M-S.; Park, S-H.; Han, H-S.; Kwonb, Y-J.; Sin, H-S. *Bioorg. Med. Chem.* **2003**, *11*, 5345–5352. <sup>I</sup>Barlocco, D.; Barrett, D.; Edwards, P.; Langston, S.; Pérez-Pérez, M. J.; Walker, M.; Weidner, J.; Westwell, A. *Drug Disc. Today.* **2003**, *8*, 372–373. <sup>J</sup>Koketsu, M.; Choi, S.Y.; Ishihara, H.; Lim B. O.; Kim, H.; Kim, S., Y. *Chem. Pharm. Bull. (Tokyo).* **2002**, *12*, 1594–1596. <http://www.thecosmeticsite.com/formulating/959621.html> (April-00). <sup>L</sup>Curto, E. V.; Kwong, C.; Hermersdorfer, H.; Glatt, H.; Santis, C.; Virador, V.; Hearing, V. J.; Dooley, T. P. *Biochem. Pharmacol.* **1999**, *57*, 663–672. <sup>M</sup>Wood, J. M.; Schallreuter-Wood, K. U.; Lindsey, N. J.; Callaghan, S.; Gardner, M. L.G. *Biochem. Biophys. Res. Commun.* **1995**, *206*, 480–485. <sup>N</sup>Hori, I.; Nihei, K-I.; Kubo, I. *Phytother. Res.* **2004**, *18*, 475–479. <sup>O</sup>Naish-Byfield, S.; Cooksey, C. J.; Riley, P. A. *Biochem. J.* **1994**, *304*, 155–162. <sup>P</sup>Nazzaro-Porro, M.; Passi, S. *J. Invest. Dermatol.* **1978**, *71*, 205–208. <sup>Q</sup>Sharma, V. K.; Choi, J.; Sharma, N.; Choi, M.; Seo, S-Y. *Phytotherapy Res.* **2004**, *18*, 841–844. <sup>R</sup>Kang, H. S.; Choi, J. H.; Cho, W. K.; Park, J. C.; Choi, J. S. *Arch Pharm Res.* **2004**, *7*, 742–750. <sup>S</sup>Sakuma, K.; Ogawa, M.; Sugibayashi, K.; Yamada, K.; Yamamoto, K. *Arch Pharm Res.* **1999**, *4*, 335–339. <sup>T</sup>Lovstad, R. A. *Biochem. Pharmacol.* **1976**, *25*, 533–535. <sup>U</sup>Kubo, I.; Kinst-Hori, I.; Yokokawa, Y. *J. Nat. Prod.* **1994**, *57*, 545–551. <sup>V</sup>Regev-Shoshani, G.; Shoseyov, O.; Bilkis, I.; Kerem, Z. *Biochem. J.* **2003**, *374*, 157–163. <sup>W</sup>Bernard, P.; Berthon, J-Y. *Int. J. Cosmetic Sci.* **2000**, *22*, 219–226. <sup>X</sup>Imada, C.; Sugimoto, Y.; Makimura, T.; Kobayashi, T.; Hamada, N.; Watanabe, E. *Fish. Sci.* **2001**, *67*, 1151–1156. <sup>Y</sup>Espin, J. C.; Wichers, H. J. *Biochim. Biophys. Acta.* **2001**, *1544*, 289–300. <sup>Z</sup>Fuller, B. B.; Drake, M. A.; Spaulding, D. T.; Chaudry, F. J. *Invest. Dermatol.* **2000**, *114*, 2680–276. <sup>aa</sup>Borojerdi, S. S.; Haghbeen, K.; Karkhane, A. A.; Fazli, M.; Sabouryc, A. A. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 925–930. <sup>ab</sup>Kong, K-H.; Hong, M-P.; Choi, S-S.; Kim, Y-T.; Cho, S-H. *Biotechnol. Appl. Biochem.* **2000**, *31*, 113–118. <sup>ac</sup>Kubo, I.; Chen, Q-X.; Nihei, K-I. *Food Chem.* **2003**, *81*, 241–247. <sup>ad</sup>Shimizu, K.; Kondo, R.; Sakai, K. *Planta Medica.* **2000**, *66*, 11–15. <sup>ae</sup>Yagi, A.; Kanbara, T.; Morinobu, N. *Planta Medica.* **1987**, *515*–517. <sup>af</sup>Shiino, M.; Watanabe, Y.; Umezawa, K. *Bioorg. Med. Chem.* **2001**, *9*, 1233–1240. <sup>ag</sup>No, J. K.; Soung, D. Y.; Kim, Y. J.; Shim, K. H.; Jun, Y. S.; Rhee, S. H.; Yokozawa, T.; Chung, H. Y. *Life Sci.* **1999**, *65*, 241–246. <sup>ah</sup>Shin, N-H., Ryu, S.; Choi, Y. E. J.; Kang, S-H.; Chang, I-M.; Min, K. R.; Kim, Y. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 801–803. <sup>ai</sup>Kim, Y. M.; Yun, J.; Lee, C-K.; Lee, H.; Min, K. R.; Kim, Y. *J. Biol. Chem.* **2002**, *277*, 16340–16344. <sup>aj</sup>Komori, K.; Yatagai, K.; Tatsuma, T. *J. Biotechnol.* **2004**, *108*, 11–16. <sup>ak</sup>Li, C-Y.; Wu, T-S. *J. Nat. Prod.* **2002**, *65*, 1452–1456. <sup>al</sup>Gilly, R.; Mara, D.; Oded, S.; Zohar, K. *J. Agric. Food Chem.* **2001**, *49*, 1479–1485. <sup>am</sup>Miyazawa, M.; Oshima, T.; Koshio, K.; Itsuzaki, Y.; Anzai, J. *J. Agric. Food Chem.* **2003**, *51*, 6953–6956. <sup>an</sup>McEwan, M.; Garsons, P. G. *J. Invest. Dermatol.* **1990**, *89*, 82–86. <sup>ao</sup>Ley, J. P.; Bertram, H-J. *Bioorg. Med. Chem.* **2001**, *9*, 1879–1885. <sup>ap</sup>Şabudak, T.; Khan, M. T. H.; Choudhary, M. I.; Oksuz, S. E. *Nat. Prod. Res.* Accepted for publication

reported with the new biological activity, also taken into account that this kind of chemicals have well-established methods of synthesis, as well as their toxicological, pharmacodynamical and pharmaceutical properties are well known.

#### *Biosilico* identification of novel tyrosinase inhibitors and experimental corroboration

The entire algorithm describes in the above sections, was make up with the main objective to explore the possibilities of the current *in silico* approach for the identification of hits from largely databases. In this sense an *in silico* screening of novel compounds looking for the biological activity concern to this work was performed. To make this, a pool of compounds never described in the literature as tyrosinase inhibitors was chosen. Later the *in silico* essays were done using all the models developed inside this report, to

find bioactive chemicals that present tyrosinase inhibitory activity.

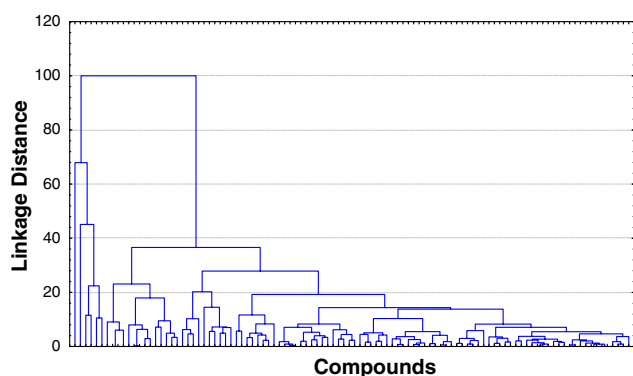
Here, 24 tetraketones were evaluated with the LDA-based QSAR models, and the *in vitro* assays of the synthesized compounds were done to corroborate the *in silico* predictions. The values of the posterior classification probabilities ( $\Delta P\%$ ) obtained with all the equations for the data are shown in Table 7. Hence here we can see that exits a good concordance among the theoretical predictions and the experimental results for all the organic-chemicals, and all were active against the tyrosinase enzyme in the *in vitro* assays. It is important to stand out that the majority of compounds showed values of activity higher than Kojic acid (standard tyrosinase inhibitor:  $IC_{50} = 16.67 \mu M$ ) with the exception of **TK2** ( $IC_{50} = 26.63 \mu M$ ), **TK4** ( $IC_{50} = 16.99 \mu M$ ), **TK7** ( $IC_{50} = 19.73 \mu M$ ) and **TK19** ( $IC_{50} = 71.47 \mu M$ ). By other way, four chemicals **TK10** ( $IC_{50} = 2.09 \mu M$ ),



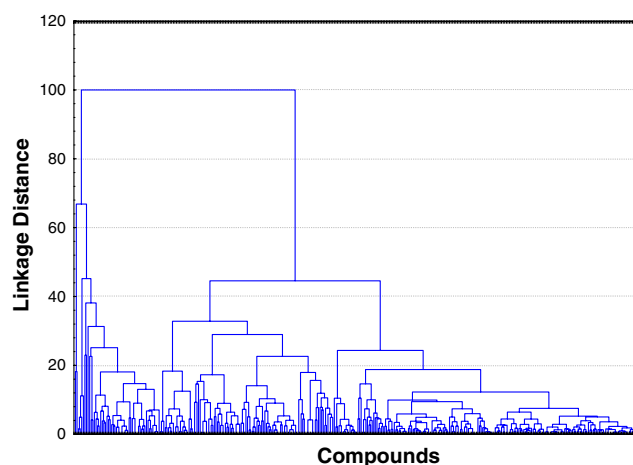
**Table 8** Molecular structure of the new tetraketones

Compound	X	Y	IC <sub>50</sub> (μM)	Compound	X	Y	IC <sub>50</sub> (μM)
<b>TK1</b>	-H		6.55	<b>TK13</b>	-CH <sub>3</sub>		14.58
<b>TK2</b>	-H		26.63	<b>TK14</b>	-CH <sub>3</sub>		9.06
<b>TK3</b>	-H		12.31	<b>TK15</b>	-CH <sub>3</sub>		3.70
<b>TK4</b>	-H		16.99	<b>TK16</b>	-CH <sub>3</sub>		12.82
<b>TK5</b>	-H		11.77	<b>TK17</b>	-CH <sub>3</sub>		15.36
<b>TK6</b>	-H		4.83	<b>TK18</b>	-CH <sub>3</sub>		6.59
<b>TK7</b>	-H		19.73	<b>TK19</b>	-CH <sub>3</sub>		71.47
<b>TK8</b>	-H		4.80	<b>TK20</b>	-CH <sub>3</sub>		4.01
<b>TK9</b>	-H		6.77	<b>TK21</b>	-CH <sub>3</sub>		2.06
<b>TK10</b>	-H		2.09	<b>TK22</b>	-CH <sub>3</sub>		13.40
<b>TK11</b>	-CH <sub>3</sub>		2.61	<b>TK23</b>	-CH <sub>3</sub>		3.19
<b>TK12</b>	-CH <sub>3</sub>		4.13	<b>TK24</b>	-CH <sub>3</sub>		12.68

As a final point, a hierarchical cluster analysis was performed for all the active compounds of the training, test, virtual screening and the new tetraketones (Fig. 8). The aim of this  $k$ -NNCA was compared if there was any similarity between the novel bioactive chemicals and some subsystems in the rest of the active database. After an exhaustive analysis to each cluster, we observed that these tetraketones were distributed in many clusters, which is reasonable because this class of compounds don't have common structural features with none of the compounds in the active database. Therefore, they can be selected to make a structural optimization with the objective to find a more potent tyrosinase inhibitory activity, and afterward a complete study of ADMET properties should be carried out to entering these organic-chemicals discovered into the pipeline of the drug market development.



**Fig. 7** A dendrogram illustrating the results of the hierarchical  $k$ -NNCA of the set of active/inactive chemicals used for evaluating the predictive ability of the QSAR models for ligand-based virtual screening



**Fig. 8** A dendrogram illustrating the results of the hierarchical  $k$ -NNCA of the set of all active chemicals (tyrosinase inhibitors) included in training, test, virtual screening and new active tetraketones discovery in the present work

## Summary and outlook

Many studies in the area of tyrosinase inhibitory activity are involved to finding novel inhibitors from different sources, due to its wide applications as food additives, depigmentation agents, in the treatment of melanogenesis disorders, to control insect pests and so on. The interest of pharmaceutical, cosmetic, and agricultural sciences in this kind of chemicals is referred to its broad spectrum of applications, and wide distribution through all the phylogenetic scale.

The advent of virtual High Through Screening (vHTS) encompassing *in silico* techniques in the drug discovery, are solutions that enable research to proceed faster and more efficiently. These new algorithms starting from the convergence of information technology and drug discovery, can be useful to resolve the question of accelerate the pace of drug discovery in the identification of higher quality compounds. Nevertheless, in this case, the process of searching of new tyrosinase inhibitor compounds until now is through trial-error traditional methods [84, 85].

Taken all these into consideration, we made use of the *non-stochastic and stochastic bond-based bilinear indices*, a new set of MDs, together with pattern recognition techniques to discriminate active compounds from inactive ones. QSAR models found here were used in a virtual screening to arising from the *in silico* to 'real' world applications. Besides, is reported the *biosilico* identification of a novel tetraketone family as tyrosinase inhibitors using the new molecular fingerprints. The experimental *in vitro* assays were also carried out to prove the usefulness of the **TOMO-COMD-CARDD** descriptors for the rational design of new bioactive agents.

These kinds of works are in the light of new challenges for the pharmaceutical industries because a research in modern drug discovery needs training and experience in multiple life science domain areas as well as in computer science [86]. Finally, the present report could permit us to look forward to many exciting new insights in the field of tyrosinase inhibitor compounds research for the treatment of hyperpigmentation and melanogenesis disorders in the years ahead.

## Supporting information available

The complete list of compounds used in training and prediction sets, as well as their structures, posterior classification and scores according to LDA-based QSAR models, chemistry and data analysis of the

obtained chemicals is available free of charge via Internet at ...

**Acknowledgments** One of the authors (M-P. Y) thanks the program 'Estades Temporals per a Investigadors Convidats' for a fellowship to work at Valencia University (2006–2007). M-P. Y thanks are also given to the Generalitat Valenciana, (Spain) for partial financial support as well as support from Spanish MEC (Project Reference: SAF2006-04698). MTHK is the recipient of a grant from MCBN-UNESCO (grant no. 1056), and fellowships from CIB (Italy) and Associazione Veneta per la Lotta alla Talassemia (AVTL, Italy). 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).

## References

- Robb DA (1984) Copper proteins and copper enzymes. CRC Press, Boca Raton, FL
- Baurin N, Arnoult E, Scior T, Do QT, Bernard P (2002) J Ethnopharmacol 82:155
- Prota G (1988) Med Res Rev 8:525
- Prota G (1992) Melanins and melanogenesis. Academic Press, San Diego, CA
- Riley PA (1996) In: Hori Y, Hearing VJ, Nakayama J (eds) Melanogenesis and malignant melanoma: biochemistry, cell biology, molecular biology, pathophysiology, diagnosis and treatment. Elsevier, Amsterdam
- Riley PA (2003) Pigment Cell Res 16:548
- Fitzpatrick TB, Seiji M, McGugan AD (1961) New Engl J Med 265:374
- Maeda K, Fukuda M (1996) J Pharmacol Exp Ther 276:765
- Li W, Kubo I (2004) Bioorg Med Chem 12:701
- Casañola-Martin GM, Khan MT, Marrero-Ponce Y, Ather A, Sultankhodzhaev MN, Torrens F (2006) Bioorg Med Chem Lett 16:324
- Marrero-Ponce Y, Khan MTH, Casañola-Martín GM, Ather A, Sultankhodzhaev MN, Torrens F (2006) QSAR Comb Sci DOI: 10.1002/qsar.200610156
- Marrero-Ponce Y (2003) Molecules 8:687
- Marrero-Ponce Y, Huesca-Guillen A, Ibarra-Velarde F (2005) J Mol Struct (Theochem) 717:67
- Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, Montero-Torres A, Pereira DM, Escario JA, Nogal-Ruiz JJ, Ochoa C, Aran VJ, Martinez-Fernandez AR, Garcia Sanchez RN (2005) Bioorg Med Chem Lett 15:3838
- Vega MC, Montero-Torres A, Marrero-Ponce Y, Rolon M, Gomez-Barrio A, Escario JA, Aran VJ, Nogal JJ, Meneses-Marcel A, Torrens F (2006) Bioorg Med Chem Lett 16:1898
- Marrero-Ponce Y, Medina-Marrero R, Martinez Y, Torrens F, Romero-Zaldivar V, Castro EA (2006) J Mol Mod 12:255
- Marrero-Ponce Y, Nodarse D, González HD, Ramos de Armas R, Romero-Zaldivar V, Torrens F, Castro E (2004) Int J Mol Sci 5:276
- Marrero Ponce Y, Castillo Garit JA, Nodarse D (2005) Bioorg Med Chem 13:3397
- Marrero-Ponce Y, Iyarreta-Veitia M, Montero-Torres A, Romero-Zaldivar V, Brandt CA, Avila PE, Kirchgatter K, Machado Y (2005) J Chem Inf Model 45:1082
- Marrero-Ponce Y, Romero V, TOMOCOMD software, Central University of Las Villas (2002) TOMOCOMD (TOPological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version can be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es
- Rouvray DH (1976) In: Balaban AT (ed) Chemical applications of graph theory. Academic Press, London, pp 180–181
- Trinajstić N (1983) Chemical graph theory. CRC Press, Boca Raton FL
- Estrada E (1995) J Chem Inf Comput Sci 35:31
- Estrada E, Ramirez A (1996) J Chem Inf Comput Sci 36:837
- Estrada E (1996) J Chem Inf Comput Sci 36:844
- Estrada E, Guevara N, Gutman I (1998) J Chem Inf Comput Sci 38:428
- Estrada E (1999) J Chem Inf Comput Sci 39:1042
- Estrada E, Molina E (2001) J Mol Graph Model 20:54
- Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Wiley-VCH, Germany
- Ivanciuc O, Balaban AT (1999) In: Devillers J, Balaban AT (eds) Topological indices and related descriptors in QSAR and QSPR. Gordon and Breach, The Netherlands, 73 p
- Edwards CH, Penney DE (1988) Elementary linear algebra. Prentice-Hall, Englewood Cliffs, New Jersey, USA
- Marrero-Ponce Y (2004) Bioorg Med Chem 12:6351
- Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castañedo N, Ibarra-Velarde F, Huesca-Guillen A, Jorge E, del Valle A, Torrens F, Castro EA (2004) J Comput Aided Mol Des 18:615
- Marrero-Ponce Y, Medina-Marrero R, Torrens F, Martinez Y, Romero-Zaldivar V, Castro EA (2005) Bioorg Med Chem 13:2881
- Marrero-Ponce Y, Díaz HG, Romero V, Torrens F, Castro EA (2004) Bioorg Med Chem 12:5331
- Marrero-Ponce Y, Cabrera MA, Romero V, Ofori E, Montero LA (2003) Int J Mol Sci 4:512
- Marrero-Ponce Y, Cabrera MA, Romero V, González DH, Torrens F (2004) J Pharm Pharmaceut Sci 7:186
- Marrero-Ponce Y, Cabrera MA, Romero-Zaldivar V, Bermejo M, Siverio D, Torrens F (2005) Internet Electron J Mol Des 4:124
- Marrero-Ponce Y, Medina R, Castro EA, de Armas R, González H, Romero V, Torrens F (2004) Molecules 9:1124
- Marrero Ponce Y (2004) J Chem Inf Comput Sci 44:2010
- Marrero-Ponce Y, Castillo-Garit JA, Torrens F, Romero-Zaldivar V, Castro E (2004) Molecules 9:1100
- Marrero-Ponce Y, Montero-Torres A, Zaldivar CR, Veitia MI, Perez MM, Sanchez RN (2005) Bioorg Med Chem 13:1293
- Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castanedo N, Ibarra-Velarde F, Huesca-Guillen A, Sanchez AM, Torrens F, Castro EA (2005) Bioorg Med Chem 13:1005
- Marrero-Ponce Y, Castillo-Garit JA (2005) J Comput Aided Mol Des 19:369
- Estrada E, Vilar S, Uriarte E, Gutierrez Y (2002) J Chem Inf Comput Sci 42:1194
- Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E (2000) J Med Chem 43:1975
- Estrada E, Peña A, Garcia-Domenech R (1998) J Comput Aided Mol Des 12:583
- Potapov VM (1978) Stereochemistry. Mir Moscow
- Wang R, Gao Y, Lai L (2000) Perspect Drug Dis Des 19:47
- Ertl P, Rohde B, Selzer P (2000) J Med Chem 43:3714
- Ghose AK, Crippen GM (1987) J Chem Inf Comput Sci 27:21
- Miller KJ (1990) J Am Chem Soc 112:8533
- Gasteiger J, Marsili M (1978) Tetrahedron Lett 19:3181

54. Pauling L (1939) The nature of chemical bond. Cornell University Press, Ithaca, New York
55. Kier LB, Hall LH (1986) Molecular connectivity in structure–activity analysis. Research Studies Press, Letchworth, UK
56. Consonni V, Todeschini R, Pavan M (2002) *J Chem Inf Comput Sci* 42:682
57. Todeschini R, Gramatica P (1998) *Perspect Drug Dis Des* 9–11:355
58. Jacobson N (1985) In: Freeman WHC (ed) *Basic algebra I*. New York, pp 343–361
59. Riley KF, Hobson MP, Vence SJ (1998) *Mathematical methods for physics and engineering*. Cambridge University Press
60. Werner G (1981) *Linear algebra*, 4th edn Springer-Verlag, New York
61. Walker PD, Mezey PG (1993) *J Am Chem Soc* 115:12423
62. Klein DJ (2003) *Internet Electron J Mol Des* 2:814
63. Negwer M (1987) *Organic-chemical drugs and their synonyms*. Akademie-Verlag, Berlin
64. STATISTICA (data analysis software system), v S I (2001) [www.statsoft.com](http://www.statsoft.com)
65. Xu J, Hagler A (2002) *Molecules* 7:566
66. Mc Farland JW, Gans DJ (1995) In: Waterbeemd H (ed), *Chemometric Methods in Molecular Design*. VCH Publishers, New York, pp 295–307
67. Johnson RA, Wichern DW (1988) *Applied multivariate statistical analysis*. Prentice-Hall, New Jersey
68. Duart MJ, Garcia-Domenech R, Anton-Fos GM, Galvez J (2001) *J Comput Aided Mol Des* 15:561
69. van de Waterbeemd H (1995) In: van de Waterbeemd H (ed) *Chemometric methods in molecular design*. VCH Publishers, Weinheim, pp 265–288
70. de Julian-Ortiz JV, de Alapont CG, Ríos-Santamarina I, Garcia-Domenech R, Galvez E (1998) *J Mol Graphics Mod* 16:14
71. Gálvez J, García R, Salabert MT, Soler R (1994) *J Chem Inf Comput Sci* 34:520
72. Gonzales-Diaz H, Marrero Ponce Y, Hernandez I, Bastida I, Tenorio E, Nasco O, Uriarte E, Castanedo N, Cabrera MA, Aguila E, Marrero O, Morales A, Perez M (2003) *Chem Res Toxicol* 16:1318
73. Estrada E, Peña A (2000) *Bioorg Med Chem* 8:2755
74. Randić M (1991) *J Mol Struct (Theochem)* 233:45
75. Randić M (1991) *J Chem Inf Comput Sci* 31:311
76. Randić M (1991) *New J Chem* 15:517
77. Lučić B, Nikolić S, Trinajstić N, Jurić D (1995) *J Chem Inf Comput Sci* 35:532
78. Klein DJ, Randić M, Babić D, Lučić B, Nikolić S, Trinajstić N (1997) *Int J Quantum Chem* 63:215
79. Estrada E, Uriarte E (2001) *Curr Med Chem* 8:1573
80. Khan KM, Maharvi GM, Khan MT, Jabbar Shaikh A, Perveen S, Begum S, Choudhary MI (2006) *Bioorg Med Chem* 14:344
81. Hearing VJ (1987) *Methods in enzymology*. Academic Press, New York
82. Wold S, Erikson L (1995) In: van de Waterbeemd H (ed) *Chemometric methods in molecular design*. VCH Publishers, New York, pp 309–318
83. Golbraikh A, Tropsha A (2002) *J Mol Graph Model* 20:269
84. Okombi S, Rival D, Bonnet S, Mariotte AM, Perrier E, Boumendjel A (2006) *J Med Chem* 49:329
85. Zhang J-P, Chen Q-X, Song K-K, Xie J-J (2006) *Food Chem* 92:579
86. Apweiler R (2003) *Biosilico* 1:5