Prediction of Tyrosinase Inhibition Activity Using Atom-Based Bilinear Indices

Yovani Marrero-Ponce,*[a, b] Mahmud Tareq Hassan Khan,[c, d] Gerardo M. Casañola Martín,[b, e] Arjumand Ather,[f] Mukhlis N. Sultankhodzhaev,[g] Francisco Torrens,[a] and Richard Rotondo[h]

A set of novel atom-based molecular fingerprints is proposed based on a bilinear map similar to that defined in linear algebra. These molecular descriptors (MDs) are proposed as a new means of molecular parametrization easily calculated from 2D molecular information. The nonstochastic and stochastic molecular indices match molecular structure provided by molecular topology by using the kth nonstochastic and stochastic graph-theoretical electronic-density matrices, M^k and S^k, respectively. Thus, the kth nonstochastic and stochastic bilinear indices are calculated using M^k and S^k as matrix operators of bilinear transformations. Chemical information is coded by using different pair combinations of atomic weightings (mass, polarizability, vdW volume, and electronegativity). The results of QSAR studies of tyrosinase inhibitors using the new MDs and linear discriminant analysis (LDA) demonstrate the ability of the bilinear indices in testing biological properties. A database of 246 structurally diverse tyrosinase inhibitors was assembled. An inactive set of 412 drugs with other clinical uses was used; both active and inactive sets were processed by hierarchical and partitional cluster analyses to design training and predicting sets. Twelve LDA-based QSAR models were obtained, the first six using the nonstochastic total and local bilinear indices and the last six with the stochastic MDs. The discriminant models were applied; globally good classifications of 99.58 and 89.96 % were observed for the best nonstochastic and stochastic bilinear indices models in the training set along with high Matthews correlation coefficients (C) of 0.99 and 0.79, respectively, in the learning set. External prediction sets used to validate the models obtained were correctly classified, with accuracies of 100 and 87.78 %, respectively, yielding C values of 1.00 and 0.73. This subset contains 180 active and inactive compounds not considered to fit the models. A simulated virtual screen demonstrated this approach in searching tyrosinase inhibitors from compounds never considered in either training or predicting series. These fitted models permitted the selection of new cycloartane compounds isolated from herbal plants as new tyrosinase inhibitors. A good correspondence between theoretical and experimental inhibitory effects on tyrosinase was observed; compound **CA6** ($IC_{50} = 1.32$ μM) showed higher activity than the reference compounds kojic acid ($IC_{50} = 16.67$ μM) and L-mimosine ($IC_{50} = 3.68$ μM).

"Indeed, we are still waiting for the rise of computational drug discovery."

R. Apweiler, *BioSilico* **2003**, *I*, 5–6.

1. Introduction

Melanin synthesis, the main process involved in skin pigmentation, is largely regulated by the melanogenic enzyme tyrosinase.[1] This is a bifunctional enzyme that catalyzes the hydroxylation of tyrosine to DOPA and promotes the oxidation of DOPA to DOPA quinone.[2] Its role is to protect the skin from ultraviolet (UV) radiation damage caused by sunlight and to remove reactive oxygen species (ROS).[3] Notwithstanding, disturbances in the amount and distribution of melanin pigments might ultimately provide clues to several diseases. Albinism is a genetic abnormality caused by a deficiency in melanin bio-

[a] Prof. Dr. Y. Marrero-Ponce, Prof. Dr. F. Torrens
Institut Universitari de Ciència Molecular
Universitat de València, Edifici d'Instituts de Paterna
Poligon la Coma s/n (detras de Canal Nou) P.O. Box 22085
46071 Valencia (Spain)
Fax: (+ 53) 42-281130 or 42-281455 (Cuba) and
Fax: (+ 96) 354-3156 (Spain)
E-mail: yovanimp@qf.uclv.edu.cu

[b] Prof. Dr. Y. Marrero-Ponce, Dr. G. M. Casañola Martín
Unit of Computer-Aided Molecular "BioSilico" Discovery and
Bioinformatic Research (CAMD-BIR Unit), Department of Pharmacy
Faculty of Chemistry–Pharmacy and Department of Drug Design
Chemical Bioactive Center
Central University of Las Villas, Santa Clara, 54830 Villa Clara (Cuba)

[c] Prof. Dr. M. T. H. Khan
Pharmacology Research Laboratory, Faculty of Pharmaceutical Sciences
University of Science and Technology, Chittagong (Bangladesh)

[d] Prof. Dr. M. T. H. Khan
Department of Pharmacology, Institute of Medical Biology
University of Tromsø, 9037 Tromsø (Norway)

[e] Dr. G. M. Casañola Martín
Department of Biological Sciences, Faculty of Agricultural Sciences
University of Ciego de Avila, 69450 Ciego de Avila (Cuba)

[f] Prof. Dr. A. Ather
The Norwegian Structural Biology Centre (NorStruct)
University of Tromsø, 9037 Tromsø (Norway)

[g] M. N. Sultankhodzhaev
S. Yunusov Institute of Chemistry of Plant Substances
Academy of Sciences, Uzbekistan (Tashkent)

[h] R. Rotondo
Advanced Medisyns, Inc.
601 Carlson Parkway, Suite 1050, Minnetonka, MN 55305 (USA)

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

synthesis which manifests as hypopigmentation of the skin, hair, and eyes.[4] However, abnormal accumulation of melanin pigments is responsible for hyperpigmentations including melasma, freckles, and senile lentigines, which show satisfactory subjective improvement upon treatment with de-pigmenting agents such as hydroquinone (carcinogenic), ascorbic acid derivatives, azelaic acid, retinoids, arbutin, and kojic acid.[5–8] However, many of the most popular de-pigmenting agents in use today exhibit toxicity toward melanocytes and are known to produce adverse side effects.[9–11] Inhibitors that target tyrosinase, the rate-limiting enzyme in melanin production, promise to be safer alternatives than melanocytolytic compounds.[12] Therefore, compounds isolated from plant extracts that have an inhibitory effect on melanin formation may be good candidates for this purpose because of their relatively low side effects.[13]

In this context, one of the efforts in our research group has been focused on finding new potent tyrosinase inhibitors through trial-and-error techniques. Recently, Khan et al. reported 2,5-disubstituted-1,3,4-oxadiazole analogues[14] with strong inhibitory activity against the enzyme. In another publication, the same research group reported that (+)-androst-4-ene-3,17-dione as well as its five metabolic analogues having steroidal skeletons, namely androsta-1,4-diene-3,17-dione, 17β-hydroxy-androsta-1,4-dien-3-one, 11α-hydroxyandrost-4-ene-3,17-dione, 11α,17β-dihydroxyandrost-4-en-3-one, and 15α-hydroxyandrosta-1,4-dien-17-one, exhibited moderate inhibitory activities against tyrosinase.[15] In 2004, Ahmad et al.[16] reported that a new coumarinolignoid, 8'-*epi*-cleomiscosin A, together with the new glycoside, 8-*O*-β-D-glucopyranosyl-6-hydroxy-2-methyl-4*H*-1-benzopyrane-4-one exhibited strong inhibition against tyrosinase relative to the standard tyrosinase inhibitors kojic acid and L-mimosine. The new coumarinolignoid exhibited twofold greater potency than that of the standard potent inhibitor L-mimosine.[16]

Because the experimental tests (based on trial-and-error screening) that must be performed, especially pharmacological and toxicological tests, are usually expensive and time consuming, the pharmaceutical industry has reoriented its research strategy over the past two decades toward the development of methods that enable the rational selection or design of novel agents with desired properties.[17] To decrease costs, pharmaceutical companies must find new technologies to replace the old "hand-crafted" synthesis and test new chemical entities (NCE) approaches.[18] In this sense, cheminformatics can be used to analyze data from high-throughput screening (HTS) and other forms of chemistry, thereby aiding the identification of optimal lead structures.[19]

Virtual screening techniques may be classified according to their particular modeling of molecular recognition and the type of algorithm used in database searching.[18, 19] If the target 3D structure is known (or at least the active site), one of the structure-based virtual screening methods can be applied. These methods are based on the principle of complementarity: the receptor of a biologically active compound is complementary to the compound itself (the lock-and-key model).[18, 19] On the other hand, ligand-based methods are related to the principle of similarity; that is, similar compounds are assumed to produce similar effects. In this case, if one or more active compounds are known, it is possible to search a database for similar but more potent molecules. The selection of the method depends on the knowledge of the active molecules and their receptor. In the case of tyrosinase, its NMR structure is still not available, thus limiting structure-based methods.[20]

Therefore, ligand-based methods are appropriate for this case because many tyrosinase inhibitors are known; they can establish QSAR (quantitative structure–activity relationship) models that describe the biological activity of new compounds and that predict their ability to inhibit tyrosinase activity. In addition, if computational approaches based on discrimination functions are used, it is possible to distinguish between active and inactive compounds, and to predict the biological activity of new leads. Therefore, cheminformatic in silico methods appear to be particularly rewarding in terms of both cost and time benefits and are easily integrated into the modern drug-discovery process.[21, 22] Moreover, several authors have reported a high incidence in the use of novel molecular descriptors to develop QSAR studies for in silico virtual drug screening.[21–25] The definition of novel molecular descriptors is a promising field in medicinal chemistry as well as the veterinary, agricultural, and pharmaceutical sciences.

Recently, a novel scheme for the rational in silico molecular design (or selection and identification of compounds) and for QSAR/QSPR studies was introduced by others of our research team: TOpological MOlecular COMputer Design–Computer-Aided Rational Drug Design (TOMOCOMD–CARDD).[26] This approach, based on principles of novel methods for on chemical graph and algebraic theory, has been successfully used in the description of different physical, physicochemical, and chemical properties of organic compounds.[27–29] The prediction of many biological activities were also effectively modeled with TOMOCOMD–CARDD descriptors,[30–33] including studies related to proteomics,[34, 35] nucleic acid–drug interactions,[36, 37] and the fast-track discovery of novel antimalarial compounds.[38]

Herein we propose a novel set of molecular descriptors (MDs), namely nonstochastic and stochastic bilinear indices. Other aspects presented include the use of these new MDs and a linear discriminant analysis (LDA) strategy to find quantitative models that discriminate tyrosinase inhibitors from inactive compounds. We also describe an experiment of ligand-based virtual screening that was carried out to simulate the discovery of new leads from a database of marketable drugs. As a final point, the in silico identification, isolation, and pharmacological testing of new hits and lead compounds are presented.

2. Methods

In earlier work, we outlined outstanding features concerned with the theory of 2D atom-based TOMOCOMD–CARDD descriptors. In this case, the atom, atom type, and total bilinear indices of the molecular pseudograph's atom adjacency matrix for small-to-medium sized organic compounds are explained

in some detail elsewhere.[39,40] However, an overview of this approach is given below.

This method codes molecular structure by means of mathematical quadratic, linear, and bilinear transformations. To calculate these algebraic maps for a molecule, the atom-based molecular vector, $\bar{x}$ (vector representation), and kth "nonstochastic and stochastic graph-theoretical electronic-density matrices" M^k and S^k (respective matrix representations) are constructed.[27–33,38,41–49] Such atom-adjacency relationships and chemical information coding are applied in this study to generate a series of atom-based MDs: atom, group, atom-type, and total bilinear indices to be used in drug design and chemoinformatic studies.

Therefore, the structure of this section is as follows: 1) a background in atom-based molecular vector as well as nonstochastic and stochastic graph-theoretical electronic-density matrices are described respectively in subsections 2.1 and 2.2, and 2) an outline of the mathematical definition of bilinear maps and a definition of our procedures are developed correspondingly in subsections 2.3 and 2.4.

2.1. Chemical information and atom-based molecular vector

The atom-based molecular vector ($\bar{x}$) used to represent small-to-medium-sized organic compounds has been explained elsewhere in some detail.[17,27–33,38,41–49] The components (x) of $\bar{x}$ are numerical values, which represent a certain standard atomic property (atom label). Therefore, these weights correspond to different atomic properties for organic molecules. Thus, a molecule with 5, 10, 15,…, n atomic nuclei can be represented by vectors with 5, 10, 15,…, n components, respectively, belonging to the spaces $\Re^5$, $\Re^{10}$, $\Re^{15}$,…, $\Re^n$, for which n is the dimension of the real set ($\Re^n$). That is to say, $\bar{x}$ is the n-dimensional vector property of the atoms (atomic nuclei) in a molecule.

This approach allows us to encode organic molecules such as 3-sulfanylisonicotinaldehyde through the molecular vector $\bar{x} = [x_{N1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}, x_{C6}, x_{C7}, x_{O8}, x_{S9}]$ (see also Table 1 for molecular structure). This vector belongs to the product space $\Re^9$. However, diverse kinds of atomic weights (x) can be used for coding information related to each atomic nucleus in the molecule. These atomic labels are chemically meaningful numbers or their contributions derived by atom-to-atom analysis such as atomic $\log P_i$,[50] surface contributions of polar atoms,[51] atomic molar refractivities,[52] atomic hybrid polarizabilities,[53] Gasteiger–Marsilli atomic charge,[54] atomic masses (M),[55] van der Waals volumes (V),[55] atomic polarizabilities (P),[55] atomic electronegativities (K) in Mulliken scale,[54] and so on.

If one is interested in coding the chemical information by means of two different molecular vectors, for instance, $\bar{x} = [x_1,…,x_n]$ and $\bar{y} = [y_1,…,y_n]$, then different combinations of molecular vectors ($\bar{x} \neq \bar{y}$) are possible if a weighting scheme is used. In the study reported herein, we characterized each atomic nucleus with the following parameters: atomic mass (M),[55] van der Waals volume (V),[55] atomic polarizability (P),[55] and atomic Mulliken electronegativity (K).[54] The values of these atomic labels are shown in Table 2. From this weighting scheme, six (or twelve if $\bar{x}_M - \bar{y}_V \neq \bar{x}_V - \bar{y}_M$) combinations (pairs)

of molecular vectors ($\bar{x}$, $\bar{y}$; $\bar{x} \neq \bar{y}$) can be computed: $\bar{x}_M - \bar{y}_V$, $\bar{x}_M - \bar{y}_P$, $\bar{x}_M - \bar{y}_K$, $\bar{x}_V - \bar{y}_P$, $\bar{x}_V - \bar{y}_K$, and $\bar{x}_P - \bar{y}_K$. Here, we used the symbols $\bar{x}_W - \bar{y}_Z$, for which the subscripts $_W$ and $_Z$ mean two different atomic properties from our weighting scheme, and a dash ($-$) expresses the combination (pair) of two selected atom-label chemical properties. To illustrate this, let us consider the same organic molecule as in the example above (3-sulfanylisonicotinaldehyde) and the following weighting scheme: M and V ($\bar{x}_M - \bar{y}_V = \bar{x}_V - \bar{y}_M$). The following molecular vectors, $\bar{x} = [14.01, 12.01, 12.01, 12.01, 12.01, 12.01, 12.01, 16.0, 32.07]$ and $\bar{y} = [15.599, 22.449, 22.449, 22.449, 22.449, 22.449, 22.449, 11.494, 24.429]$ are obtained if we use M and V as chemical weights for coding each atom in the example molecule in $\bar{x}$ and $\bar{y}$ vectors, respectively.

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

In molecular topology, molecular structure is generally expressed by the hydrogen-suppressed graph, and therefore, a molecule is represented by a graph. Informally, a graph G is a collection of vertices (points) and edges (lines or bonds) connecting these vertices.[56–58]

Earlier we introduced new molecular matrices that describe time-dependent changes in the electronic distribution throughout the molecular backbone. The $n \times n$ kth nonstochastic graph-theoretical electronic-density matrix of the molecular pseudograph (G), M^k, is a symmetric and square matrix for which n is the number of atoms (atomic nuclei) in the molecule.[17,27–33,38,41–49] The coefficients $^k m_{ij}$ are the elements of the kth power of $M(G)$ and are defined as follows:

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

where E(G) represents the set of edges of G. P_{ij} is the number of edges (bonds) between vertices (atomic nuclei) v_i and v_j, and L_{ii} is the number of loops in v_i. The elements $m_{ij} = P_{ij}$ of such a matrix represent the number of chemical bonds between an atomic nucleus i and other j. The matrix M^k provides the numbers of walks of length k that link every pair of vertices v_i and v_j. For this reason, each edge in M^1 represents two electrons belonging to the covalent bond between atomic nuclei i and j; for example, the inputs of M^1 are equal to 1, 2, or 3 when single, double, or triple bonds, correspondingly, appear between vertices v_i and v_j. On the other hand, molecules containing aromatic rings with more than one canonical structure are represented by pseudographs. This happens for substituted aromatic compounds such as pyridine, naphthalene, and quinoline, where the presence of π electrons is accounted by means of loops in each atomic nucleus of the aromatic ring. Conversely, aromatic rings with only one canonical structure, such as furan, thiophene, and pyrrole are represented by a multigraph. To illustrate the calculation of these matrices, let us consider the same molecule selected in the previous

Table 1. A) Chemical structure of 3-sulfanylisonicotinaldehyde and its labeled molecular pseudograph, G; the zero (k=0), first (k=1), second (k=2), and third (k=3) powers of the B) nonstochastic and C) stochastic graph-theoretical electronic-density matrices of G.

A)

Structure

H-atom-suppressed pseudograph[a]

vertices { ● heteroatoms / ○ C atoms }

B) kth nonstochastic graph-theoretical electronic-density matrices, M^k (k=0–3)

(k=0), (k=1), (k=2), (k=3) — matrices

(k=3):
$$
\begin{bmatrix}
7 & 6 & 3 & 2 & 3 & 6 & 0 & 0 & 1 \\
6 & 7 & 7 & 3 & 2 & 3 & 1 & 0 & 2 \\
3 & 7 & 9 & 8 & 3 & 2 & 2 & 2 & 4 \\
3 & 2 & 3 & 8 & 7 & 3 & 8 & 2 & 2 \\
3 & 2 & 3 & 7 & 7 & 6 & 2 & 2 & 2 \\
6 & 3 & 2 & 3 & 6 & 7 & 1 & 1 & 0 \\
0 & 1 & 2 & 8 & 2 & 1 & 1 & 10 & 1 \\
0 & 0 & 2 & 2 & 2 & 0 & 10 & 0 & 1 \\
1 & 2 & 4 & 2 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
$$

C) stochastic graph-theoretical electronic-density matrices, S^k (k=1–3)[b,c]

(k=3):
$$
\begin{bmatrix}
0.25 & 0.2142 & 0.1071 & 0.0714 & 0.1071 & 0.2143 & 0 & 0 & 0.0357 \\
0.1935 & 0.2258 & 0.2258 & 0.0967 & 0.0645 & 0.0967 & 0.0323 & 0 & 0.0645 \\
0.075 & 0.175 & 0.225 & 0.2 & 0.075 & 0.05 & 0.05 & 0.05 & 0.1 \\
0.0455 & 0.0682 & 0.1818 & 0.2045 & 0.1591 & 0.0682 & 0.1818 & 0.0455 & 0.0455 \\
0.0909 & 0.0606 & 0.0909 & 0.2121 & 0.2121 & 0.1818 & 0.0606 & 0.0606 & 0.0303 \\
0.2143 & 0.1071 & 0.0714 & 0.1971 & 0.2143 & 0.25 & 0.0357 & 0.0357 & 0 \\
0 & 0.0385 & 0.0714 & 0.3076 & 0.0769 & 0.0385 & 0.0385 & 0.3846 & 0.0385 \\
0 & 0 & 0.125 & 0.125 & 0.125 & 0 & 0.625 & 0 & 0 \\
0.0833 & 0.1666 & 0.333 & 0.1666 & 0.0833 & 0.0833 & 0 & 0 & 0.0833
\end{bmatrix}
$$

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

Table 2. Values of the atomic weights used for bilinear indices calculation.[55,72–74]

ID	Atomic Mass [Da]	V_{vdW} [Å³][a]	P [Å³][b]	K[c]
H	1.01	6.709	0.667	2.2
B	10.81	17.875	3.030	2.04
C	12.01	22.449	1.760	2.55
N	14.01	15.599	1.100	3.04
O	16.00	11.494	0.802	3.44
F	19.00	9.203	0.557	3.98
Al	26.98	36.511	6.800	1.61
Si	28.09	31.976	5.380	1.9
P	30.97	26.522	3.630	2.19
S	32.07	24.429	2.900	2.58
Cl	35.45	23.228	2.180	3.16
Fe	55.85	41.052	8.400	1.83
Co	58.93	35.041	7.500	1.88
Ni	58.69	17.157	6.800	1.91
Cu	63.55	11.494	6.100	1.9
Zn	65.39	38.351	7.100	1.65
Br	79.90	31.059	3.050	2.96
Sn	118.71	45.830	7.700	1.96
I	126.90	38.792	5.350	2.66

[a] van der Waals volume. [b] Polarizability. [c] Electronegativity.

section. Table 1 depicts the molecular structure of this compound and its labeled molecular pseudograph. The zero ($k=0$), first ($k=1$), second ($k=2$), and third ($k=3$) powers of the nonstochastic graph-theoretical electronic-density matrices are also given in this Table.

As can be observed, M^k are graph-theoretical electronic-structural models, like an "extended Hückel theory (EHT) model". The M^1 matrix considers all valence-bond electrons (σ and π networks) in one step, and its powers ($k=0, 1, 2, 3,…$) can be considered as interacting-electron chemical-network models in the k step. The complete model can be considered as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.[59]

The present approach is based on a simple model for the intramolecular movement of all outer-shell electrons. Let us consider a hypothetical situation in which a set of atoms is free in space at an arbitrary initial time (t_0). At this time, the electrons are distributed around the atomic nuclei. Alternatively, these electrons can be distributed around cores at discrete intervals of time t_k. In this sense, the electron in an arbitrary atom i can move (step-by-step) to other atoms at different discrete time periods t_k ($k=0, 1, 2, 3,…$) throughout the chemical-bonding network.

On the other hand, the kth stochastic graph-theoretical electronic-density matrix of G, S^k, can be directly obtained from M^k. Here, $S^k = [^k s_{ij}]$ is a square matrix of order n ($n=$number of atomic nuclei), and the elements $^k s_{ij}$ are defined as follows:[38,46,48,49]

$$^k s_{ij} = \frac{^k m_{ij}}{^k SUM_i} = \frac{^k m_{ij}}{^k \delta_i} \tag{2}$$

in which $^k m_{ij}$ are the elements of the kth power of M and the SUM of the ith row of M^k are named the k-order vertex

degree of atom i, $^k \delta_i$. Notably, the matrix S^k in Equation (2) has the property that the sum of the elements in each row is 1. An $n \times n$ matrix with nonnegative entries having this property is called a "stochastic matrix".[44] The kth s_{ij} elements are the transition probabilities with the electrons moving from atom i to j in the discrete time periods t_k. It should be also pointed out that the kth element s_{ij} takes into consideration the molecular topology in the k step throughout the chemical-bonding (σ and π) network. In this sense, the $^2 s_{ij}$ values can distinguish between hybrid states of atoms in bonds. For instance, the self-return probability of second order ($^2 s_{ij}$) (that is, the probability with which an electron returns to the original atom at t_2) varies regularly according to the different hybrid states of atom i in the molecule; for example, an electron will have a higher probability of returning to the sp C atom than to the sp² (or sp³) C atom in t_2 $[p(C_{sp}) > p(C_{sp}^2) > p(C_{sp\ arom}^2) > p(C_{sp}^3)]$ (see Table 1 for more details). This is a logical result if the electronegativity scale of these hybrid states is taken into account.

2.3. Mathematical bilinear forms: a theoretical framework

In mathematics, a bilinear form in a real vector space is a mapping $b : V \times V \to \Re$, which is linear in both arguments.[60–65] Therefore, this function satisfies the following axioms for any scalar α and any choice of vectors $\bar{v}$, $\bar{w}$, $\bar{v}_1$, $\bar{v}_2$, $\bar{w}_1$, and $\bar{w}_2$.

1. $b(\alpha \bar{v}, \bar{w}) = b(\bar{v}, \alpha \bar{w}) = \alpha b(\bar{v}, \bar{w})$
2. $b(\bar{v}_1 + \bar{v}_2, \bar{w}) = b(\bar{v}_1, \bar{w}) + b(\bar{v}_2, \bar{w})$
3. $b(\bar{v}, \bar{w}_1 + \bar{w}_2) = b(\bar{v}, \bar{w}_1) + b(\bar{v}, \bar{w}_2)$

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

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

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

and

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

Subsequently,

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

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

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

Then

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

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

Finally, we introduce the formal definition of symmetric bilinear form. Let V be a real vector space and b be a bilinear function in $V \times V$. The bilinear function b is called symmetric if $b(\bar{x}, \bar{y}) = b(\bar{y}, \bar{x}), \forall \bar{x}, \bar{y} \in V$.[60–65] Then,

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

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

The kth nonstochastic and stochastic bilinear indices for a molecule, $b_k(\bar{x}, \bar{y})$ and $^s b_k(\bar{x}, \bar{y})$,[66] respectively, are computed from these kth nonstochastic and stochastic graph-theoretical electronic-density matrices, M^k and S^k as shown in Equations (9) and (10):

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

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

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

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

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

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

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

because components of the molecular vectors are values of some atomic property that characterizes each kind of atomic nucleus in the molecule.

Notably, nonstochastic and stochastic bilinear indices are symmetric and nonsymmetric bilinear forms, respectively. Therefore, if in the following weighting scheme, M and V are used as atomic weights to compute these MDs, two different sets of stochastic bilinear indices, $^{M-V\,s}b_k{}^H(x,y)$ and $^{V-M\,s}b_k{}^H(x,y)$ [because $\bar{x}_M - \bar{y}_V \neq \bar{x}_V - \bar{y}_M$], can be obtained, and only one group of nonstochastic bilinear indices [$^{M-V\,s}b_k{}^H(x,y) = {}^{V-M\,s}b_k{}^H(x,y)$ because in this case, $\bar{x}_M - \bar{y}_V = \bar{x}_V - \bar{y}_M$] can be calculated.[66]

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

Several year ago, Randić[67] proposed a list of desirable attributes for a MD. This list can be considered as a methodological guide for the development of new topological indices (TIs). One of the most important criteria is the possibility of defining the descriptors locally. This attribute refers to the fact that the index could be calculated not only for the molecule as a whole, but also over certain fragments of the structure itself.

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

- Research on drugs, toxics, or generally any organic molecules with a common skeleton that is responsible for the activity or property under study.
- The study of the reactivity of specific sites of a series of molecules that can undergo a chemical reaction or enzymatic metabolism.
- The study of molecular properties such as spectroscopic measurements that are obtained experimentally in a local way.
- Any general case in which it is necessary to study not only the molecule as a whole, but also some local properties of certain fragments; then the definition of local descriptors could be necessary.

Therefore, in addition to *total bilinear indices* computed for the whole molecule,[66] a local-fragment (atomic, group, or atom-type) formalism can be developed. These descriptors are termed *local nonstochastic* and *stochastic bilinear indices*, $b_{kL}(\bar{x}, \bar{y})$ and $^s b_{kL}(\bar{x}, \bar{y})$, respectively.[66] The definition of these descriptors is as follows:

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

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

in which $^k m_{ijL}$ [$^k s_{ijL}$] is the kth element of the row "i" and column "j" of the local matrix M^k_L [S^k_L]. This matrix is extracted from the M^k [S^k] matrix and contains information that refers to the pairs of vertices (atomic nuclei) of the specific molecular fragments and also of the molecular environment in the k step. The matrix M^k_L [S^k_L] with elements $^k m_{ijL}$ [$^k s_{ijL}$] is defined as follows:

$^k m_{ijL}$ [$^k s_{ijL}$] = $^k m_{ij}$ [$^k s_{ij}$] *if* both v_i and v_j are atomic nuclei

contained within the molecular fragment,

$= \frac{1}{2}$ $^k m_{ij}$ [$^k s_{ijL}$] *if* either v_i or v_j is an atomic nucleus

contained   (15)

within the molecular fragment,

$= 0$ otherwise

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

$$b_L(\bar{x}, \bar{y}) = [X]^T M^k_L [Y] \tag{16}$$

$$^s b_L(\bar{x}, \bar{y}) = [X]^T S^k_L [Y] \tag{17}$$

It should be mentioned that the scheme above follows the spirit of a Mulliken population analysis of atomic net charges.[68] It should be also pointed out that for every partitioning of a molecule into Z molecular fragments, there will be Z local molecular fragmental matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrix M^k [S^k] can be correspondingly partitioned into Z local matrices M^k_L [S^k_L], $L = 1,\dots Z$, and the kth power of matrix M [S] is exactly the sum of the kth powers of the local Z matrices. In this way, the total nonstochastic and stochastic bilinear indices are the sum of the nonstochastic and stochastic bilinear indices, respectively, of the Z molecular fragments:

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

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

Atomic, group, and atom-type bilinear fingerprints are specific cases of local bilinear indices.[66] Atomic bilinear indices, $b_{kL}(\bar{x}_i, \bar{y}_i)$ and $^s b_{kL}(\bar{x}_i, \bar{y}_i)$, can be computed for each atom i in the molecule and contain electronic and topological structural information from all other atoms within the structure. The values of atom-level bilinear indices for the common scaffold atoms can be directly used as variables in seeking a QSPR/QSAR model, as long as these atoms are numbered in the same way in all molecules in the database.

In addition, the atom-type bilinear indices can also be calculated. In the same way as atom-type E-state values,[69] for all data sets (including those with a common skeletal core as well as those with very diverse structures), these novel local MDs provide much useful information. This approach therefore pro-

vides the basis for application to a wider range of problems to which the atomic bilinear indices formalism is applicable without the need for superposition.[70,71] For this reason, the present method represents a significant advantage over traditional QSAR methods. The atom-type bilinear descriptors are calculated by adding the kth atomic bilinear indices for all atoms of the same type in the molecule. This atom-type index lends itself to use in a group-additive-type scheme, in which an index appears for each atom type in the molecule. In the atom-type bilinear indices formalism, each atom in the molecule is classified into an atom type (fragment), such as $-F$, $-OH$, $=O$, $-CH_3$, and so on.[69–71] That is to say, each atom in the molecule is categorized according to a valence-state classification scheme including the number of attached H atoms.[69] The atom-type descriptors combine three important aspects of structural information: 1) collective electron and topological accessibility to the atoms of the same type (for each structural feature: atom or hybrid group such as $-Cl$, $=O$, $-CH_2-$, etc.), 2) the presence or absence of the atom type (structural features), and 3) a count of the atoms in the atom-type sets.

Finally, these local MDs can be calculated by a functional group in the molecule, such as heteroatoms (O, N, and S in all valence states as well as including the number of attached H atoms), hydrogen bonding to heteroatoms (O, N, and S in all valence states), halogen atoms (F, Cl, Br, and I), all aliphatic carbon chains (several atom types), all aromatic atoms (aromatic rings), and so on.[66] The group-level bilinear indices are the sum of the individual atom-level bilinear indices for a particular group of atoms. For all data set structures, the kth group-based bilinear indices provide important information for QSAR/QSPR studies.

2.6. Sample calculation
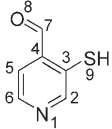
It is useful to perform a calculation on a molecule to illustrate the effect of structure on the values of atomic and global bilinear indices. Thus, we use the 3-sulfanylisonicotinaldehyde molecule. The labeled (atom numbering) molecular structure of this compound and the nonstochastic and stochastic (atom-level, group, and atom-type as well as total) atom-based bilinear indices are shown in Table 3.

3. Experimental Section

3.1. Computational strategies

Molecular fingerprints were generated by the interactive program for molecular design and bioinformatic research TOMO-COMD.[26] It is composed of four subprograms; each one allows both drawing the structures (drawing mode) and calculating molecular 2D/3D descriptors (calculation mode). The modules are named CARDD (Computer-Aided "Rational" Drug Design), CAMPS (Computer-Aided Modeling in Protein Science), CANAR (Computer-Aided Nucleic Acid Research) and CABPD (Computer-Aided Bio-Polymers Docking). The CARDD module was selected for drawing all structures and for the computation of nonstochastic and stochastic bilinear indices. The main steps

Table 3. Values of atom-based bilinear indices for 3-sulfanylisonicotinaldehyde.

	$k=0$	$k=1$	$k=2$	$k=3$	$k=8$	$k=15$
			*k*th Nonstochastic MDs, $b_{kL}(\bar{x}_i, \bar{y}_i)^{[a]}$			
Atom (*i*)			Atom-level bilinear indices			
N₁	17,1589	69,30704	207,92112	677,6482	217716,1986	711876800,9
C₂	39,51024	105,09455	382,76839	1173,27708	380330,8902	1242200889
C₃	39,51024	172,57929	475,22504	1557,467468	498904,5116	1625489636
C₄	39,51024	158,04096	526,894748	1699,867938	545797,7789	1784733323
C₅	39,51024	118,53072	381,66623	1237,280798	406509,5989	1327324524
C₆	39,51024	105,09455	328,71982	1025,6697	337449,5327	1105114810
C₇	39,51024	77,743778	316,08192	837,86986	288840,4172	930027508,4
O₈	9,218188	38,233538	75,10629	305,868304	85899,64937	284074052,7
S₉	70,8441	54,04857	232,98981	647,3843	213952,7799	698815935,8
Total (Sum)	334,282628	898,672996	2927,373368	9162,333648	2975401,357	9709657480
Group (*i*)			Group bilinear indices			
Heteroatoms[b]	97,221188	161,589148	516,01722	1630,900804	517568,6278	1694766789
Atom-type (*i*)			Atom-type bilinear indices			
CH-arom[c]	118,53072	328,71982	1093,15444	3436,227578	1124290,022	3674640223
			*k*th Stochastic MDs, $^{s}b_{kL}(\bar{x}_i, \bar{y}_i)^{[a]}$			
	$k=0$	$k=1$	$k=2$	$k=3$	$k=8$	$k=15$
Atom (*i*)			Atom-level bilinear indices			
N₁	17,1589	23,10234667	22,34781083	23,66748208	23,89914186	24,01863153
C₂	39,51024	33,38525667	40,80307944	39,00990209	38,9415613	38,95862823
C₃	39,51024	60,9142225	44,90684156	47,06658972	45,29450781	45,13795589
C₄	39,51024	42,80276	47,40669467	47,62823907	47,72612389	47,70941996
C₅	39,51024	37,86398	37,79391122	39,18662953	40,21791136	40,32910683
C₆	39,51024	35,03151667	35,08198711	34,64609483	36,44270378	36,7396616
C₇	39,51024	27,639906	37,134922	32,10412486	34,42791207	33,95447149
O₈	9,218188	16,116086	10,803039	14,98191176	13,79473544	14,04362415
S₉	70,8441	29,6352825	42,1269325	34,8188677	34,3736933	34,24319914
Total (Sum)	334,282628	306,491357	318,4052183	313,1098416	315,1182908	315,1346988
Group (*i*)			Group bilinear indices			
Heteroatoms[b]	97,221188	68,85371517	75,27778233	73,46826154	72,0675706	72,30545482
Atom type (*i*)			Atom-type bilinear indices			
CH-arom[c]	118,53072	106,2807533	113,6789778	112,8426264	115,6021765	116,0273967

[a] Calculation development using van der Waals volume (V) and polarizability (P) (see Table 2) as combinations (pairs) of two atom-label chemical properties from our weighting schemes. [b] Sum of local-index values for N, O, and S atoms. [c] Sum of local-index values for C atoms 2, 5, and 6.

jacency 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 atom-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, such as linear discrimination analysis. That is to say, one can find a quantitative relationship between an activity A and the bilinear fingerprints having, for example, the following appearance:

$$A = a_0 b_0(x,y) + a_1 b_1(x,y) \\ + a_2 b_2(x,y) + \dots + a_k b_k(x,y) + c \quad (20)$$

where A is the measured activity, $b_k(x,y)$ are the *k*th nonstochastic total bilinear indices, and the a_k values 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.

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 of the appropriate weights to differentiate the molecular atoms. The weights used in this study are those previously proposed for the calculation of the DRAGON descriptors,[55,72,73] that is, atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), and van der Waals atomic volume (V).[55] The values of these atomic labels are shown in Table 2.[55,72–74]
3. Computation of the total and local (atom and atom-type) bilinear indices of the molecular pseudograph's atom ad-

The following descriptors were calculated in this study:

1. $b_k(x,y)$ and $b_k^H(x,y)$ are the *k*th total bilinear indices not considering and considering H atoms in the molecule, respectively.
2. $b_{kL}(x_E,y_E)$ and $b_{kL}^H(x_E,y_E)$ are the *k*th local (group = heteroatoms: S, N, O) bilinear indices not considering and considering H atoms in the molecule, correspondingly. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.

3. $b_{kL}^H(x_{E-H}, y_{E-H})$ are the *k*th local (group = H atoms bonding to heteroatoms: S, N, O) bilinear indices considering H atoms in the molecule. These local descriptors are putative H-bonding donors (H-bonding capacity), lipophilicity, and so on.

4. The *k*th total [$^sb_k(x,y)$ and $^sb_k^H(x,y)$] and atom-type [$^sb_k(x_E,y_E)$, $^sb_k^H(x_E,y_E)$, and $^sb_k^H(x_{E-H},y_{E-H})$] stochastic bilinear indices were also computed.

3.2. Database selection

The general data set used in this study consists of 658 organic compounds of great structural variation, 246 of which have reported activity against tyrosinase; the remainder are inactive. The database of active compounds was chosen considering a representation of most of the different inhibition modes in the case of the compounds with tyrosinase inhibitory activity. It includes compounds that belong to different subsystems such as derivatives of chalcone,[75,76] new phenolic compounds,[77]

azobenzene derivatives,[78] kojic acid tripeptide library members,[79] glycyrrhetinic acid derivatives,[80] novel N-substituted *N*-nitrosohydroxylamines,[81,82] catechins,[83] gentisic acid esters,[10] hydroxystilbene compounds,[84] benzaldoximes,[85] cinnamic acid derivatives,[86,87] vitamin B_6 compounds,[88] flavonoids,[89] phlorotannins,[90] disubstituted oxadiazole analogues,[14] longifolene derivatives,[91] androstadienone derivatives,[15] steroids,[92] and so on. Figure 1 shows some representative compounds from these data. The names of compounds in the database together with their experimental data taken from the literature are given in Table 4. The molecular structures of these 246 tyrosinase inhibitors are listed in Table 1 of the Supporting Information.

In this sense, it is remarkable that the wide variability of drugs and mechanisms of inhibition of active compounds in the training and prediction sets assures adequate extrapolation power and increases the possibilities of the discovery of new lead compounds with novel mechanisms of tyrosinase inhibition, one of the most critical aspects in the construction of non-congeneric data. In this way, this dataset provides a useful
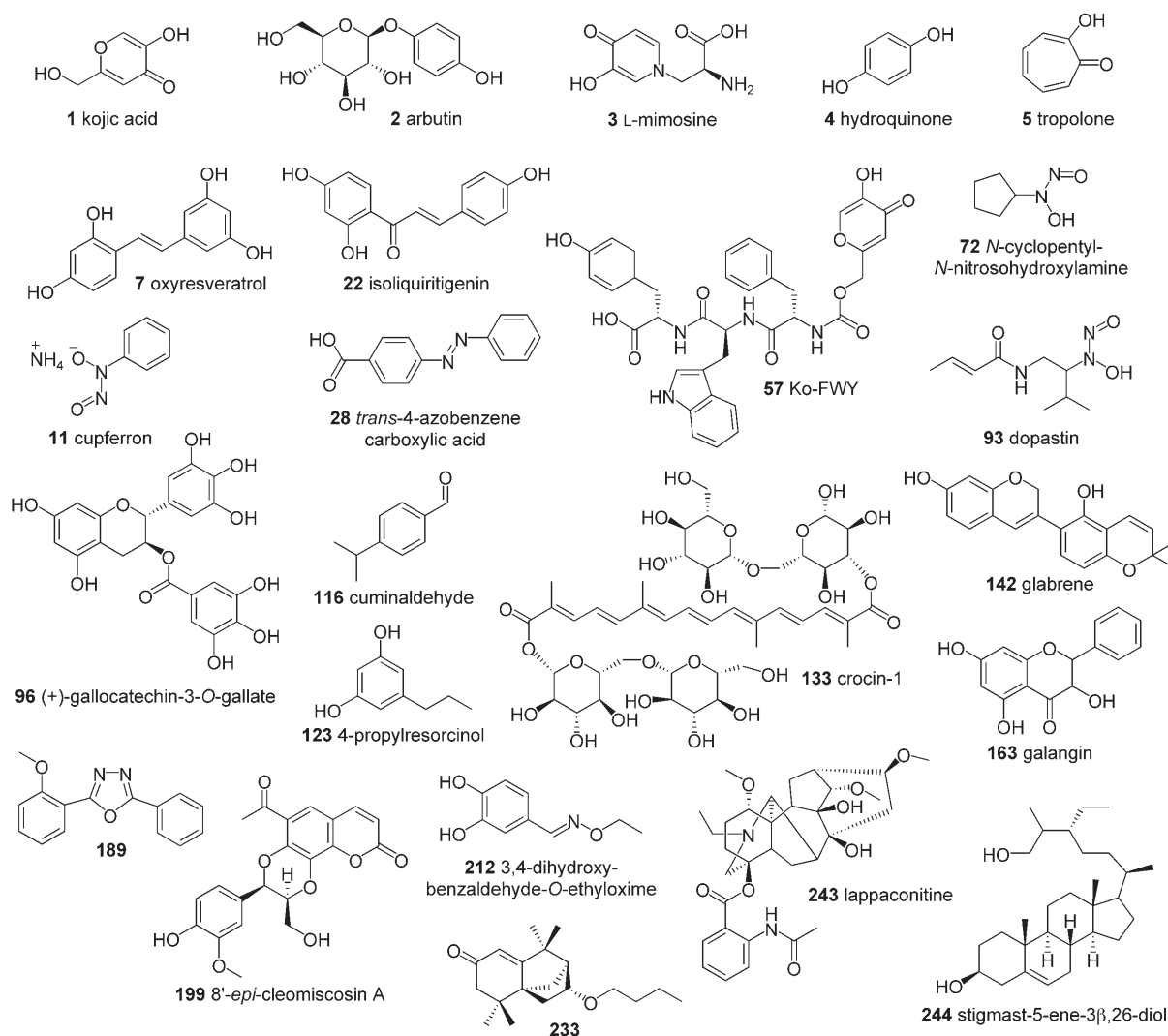
Figure 1. Some examples of the molecular families of tyrosinase inhibitors studied herein.

Table 4. Databases of tyrosinase inhibitors taken from different sources.

Compound[a]	IC$_{50}$ [µM][b]	Inhibition [%][c]	K_i [µM][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
1 Kojic acid		29.71		Mu	L-D		[117]
	22.94	35			L-T		[79]
	60			Mh	L-D		[80]
	500						
	<u>500</u>						
	934.5			Mh		C	[118]
				Mh	L-D		[82]
				Mh			[81]
	4.36			Ma			[10]
	40.1	76.7 (100)		Mh	L-T		[84]
	> 100	43.0 (100)		Mu	L-T		[84]
	9.85			Mh	L-T		[119]
	280			Mh	L-D		[120]
	22			Mh	L-D		[85]
	8.66				L-T		[76]
	22				L-D		[76]
	27				L-D		[121]
	235			Mh	L-D		[122]
		64 (0.14)			L-T		[3]
		72 (0.14)			L-D		[3]
							[123]
		small			L-D		[124]
	250				L-D		[125]
	10			Mh	L-D		[12]
	131			Mu	L-D		[12]
	410			Hu	L-D		[12]
	3.51						[126]
	79.51				L-D		[127]
	4.41			Mh	L-T		[128]
		50 (1)		Th	L-D		[129]
		24 (0.1)		Th	L-D		[129]
	16.67						[14]
2 Arbutin				Mh			[130]
	54.72			Ma			[10]
	5.51			Mh	L-T		[119]
	8400			Mh	L-D		[120]
	24 000				L-D	C	[131]
	790			Mh	L-T		[131]
	14 700			Mh	L-D		[131]
	24 000				L-D		[122]
		13 (90)		Mh	L-T		[87]
		0 (90)		Mh	L-D		[87]
	17 000				L-D		[125]
						C	[125]
	11 130			Mh	L-D		[12]
	205			Mu	L-D		[12]
	3020			Hu	L-D		[12]
	41.13			Mh	L-T		[128]
3 L-Mimosine							[123]
				Mh			[81]
		small			L-D		[124]
							[132]
	3.68						[14]
4 Hydroquinone	25.88			Mh	L-T		[83]
	65.39			Ma			[10]
	4500				L-D		[122]
	4500				L-D		[125]
							[133]
5 L-Tropolone				Mh	L-D		[82]
				Mh			[81]
	0.13			Mh	L-D		[85]
							[123]
		100 (100)			L-D		[124]
		60 (1000)		Pa	L-D		[134]
		15 (200)		Pa	L-D		[134]
	1			Mh			[135]

Table 4. (Continued)							
Compound[a]	IC$_{50}$ [µM][b]	Inhibition [%][c]	K_i [µM][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
				Mh			[136]
6 Ascorbic acid		23 (90)		Mh	L-T		[87]
		34 (90)		Mh	L-D		[87]
	28.38						[126]
		57 (50)					[137]
		39 (100)					[137]
	30			Mh	L-D		[80]
	300						
	<u>500</u>						
7 Oxyresveratrol						N	[117]
	1	95		Mh	L-D		[138]
	1.2	97.3		Mh	L-T		[84]
	52.7	63.3		Mu	L-T		[84]
	1				L-D	N	[125]
8 Quercetin	130		38.6	Mh	L-D	C	[139]
							[123]
	50		29	Mh	L-D	C	[89]
9 Benzoic acid							[140]
							[141]
	710			Mh	L-D	C	[86]
			14	Ne	L-D	C	[142]
							[135]
							[143]
10 Benzaldehyde	830			Mh	L-D	N	[86]
				Mh	L-D	C	[144]
11 Cupferron	1.1			Mh	L-D		[81]
				Mh	L-D	C	[82]
	0.52		0.20	Mh	L-T	C	[145]
	0.84		0.48	Mh	L-D	C	[145]
12 Aloesin						N	[125]
	193			Mh	L-D		[12]
	167			Mu	L-D		[12]
	710			Hu	L-D		[12]
13 *trans*-Resveratrol				Mu			[117]
				Ca			[146]
	155						[3]
	250				L-D	N	[125]
	155	78		Mh			[138]
	54.6	63.8		Mh	L-T		[84]
	100	32.7		Mu	L-T		[84]
14 Anisaldehyde				Mh	L-D	C	[144]
	380			Mh	L-D	N	[86]
					L-D		[144]
15 Cinnamic acid	320			Mh	L-D	N	[147]
				Mh			[130]
	700			Mh	L-D	M	[86]
16 Gnetol		99		Mu	L-D		[117]
17 Dihydrognetol		18		Mu	L-D		[117]
18 3,3′,4-Hydroxy-*trans*-stilbene	87.7	74.3		Mu	L-D		[117]
19 3,3′,4,4′-Hydroxy-*trans*-stilbene	29.1	98.3		Mu	L-D		[117]
20 3-Amino-L-tyrosine	14						[140]
							[148]
21 2-Aminophenol	2						[140]
22 Isoliquiritigenin	8.1	67 (50)					[75]
23 4-Hydroxychalcone	21.8	71 (50)					[75]
24 Butein	29.3	77 (50)					[75]
25 4′-Hydroxychalcone		11 (50)					[75]
26 2′,4′-Dihydroxychalcone		14 (50)					[75]
27 2′,4-Dihydroxychalcone		10 (50)					[75]
28 *trans*-4-Azobenzene carboxylic acid		67 (1000)		Mh			[78]
29 *cis*-4-Azobenzene carboxylic acid		86 (1000)		Mh			[78]
30 *trans*-4,4′-Azobenzene dicarboxylic acid		72 (1000)		Mh			[78]
31 *cis*-4,4′-Azobenzene dicarboxylic acid		53 (1000)		Mh			[78]
32 Castanospermine	2.64	50					[149]
33 Deosynojirimycin		50					[149]
34 Ko-YGC		52.5		Mh	L-T		[79]

Compound[a]	IC$_{50}$ [μm][b]	Inhibition [%][c]	K_i [μm][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
Table 4. (Continued)							
35 Ko-YGV		54		Mh	L-T		[79]
36 Ko-YGE		55		Mh	L-T		[79]
37 Ko-YGT		52		Mh	L-T		[79]
38 Ko-YGL		55		Mh	L-T		[79]
39 Ko-YGW		58		Mh	L-T		[79]
40 Ko-YGF		57		Mh	L-T		[79]
41 Ko-YGH		57		Mh	L-T		[79]
42 Ko-YGN		52		Mh	L-T		[79]
43 Ko-YGD		54		Mh	L-T		[79]
44 Ko-YGG		57		Mh	L-T		[79]
45 Ko-YIG		50		Mh	L-T		[79]
46 Ko-YYG		60		Mh	L-T		[79]
47 Ko-YSG		55		Mh	L-T		[79]
48 Ko-YMG		56		Mh	L-T		[79]
49 Ko-YQG		54		Mh	L-T		[79]
50 Ko-YRG		61		Mh	L-T		[79]
51 Ko-YHG		52		Mh	L-T		[79]
52 Ko-YNG		53		Mh	L-T		[79]
53 Ko-YDG		54		Mh	L-T		[79]
54 Ko-FIY		88		Mh	L-T		[79]
55 Ko-FRY		91		Mh	L-T		[79]
56 Ko-FYY	0.39	93		Mh	L-T		[79]
57 Ko-FWY	0.24	95		Mh	L-T		[79]
58 Ko-FFY	0.33	94		Mh	L-T		[79]
59 Ko-KWY		67		Mh	L-T		[79]
60 Ko-KRY		59		Mh	L-T		[79]
61 Ko-KKY		66		Mh	L-T		[79]
62 Ko-KIY		65		Mh	L-T		[79]
63 Ko-FWW	6.17			Mh	L-T		[79]
64 Ko-FWF	4.48			Mh	L-T		[79]
65 Ko-FWI	2.18			Mh	L-T		[79]
66 Ko-FWD	2.13			Mh	L-T		[79]
67 Ko-WWY	0.78			Mh	L-T		[79]
68	**500**			Mh	L-D		[80]
	100						
	_70_						
69	**500**			Mh	L-D		[80]
	80						
	_35_						
70	**500**			Mh	L-D		[80]
	150						
	_100_						
71 Glabridin				Mh			[80]
	0.09				L-T		[76]
	0.09				L-T		[3]
	3.94				L-D		[3]
72 *N*-Cyclopentyl-*N*-nitrosohydroxylamine	0.6			Mh	L-D	N	[82]
	0.6			Mh	L-D		[81]
73 *N*-Benzyl-*N*-nitrosohydroxylamine	3.0			Mh	L-D		[82]
	3.0			Mh	L-D		[81]
74	13.7			Mh	L-D	N	[82]
75	19.2			Mh	L-D	M	[82]
76	13.1			Mh	L-D	M	[82]
77	9.7			Mh	L-D	M	[82]
78	15.8			Mh	L-D	M	[82]
79	20.3			Mh	L-D	N	[82]
80	11.6			Mh	L-D	N	[82]
81	23.8			Mh	L-D	N	[82]
82	1.3			Mh	L-D		[81]
83	23.5			Mh	L-D		[81]
84	273			Mh	L-D		[81]
85	139			Mh	L-D		[81]
86	1.2			Mh	L-D		[81]
87	1.1			Mh	L-D		[81]
88	1.5			Mh	L-D		[81]
89	2.2			Mh	L-D		[81]

Table 4. (Continued)

Compound[a]	IC$_{50}$ [μM][b]	Inhibition [%][c]	K_i [μM][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
90	19.3			Mh	L-D		[81]
91	280			Mh	L-D		[81]
92	220			Mh	L-D		[81]
93 Dopastin	20			Mh	L-D		[81]
94	46			Mh	L-D		[81]
95 (−)-Epigallocatechin-3-*O*-gallate	34.10	58		Mh	L-T		[83]
96 (+)-Gallocatechin-3-*O*-gallate	17.34	72		Mh	L-T		[83]
97 (−)-Epicatechin-3-*O*-gallate	34.58	63		Mh	L-T		[83]
98 (−)-Epigallocatechin		40		Mh	L-T		[83]
99 Magnesium L-ascorbyl-2-phosphate	71.84			Ma			[10]
100	6.66			Ma			[10]
101	10.75			Ma			[10]
102	14.53			Ma			[10]
103	29.30			Ma			[10]
104	60.88			Ma			[10]
105	89.19			Ma			[10]
106	50.46			Ma			[10]
107 3,5-Dihydroxy-4′-methoxystilbene	252	57		Mh			[138]
	78.4	57.9		Mh	L-T		[84]
	100	3.7		Mu	L-T		[84]
108 3,4′-Dimethoxy-5-hydroxystilbene	490	50		Mh			[138]
	100	0		Mh	L-T		[84]
	100	2.3		Mu	L-T		[84]
109 Piceid	500	23		Mh			
	100	10.2		Mh	L-T		[84]
	100	24.0		Mu	L-T		[84]
110 Rhapontigenin	76.2	58.7		Mh	L-T		[84]
	100	2.3		Mu	L-T		[84]
111 Rhaponticin	100	0		Mh	L-T		[84]
	100	7.7		Mu	L-T		[84]
112 Kurarinone	1.04			Mh	L-T	N	[119]
113 Kushnol F	2.12			Mh	L-T		[119]
114 4-Hydroxyanisol	12.08			Mh	L-T		[119]
							[131]
							[150]
115 2-Hydroxy-4-methoxybenzaldehyde	30			Mh	L-D		[151]
116 Cuminaldehyde	50			Mh	L-D		[151]
			120	Mh	L-T	C	[144]
			9	Mh	L-D	C	[144]
	33.74			Mh	L-D	N	[144]
117	820			Mh	L-D		[151]
118	320			Mh	L-D		[151]
119	330			Mh	L-D		[151]
120 Artocarbene	2.45			Mh	L-T		[76]
121 Chlorophorin	2.6			Mh	L-T		[76]
122 Norartocarpanone	1.76			Mh	L-T		[76]
123 4-Propylresorcinol	0.91			Mh	L-T		[76]
124 3,4-Dihydroxybenzonitrile	45			Mh	L-D		[76]
	45			Mh	L-D		[76]
125 3,4,2,4-*trans*-stilbene	1.5			Mh	L-T		[76]
126	29.3			Mh	L-T		[76]
	> 100			Mh	L-D		[76]
127	0.2			Mh	L-T		[76]
	7.5			Mh	L-D		[76]
128	31.68			Mh	L-T		[76]
	> 100			Mh	L-D		[76]
129	0.02			Mh	L-T		[76]
	90			Mh	L-D		[76]
130 Artogomezianol	68			Mh	L-D		[121]
131 Andalasin	38			Mh	L-D		[121]
132 Crocusatins H	870			Mh	L-D		[122]
133 Crocin-1	140			Mh	L-D		[122]
134 Crocin-3	960			Mh	L-D		[122]
135 3,4-Dihydroxycinnamic acid	970			Mh	L-D	N	[86]
136 4-Hydroxy-3-methoxycinnamic acid	330			Mh	L-D	N	[86]
137 Anisic acid	680			Mh	L-D	N	[86]

Table 4. (Continued)

Compound[a]	IC$_{50}$ [μM][b]	Inhibition [%][c]	K$_i$ [μM][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
	600		603		L-D	N	[152]
138 2-Methoxycinnamic acid	340			Mh	L-D	N	[86]
139 3-Methoxycinnamic acid	350			Mh	L-D	N	[86]
140 4-Methoxycinnamic acid	340			Mh	L-D	N	[86]
141 Kaempferol	230			Mh.	L-D		[153]
142 Glabrene	8.1				L-T		[3]
	7600				L-D		[3]
143 Pyridoxine		30	5200	Mh	L-D	M	[88]
144 Pyridoxamine		38	4300	Mh	L-D	M	[88]
145 Pyridoxal		30		Mh	L-D		[88]
146 Pyridoxamine-5′-phosphate		30		Mh	L-D		[88]
147 Protocatechuic acid methyl ester	280	75.4 (0.5)		Mh	L-T		[154]
148 Protocatechuic acid	420	60 (0.5)		Mh	L-T		[154]
149 *m*-Coumaric acid							[155]
150 3-Caffeoylquinic acid		27 (90)		Mh	L-T		[87]
		21 (90)		Mh	L-D		[87]
151 4-Caffeoylquinic acid		25 (90)		Mh	L-T		[87]
		22 (90)		Mh	L-D		[87]
152 5-Caffeoylquinic acid		26 (90)		Mh	L-T		[87]
		20 (90)		Mh	L-D		[87]
153 5-Feroloylquinic acid		40 (90)		Mh	L-T		[87]
		23 (90)		Mh	L-D		[87]
154 3,4-Dicaffeoylquinic acid		51 (90)		Mh	L-T		[87]
		51 (90)		Mh	L-D		[87]
155 3,5-Dicaffeoylquinic acid		48 (90)		Mh	L-T		[87]
		59 (90)		Mh	L-D		[87]
156 4,5-Dicaffeoylquinic acid		45 (90)		Mh	L-T		[87]
		56 (90)		Mh	L-D		[87]
157 Caffeic acid		28 (90)		Mh	L-T		[87]
		25 (90)		Mh	L-D		[87]
158 Fisetin	130		154	Mh	L-D	C	[89]
159 3,7,4′-Trihydroxyflavone	270		410	Mh	L-D	C	[89]
160 Morin	720		103	Mh	L-D	C	[89]
161 Luteolin	240			Mh	L-D	U	[89]
162 Apigenin	150	20 (150)		Mh	L-D		[89]
163 Galangin	101		58	Mh	L-D	C	[89]
164			50		L-T		[156]
			97		L-D		[156]
165 Diethyldithiocarbamate		100			L-D		[124]
		100 (1)					[129]
		76 (0.1)					[129]
166 Phenylthiourea		100			L-D		[124]
		57 (50)					[137]
167 2-Mercaptobenzothiazole	3			Mh	T		[135]
168	5.87			Mh	M	N	[157]
169	1.31			Mh	M	N	[157]
170 Phloroglucinol	73.58		2.3×10^{-4}	Mh	L-T	C	[128]
171 Eckstolonol	34.02		3.1×10^{-4}	Mh	L-T	C	[128]
172 Eckol	8.91		1.9×10^{-5}	Mh	L-T	N	[128]
173 Phlorofucofuroeckol	29.47		1.4×10^{-3}	Mh	L-T	N	[128]
174 Dieckol	0.29		1.5×10^{-5}	Mh	L-T	N	[128]
175 HPABS			7.17		L-T	C	[158]
176 Gluthatione		86 (1)		Th	L-D		[129]
		38 (0.1)		Th	L-D		[129]
177 β-Mercaptoethanol		100 (1)		Th	L-D		[129]
		100 (0.1)		Th	L-D		[129]
178 Protocatechualdehyde	65.74		1.1	Mh	L-T	C	[159]
179	5.15			Mh	L-D		[14]
180	3.18			Mh	L-D		[14]
181	5.23			Mh	L-D		[14]
182	6.04			Mh	L-D		[14]
183	2.18			Mh	L-D		[14]
184	3.29			Mh	L-D		[14]
185	4.05			Mh	L-D		[14]
186	3.98			Mh	L-D		[14]
187	10.40			Mh	L-D		[14]

Table 4. (Continued)

Compound[a]	IC$_{50}$ [μM][b]	Inhibition [%][c]	K_i [μM][d]	Source[e]	Substrate[f]	I.T.[g]	Ref.
188	3.23			Mh	L-D		[14]
189	8.71			Mh	L-D		[14]
190	5.16			Mh	L-D		[14]
191	7.18			Mh	L-D		[14]
192	7.82			Mh	L-D		[14]
193	7.28			Mh	L-D		[14]
194	6.21			Mh	L-D		[14]
195	6.43			Mh	L-D		[14]
196	7.81			Mh	L-D		[14]
197	1.36			Mh	L-D		[160]
198	11.68			Mh	L-D		[160]
199 8′-*epi*-Cleomiscosin A	1.33			Mh	L-D		[16]
200	256.97			Mh	L-D		[16]
201	18.69			Mh	L-D		[16]
202	15.69			Mh	L-D		[16]
203	8.65			Mh	L-D		[16]
204	64			Mh	L-D		[85]
205 3,4-Dihydroxybenzaldoxime	18			Mh	L-D		[85]
206 3-Hydroxy-4-methoxybenzaldoxime	4.6			Mh	L-D		[85]
207 3,4,5-Trihydroxybenzaldoxime	20.2			Mh	L-D		[85]
208 4-Hydroxy-3-methoxybenzaldoxime	2.3			Mh	L-D		[85]
209 3-Ethoxy-4-hydroxybenzaldoxime	3.5			Mh	L-D		[85]
210 4-Hydroxybenzaldoxime	25			Mh	L-D		[85]
211 4-Methoxybenzaldoxime	56			Mh	L-D		[85]
212 3,4-Dihydroxybenzaldehyde-*O*-ethyloxime	0.3			Mh	L-D		[85]
213 3,4-Dihydroxybenzaldehyde-*O*-(4-methylbenzyl)-oxime	3			Mh	L-D		[85]
214 3-Hydroxy-4-methoxybenzaldehyde-*O*-ethyloxime	18			Mh	L-D		[85]
215 3,4,5-Trihydroxybenzaldehyde-*O*-ethyloxime	380			Mh	L-D		[85]
216 4-Hydroxy-3-methoxybenzaldehyde-*O*-ethyloxime	4.2			Mh	L-D		[85]
217 3-Ethoxy-4-hydroxybenzaldehyde-*O*-ethyloxime	14			Mh	L-D		[85]
218 4-Hydroxybenzaldehyde-*O*-ethyloxime	43			Mh	L-D		[85]
219 4-Hydroxy-3-methylbenzaldehyde-*O*-ethyloxime	124			Mh	L-D		[85]
220 3,5-Dimethyl-4-hydroxybenzaldehyde-*O*-ethyloxime	500			Mh	L-D		[85]
221	620			Mh	L-D		[85]
222	520			Mh	L-D		[91]
223	51			Mh	L-D		[91]
224	101			Mh	L-D		[91]
225	9.64			Mh	L-D		[91]
226	6.68			Mh	L-D		[91]
227	14.05			Mh	L-D		[91]
228	22.03			Mh	L-D		[91]
229	7.45			Mh	L-D		[91]
230	68.79			Mh	L-D		[91]
231	5.79			Mh	L-D		[91]
232	9.77			Mh	L-D		[91]
233	2.70			Mh	L-D		[91]
234 (+)-Androst-4-ene-3,17-dione	3.68			Mh	L-D		[15]
235 Androsta-1,4-diene-3,17-dione	6.42			Mh	L-D		[15]
236 17β-Hydroxyandrosta-1,4-dien-3-one	139.67			Mh	L-D		[15]
237 11α-Hydroxyandrost-4-ene-3,17-dione	3.56			Mh	L-D		[15]
238 17β,11α-Dihydroxyandrost-4-en-3-one	1.25			Mh	L-D		[15]
239	2.61			Mh	L-D		[77]
240	1.85			Mh	L-D		[77]
241	0.40			Mh	L-D		[77]
242 Esculetin	43			Mh	L-D	C	[161]
243 Lappaconitine	93.33			Mh	L-D		[162]
244 Stigmast-5-ene-3β,26-diol	2.39			Mh	L-D		[92]
245 Stigmast-5-ene-3β-ol	5.25			Mh	L-D		[92]
246 Campesterol	8.90			Mh	L-D		[92]

[a] Capital letters refer to the amino acid code. The molecular structures of these 246 tyrosinase inhibitors is given in Table 1 of the Supporting Information. [b] Minimum inhibitory concentration data gathered at: **boldface**, 10 min; *italics*, 60 min; underline, 120 min. [c] Percent enzymatic inhibition (values in parentheses: concentration of substrate used [μM]). [d] Inhibition constant values. [e] Tyrosinase sources: Mh = mushroom, Mu = murine, Ma = mammalian, Ca = carignan, Ne = *Neurospora* sp., Pa = pear, Hu = human, Th = *Thermomicrobium roseum*. [f] Substrates used in assays: L-D = L-DOPA, L-T = L-tyrosine, M = 4-[(4-methylbenzo)azo]-1,2-benzenediol (MeBACat), TBC = 4-*tert*-butylcatechol, T = tyramine. [g] I.T. = inhibition type: C, competitive; N, noncompetitive; U, uncompetitive; M, mixed.

tool for scientific research in synthesis, natural product chemistry, theoretical chemistry, and other areas related to the field of tyrosinase inhibitors.

On the other hand, 412 compounds having different clinical uses such as antivirals, sedative/hypnotics, diuretics, anticonvulsants, hemostatics, oral hypoglycemics, anti-hypertensives, antihelminthics, and anticancer compounds were selected for the set of inactive compounds through random selection, guaranteeing great structural variability as well. All these compounds were taken from the Negwer Handbook,[93] in which their names, synonyms, and structural formulas can be found.

The classification of these organic compounds as 'inactive' (non-inhibitors of tyrosinase) does not guarantee that all are truly so; some of them may have inhibitory activity toward tyrosinase that is undetected. This problem can be reflected in the results of classification for the series of inactive compounds.[94]

Therefore, the developed LDA-based QSAR models can classify some of these compounds as inhibitors against tyrosinase, helping with the identification of new compounds, among drugs from large datasets with other pharmacological uses. In our case, we carried out a virtual screening experiment in another part of this work. To split the database into training and prediction series, two *k*-means cluster analyses (*k*-MCA) were carried out for both active and inactive compounds to design, in a rational representative way, the training (learning) and prediction (test) series.[95, 96]

3.3. Chemometric methods

3.3.1. Cluster analysis

Cluster analysis (CA) is the name of a group of methods used to recognize similarities among cases (objects) or among variables and single out some categories as a set of similar cases (or variables).[18] This CA comprehends a number of different 'classification algorithms' and it allows organization of the data into subsystems. These algorithms are grouped into two categories: hierarchical clustering and partitional (nonhierarchical) clustering. Hierarchical clustering rearranges objects in a tree structure (joining clustering) in an agglomerative (bottom-up) procedure. On the other hand, partitional clustering assumes that the objects have nonhierarchical characters.[18, 95–97]

The most frequently used cluster algorithms are the *k*-mean cluster algorithm (*k*-MCA) and the Jarvis–Patrick algorithm (also known as *k*-nearest-neighbor cluster algorithm; *k*-NNCA); in our case, to design the training and test series and to demonstrate the great structural variability of the present database, we carried out both kinds of cluster analyses (*k*-MCA and *k*-NNCA) for both the active and inactive series of compounds.[18, 96–98] The number of members in each cluster and the standard deviation (SD) of the variables in the cluster (kept as low as possible) were taken into account to have an acceptable statistical quality of data partition in clusters. The values of the SD between and within clusters of the respective Fisher ratio and their *p*-level of significance, were also examined.[18, 95–97] Finally, before carrying out the cluster processes,

all variables were standardized. In standardization, all values of selected variables (MDs) were replaced by standardized values, which are computed as follows: standard score = (raw score − mean)/SD.

3.3.2. Linear discriminant analysis

The aim of linear discriminant analysis (LDA), a heuristics algorithm capable of distinguishing among two or more categories of objects, is to find a linear function allowing one to discriminate between active and inactive compounds.[98] LDA was selected between many statistical methods to get classification functions owing to its simplicity.[99] LDA is one of the most currently used nowadays; moreover, its use in drug discovery and design has been widely described.[17, 22, 24, 28, 31–33, 35, 36, 38, 42, 43, 45–48, 94, 100–102] The LDA was carried out with STATISTICA software.[97] The considered tolerance parameter was the default value for minimal acceptable tolerance, which is 0.01. A forward stepwise search procedure was fixed as the strategy for variable selection; the selection method by forward stepwise exhibited the majority of applications in drug design. The construction process of the model occurs through many steps in the following way: the variables enter and are evaluated by STATISTICA in the model, the variable with greatest contribution to discriminate between groups is included in the model, and then STATISTICA continues with the next step.

The principle of parsimony (Occam's razor) was taken into account as a strategy for model selection. In this case simplicity is loosely equated with the number of parameters in the model. In this context, we selected the model with a high statistical significance but having as few parameters (a_k) as possible. The quality of the models was determined by examining Wilks' λ parameter (*U* statistic), the value of which in the overall discrimination can range from 0 (perfect discrimination) to 1 (no discrimination) and the square Mahalanobis distance (D^2), which indicates the separation of the respective groups, showing whether the model possesses an appropriate discriminatory power for differentiating between the two corresponding groups. The Fisher ratio (*F*), the corresponding *p* level [*p*(*F*)], and the percentage of good classification in the training and test sets were also examined.

The classification of cases was carried out by means of the posterior classification probabilities. Using the Mahalanobis distance values to do the classification, one can now derive probabilities. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from that group centroid, and then it can be said that posterior probability is that, based on our knowledge of the values of other variables, with which the respective case belongs to a particular group.[97]

The biological activity was coded 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*(Active)−*P*(Inactive)]100, or

otherwise as inactive. *P*(Active) and *P*(Inactive) are the probabilities with which the equations classify a compound as active or inactive, respectively.

The statistical robustness and predictive power of the obtained model were assessed using a prediction (test) set. Finally, the calculation of percentages of global good classification (accuracy), sensibility, specificity (also known as "hit rate"), false positive rate (also known as "false alarm rate"), and Matthews correlation coefficient (*C*) in the training and test sets, together with the linear discriminant canonical statistics: canonical regression coefficient (R_{can}) and chi-square (χ^2) permitted the assessment of the model.[103]

3.3.3. Orthogonalization of descriptors

The interrelation among the MDs makes the interpretation of the QSAR model difficult. To overcome this, we examined the orthogonalization of the molecular descriptors introduced by Randić and others several years ago, as a way to improve the statistical interpretation of the models by using interrelated indices.[102–108] It is well known that the interrelation among the different descriptors can result in highly unstable regression coefficients, which makes it impossible to know the relative importance of an index included in a model.

However, in some cases, strongly interrelated descriptors can enhance the quality of a model because the small fraction of a descriptor that is not reproduced by its strongly interrelated pair can provide positive contributions to modeling.[102–108] The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of their colinearity with other variables included in the model. In addition, the coefficients of the QSAR model based on orthogonal descriptors are stable to the inclusion of novel descriptors, which permits interpretation of the correlation coefficient and to evaluate the role of individual fingerprints in the QSAR model.

In this study the Randić method of orthogonalization was used.[102–108] As a first step, an appropriate order of orthogonalization was considered following the order with which the variables were selected in the forward stepwise search procedure of the statistical analysis.[108] The first variable (V_1) is taken as the first orthogonal descriptor $^1O(V_1)$, and the second one (V_2) is orthogonalized with respect to it [$^2O(V_2)$]. The residual of its correlation with $^1O(V_1)$ is that part of the descriptor V_2 not reproduced by $^1O(V_1)$. Similarly, from the regression of V_3 versus $^1O(V_1)$, the residual is the part of V_3 that is not reproduced by $^1O(V_1)$, and it is labeled $^1O(V_3)$. The orthogonal descriptor $^3O(V_3)$ is obtained by repeating this process to make it also orthogonal to $^2O(V_2)$. The process is repeated until all variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model.

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

3.4. Chemical methods

The isolation and characterization of seven cycloartane compounds, their biological studies, and cross-references has been reported by other members of our research team.[109]

3.5. In vitro determination of tyrosinase inhibitory activity

The tyrosinase inhibition assay was performed with kojic acid and L-mimosine (both from Sigma Chemical, USA) as standard inhibitors of tyrosinase in a 96-well microplate format using a SpectraMax 340 microplate reader (Molecular Devices, CA, USA) according to the method developed by Hearing.[110] 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$_{50}$ studies. Compounds were dissolved in methanol to a concentration of 2.5%. Thirty units of mushroom tyrosinase (28 nM, Sigma Chemical, USA) were first preincubated with the test compounds in 50 nM sodium 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 for 10 min by measuring the change in absorbance at 475 nm (at 37 °C) due to the formation of DOPAchrome. The percent enzyme inhibition was calculated as follows, by using a program based in Excel 2000 (Microsoft, USA) developed for this purpose:

$$\% \text{ inhibition} = [A_B - A_S/A_B]\,100 \qquad (21)$$

Here the A_B and A_S are the absorbance values for the blank and samples, respectively. After screening of the compounds, median inhibitory concentrations (IC$_{50}$) were also calculated. All studies were carried out in triplicate at least, and the results represent the mean ± SEM (standard error of the mean).

4. Results and Discussion

4.1. Similarity analysis and design of training and test sets

The quality of a classification (or any QSAR) model depends on the quality of the selected data set. One of the most critical aspects of constructing the training set is to warrant enough molecular diversity for it. Taking this into account, we selected a data set of 658 compounds having a great degree of structural variability. The 246 tyrosinase inhibitors considered in this study are representative of families with different inhibition modes and diverse structural patterns.

To demonstrate the structural diversity of these data sets, we performed a hierarchical CA of the active and inactive series.[95,96] Both dendrograms are given in Figures 2 and 3, using the Euclidean distance (*x* axis) and the complete linkage (*y* axis), illustrating the results of the *k*-NNCA developed in active and inactive sets, correspondingly. As can be observed in both binary trees, there is a great number of different subsets, which prove the molecular variability of the selected compounds in these databases.
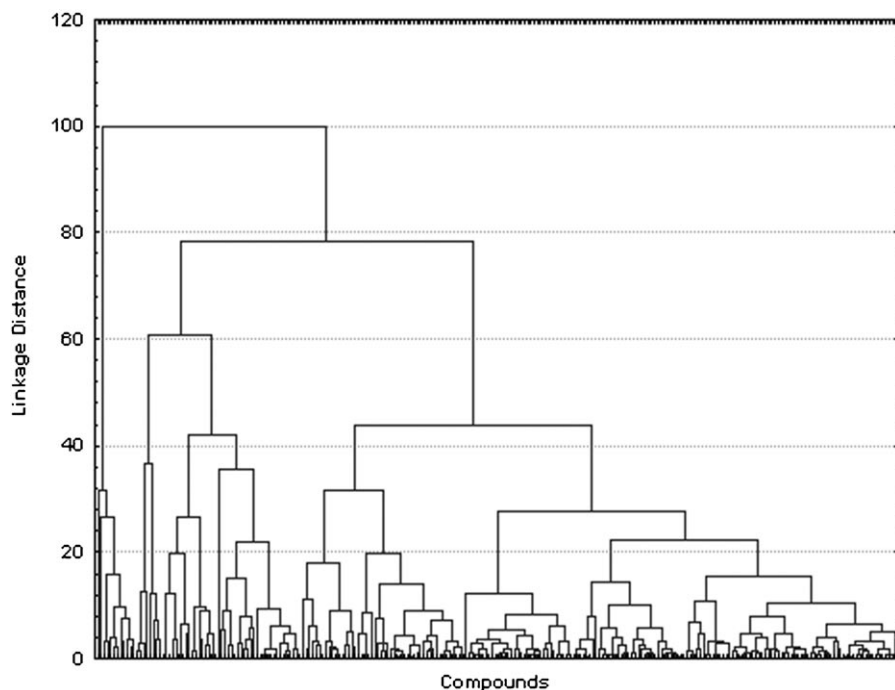
Figure 2. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of tyrosinase inhibitors used in the training and prediction sets.
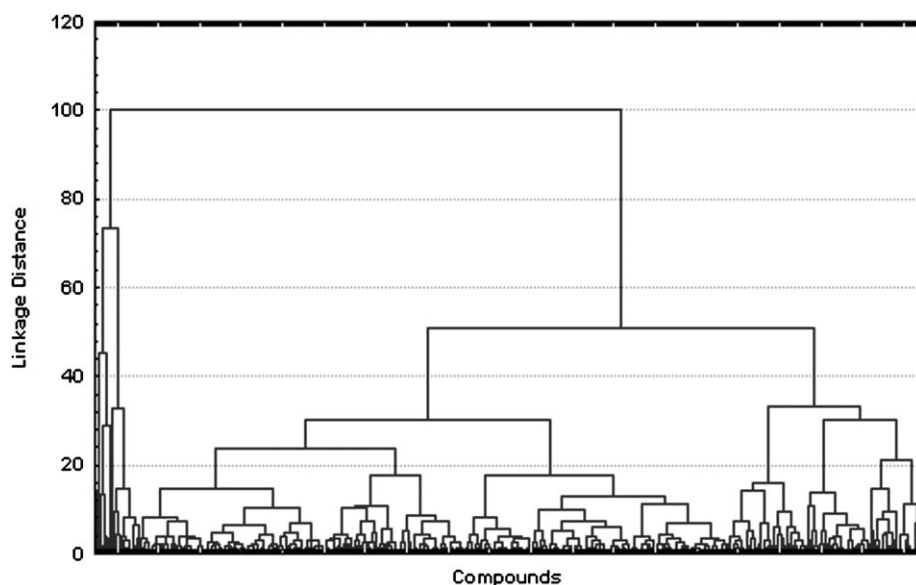

Figure 3. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of inactive compounds (non-tyrosinase inhibitors) used in the training and prediction sets.

cal class (as determined by the clusters derived from *k*-MCA) will be represented in both compound series. This "rational" design of the training and predicting series allowed us to design both sets that are representative of the whole "experimental universe".

We first carried out a *k*-MCA with active compounds and afterward, we did likewise with inactive compounds. A first *k*-MCA (*k*-MCA I) split the tyrosinase inhibitors into 10 clusters. Additionally, the inactive compound series was also partitioned into 12 clusters (*k*-MCA II). The *k*th total and atom-type nonstochastic bilinear indices were used, with all variables showing $p < 0.05$ for the Fisher test. The results are listed in Table 5.

Afterward, the selection of the training and prediction sets was performed by randomly taking compounds belonging to each cluster. From these 658 compounds, 478 were chosen at random to form the training set, 183 of which were actives, and 295 were inactives. The great structural variability of the selected training data set makes possible the discovery of lead compounds, not only with determined mechanisms of tyrosinase inhibitory activity, but also with novel modes of inhibition.

The remaining subseries, composed of 63 tyrosinase inhibitors and 117 compounds with different biological properties, were prepared as test sets for the external set validation of the models. These compounds were never used

Because of the difficulty in evaluating the output dendrogram, another kind of CA is usually performed. In this sense and also to split the whole group into two data sets (training and predicting), we performed a *k*-MCA. The main idea of this procedure consists of making a partition of either active or inactive series of compounds in several statistically representative compound classes. This procedure ensures that any chemi-

in the development of the classification models. Figure 4 is a graphic illustration of the procedure described above, where two independent cluster analyses (one for active and the other for inactive compounds) were performed to select a representative sample for the training and test sets.

Table 5. Analysis of variance results of the *k*-MCAs for tyrosinase inhibitors and inactive druglike compounds.

Variables	Between SS[a]	Within SS[b]	Fisher ratio [F]	p level[c]
	Tyrosinase inhibitors clusters (*k*-MCA I)			
$^{VK}b_0^H(x)$	74.98	6.22	316.24	0.00
$^{VK}b_0(x)$	319.25	19.50	429.20	0.00
$^{VK}b_{4L}^H(x_E)$	199.73	16.66	314.44	0.00
$^{VK}b_{5L}^H(x_E)$	159.77	16.23	258.09	0.00
	Inactives clusters (*k*-MCA II)			
$^{VK}b_0^H(x)$	382.18	53.55	318.77	0.00
$^{VK}b_0(x)$	375.92	41.96	400.18	0.00
$^{VK}b_{4L}^H(x_E)$	645.11	40.11	718.49	0.00
$^{VK}b_{5L}^H(x_E)$	687.14	43.72	702.06	0.00

[a] Variability between groups. [b] Variability within groups. [c] Level of significance.
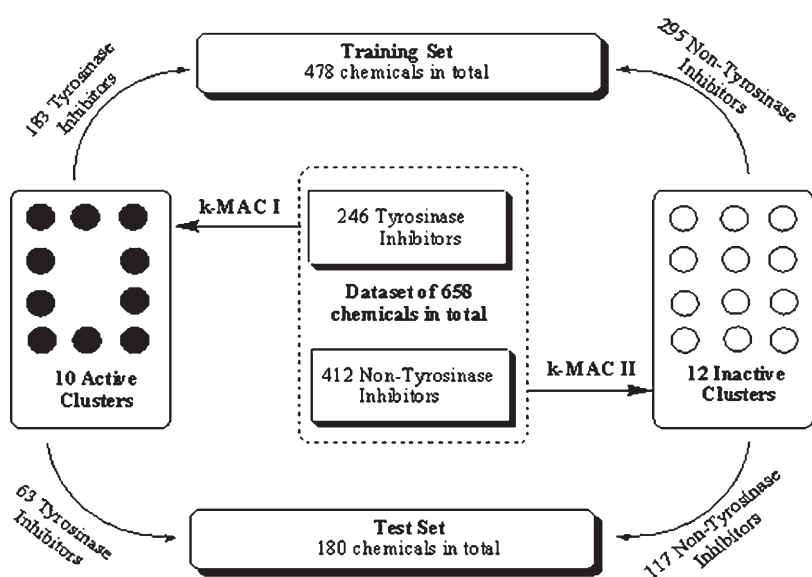
Figure 4. General algorithm used for designing the training and test sets throughout *k*-MCA.

4.2. Discriminant models

4.2.1. Development

Once we perform a representative selection of training series, they can be used in many different chemometric techniques to fit discriminant functions such as SIMCA or neural networks. In our case, however, we select the LDA given the simplicity of the method, to derive discriminant functions that permit the classification of compounds as either active (tyrosinase inhibitors) or inactive. LDA is a method extensively reported in drug design.[31–33,45–48,94,100–102]

In the present study, we developed classification functions by using the total and group nonstochastic and stochastic bilinear indices as independent variables. These MDs were computed with the TOMOCOMD–CARDD software according to

the weighting schemes proposed above in section 3.1 (see Table 2).

Thus, 12 LDA-based QSAR models were obtained. The first six models used the nonstochastic total and local bilinear indices [Eqs. (22)–(27)] and the remaining six, the stochastic molecular descriptors [Eqs. (28)–(33)]. The classification models obtained are given in Table 6, and in Table 7 the prediction performances, Wilks' statistics (λ), the square of the Mahalanobis distances (D^2), and the Fisher ratio (F) for LDA-based QSAR models with the training set are shown. All these statistical parameters measured the quality of the models determined. The equations were shown to be statistically significant ($p < 0.0001$).

As can be observed in Table 7, the fitted models for Equations (26) and (32) showed the best results; in these models, the combination of van der Waals volume (V) and Mulliken electronegativity (K) weighting schemes were used, including the nonstochastic and stochastic bilinear indices. The best models [Eq. (26)] and [Eq. (32)] correctly classified 99.58 and 89.96 % accuracy of the training set. The equations showed high Matthews correlation coefficients (C) of 0.99 and 0.79, respectively. C quantifies the strength of the linear relation between the molecular descriptors and the classifications, and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages (accuracy).[111]

The best model with the nonstochastic bilinear fingerprints [Eq. (26)] shows a percentage of false actives in this data set of 0 %, that is, no inactive compounds were classified as actives out of 295 cases. In addition, two compounds from the group of 183 actives were misclassified as inactive (1.09 % misclassification).

In the statistical analysis of model [Eq. (32)] we can show adequate results. Only 9.83 % misclassification for the inactive group was observed (29 inactive compounds were classified as active out of a total of 295). In this case, 19 compounds out of 117 (10.38 %) were false inactives. These commented results are most of the parameters commonly used in medical statistics (sensitivity, specificity, and false positive rate (also known as 'false alarm rate')) for the whole set of developed models. While the sensitivity is the probability of correctly predicting a positive example, the specificity (also known as 'hit rate') is the probability that a positive prediction is correct.[111]

These statistical parameters mentioned above, together with the linear discriminant canonical statistics: canonical regression coefficient (R_{can}) as well as chi-squared (χ^2) and its p level [$p(\chi^2)$] were checked, and the results are given in Table 7. The canonical transformations of the LDA results with the nonstochastic [Eq. (26)] and stochastic [Eq. (32)] bilinear indices give rise to canonical roots with good canonical correlation coefficients of 0.87 and 0.72. The χ^2 test permits us to asses the statistical significance of this analysis as having a p level of < 0.0001.

Table 6. Discriminant models obtained with total and local nonstochastic and stochastic bilinear indices.

LDA-based QSAR models obtained with nonstochastic bilinear indices	Equation

$Class = -1.054 + 2.701 \times 10^{-\ MV}b_1(x,y) - 3.133 \times 10^{-10\ MV}b_{12}(x,y) - 3.875 \times 10^{-4\ MV}b_{2L}{}^H(x_E,y_E) - 8.983 \times 10^{-4\ MV}b_{3L}{}^H(x_E,y_E) + 9.924 \times 10^{-4\ MV}b_{3L}(x_E,y_E) - 0.927\ {}^{MV}b_{0L}{}^H(x_{E-H},y_{E-H}) + 0.137\ {}^{MV}b_{1L}{}^H(x_{E-H},y_{E-H}) - 2.756 \times 10^{-3\ MV}b_{3L}{}^H(x_{E-H},y_{E-H}) + 6.643 \times 10^{-5\ MV}b_{6L}{}^H(x_{E-H},y_{E-H})$ ... (22)

$Class = -0.216 - 2.624 \times 10^{-3\ MP}b_2(x,y) + 1.607 \times 10^{-3\ MP}b_3(x,y) - 5.192 \times 10^{-7\ MP}b_9(x,y) - 1.060 \times 10^{-2\ MP}b_{3L}{}^H(x_E,y_E) + 9.951 \times 10^{-3\ MP}b_{3L}(x_E,y_E) + 1.845 \times 10^{-10\ MP}b_{15L}(x_E,y_E) - 7.128\ {}^{MP}b_{0L}{}^H(x_{E-H},y_{E-H}) + 1.164\ {}^{MP}b_{1L}{}^H(x_{E-H},y_{E-H}) - 7.339 \times 10^{-3\ MP}b_{5L}{}^H(x_{E-H},y_{E-H}) + 1.841 \times 10^{-3\ MP}b_{6L}{}^H(x_{E-H},y_{E-H})$ ... (23)

$Class = 4.376 \times 10^{-2} - 7.176 \times 10^{-3\ MK}b_0{}^H(x,y) - 3.660 \times 10^{-4\ MK}b_3{}^H(x,y) + 1.126 \times 10^{-2\ MK}b_1(x,y) - 1.450 \times 10^{-8\ MK}b_{11}(x,y) - 5.481 \times 10^{-3\ MK}b_{3L}{}^H(x_E,y_E) - 2.613 \times 10^{-4\ MK}b_{4L}{}^H(x_E,y_E) + 1.389 \times 10^{-10\ MK}b_{15L}(x_E,y_E) + 5.839 \times 10^{-3\ MK}b_{3L}(x_E,y_E) - 2.642\ {}^{MK}b_{0L}{}^H(x_{E-H},y_{E-H}) + 0.390\ {}^{MK}b_{1L}{}^H(x_{E-H},y_{E-H})$ ... (24)

$Class = 4.376 \times 10^{-2} - 1.817 \times 10^{-7\ VP}b_9{}^H(x,y) - 1.466 \times 10^{-2\ VP}b_0(x,y) + 4.114 \times 10^{-3\ VP}b_2(x,y) - 1.068 \times 10^{-6\ VP}b_9(x,y) + 2.357 \times 10^{-8\ VP}b_{10}(x,y) - 0.229\ {}^{VP}b_{1L}{}^H(x_E,y_E) - 9.258 \times 10^{-3\ VP}b_{3L}{}^H(x_E,y_E) + 0.236\ {}^{VP}b_{1L}(x_E,y_E) + 7.755 \times 10^{-3\ VP}b_{3L}(x_E,y_E) + 2.482 \times 10^{-10\ VP}b_{15L}(x_E,y_E) + 0.584\ {}^{VP}b_{0L}{}^H(x_{E-H},y_{E-H})$ ... (25)

$Class = -1.359 - 1.583 \times 10^{-2\ VK}b_0{}^H(x,y) + 1.808 \times 10^{-2\ VK}b_0(x,y) - 3.356 \times 10^{-4\ VK}b_{4L}{}^H(x_E,y_E) + 6.883 \times 10^{-5\ VK}b_{5L}{}^H(x_E,y_E)$ ... (26)

$Class = 0.213 - 1.353 \times 10^{-2\ PK}b_2{}^H(x,y) + 9.503 \times 10^{-2\ PK}b_1(x,y) - 6.577 \times 10^{-6\ PK}b_8(x,y) - 4.318 \times 10^{-2\ PK}b_{3L}{}^H(x_E,y_E) - 0.142\ {}^{PK}b_{0L}(x_E,y_E) + 4.071 \times 10^{-2\ PK}b_{3L}(x_E,y_E) + 5.457 \times 10^{-9\ PK}b_{14L}(x_E,y_E) + 1.838\ {}^{PK}b_{1L}{}^H(x_{E-H},y_{E-H}) - 0.737\ {}^{PK}b_{2L}{}^H(x_{E-H},y_{E-H}) + 8.283 \times 10^{-4\ PK}b_{8L}{}^H(x_{E-H},y_{E-H}) - 1.611 \times 10^{-5\ PK}b_{11L}{}^H(x_{E-H},y_{E-H})$ ... (27)

LDA-Based QSAR Models Obtained Using Stochastic Bilinear Indices	Equation

$Class = 0.244 + 2.072 \times 10^{-2\ MV}b_{5L}{}^H(x_E,y_E) - 1.655 \times 10^{-2\ MV}b_{13L}{}^H(x_E,y_E) + 7.797 \times 10^{-4\ MV}b_{0L}(x_E,y_E) - 7.756 \times 10^{-3\ MV}b_{1L}(x_E,y_E) + 3.167 \times 10^{-3\ MV}b_{15L}(x_E,y_E) + 9.155 \times 10^{-2\ MV}b_{1L}{}^H(x_{E-H},y_{E-H}) + 1.299\ {}^{MV}b_{4L}{}^H(x_{E-H},y_{E-H}) - 4.021\ {}^{MV}b_{6L}{}^H(x_{E-H},y_{E-H}) + 0.284\ {}^{MV}b_{7L}{}^H(x_{E-H},y_{E-H}) + 2.890\ {}^{MV}b_{8L}{}^H(x_{E-H},y_{E-H}) - 0.529\ {}^{MV}b_{11L}{}^H(x_{E-H},y_{E-H})$ ... (28)

$Class = 0.367 + 0.116\ {}^{MP}b_{5L}{}^H(x_E,y_E) - 6.444 \times 10^{-2\ MP}b_{13L}{}^H(x_E,y_E) - 4.929 \times 10^{-2\ MP}b_{1L}(x_E,y_E) - 1.572 \times 10^{-2\ MP}b_{2L}(x_E,y_E) - 4.250\ {}^{MP}b_{0L}{}^H(x_{E-H},y_{E-H}) + 1.127\ {}^{MP}b_{1L}{}^H(x_{E-H},y_{E-H}) + 0.740\ {}^{MP}b_{2L}{}^H(x_{E-H},y_{E-H}) + 1.273\ {}^{MP}b_{3L}{}^H(x_{E-H},y_{E-H}) - 2.742\ {}^{MP}b_{6L}{}^H(x_{E-H},y_{E-H}) - 2.402\ {}^{MP}b_{11L}{}^H(x_{E-H},y_{E-H}) + 3.582\ {}^{MP}b_{12L}{}^H(x_{E-H},y_{E-H})$ ... (29)

$Class = 7.356 \times 10^{-2} - 4.445 \times 10^{-2\ MK}b_{2L}{}^H(x_E,y_E) + 0.176\ {}^{MK}b_{4L}{}^H(x_E,y_E) - 0.149\ {}^{MK}b_{12L}(x_E) + 3.926 \times 10^{-2\ MK}b_{13L}(x_E,y_E) - 6.588 \times 10^{-2\ MK}b_{4L}(x_E,y_E) - 2.591 \times 10^{-2\ MK}b_{5L}(x_E,y_E) + 6.262 \times 10^{-2\ MK}b_{10L}(x_E,y_E) - 0.663\ {}^{MK}b_{0L}{}^H(x_{E-H},y_{E-H}) + 0.305\ {}^{MK}b_{1L}{}^H(x_{E-H},y_{E-H}) - 0.483\ {}^{MK}b_{9L}{}^H(x_{E-H}) + 0.386\ {}^{MK}b_{15L}{}^H(x_{E-H},y_{E-H})$ ... (30)

$Class = 3.706 \times 10^{-2} + 2.343 \times 10^{-2\ VP}b_0{}^H(x,y) - 8.684 \times 10^{-2\ VP}b_1{}^H(x,y) + 8.618 \times 10^{-2\ VP}b_3{}^H(x,y) - 1.971 \times 10^{-2\ VP}b_{14}{}^H(x,y) + 5.051 \times 10^{-2\ VP}b_{1L}{}^H(x_E,y_E) - 9.229 \times 10^{-2\ VP}b_{5L}{}^H(x_E,y_E) + 0.112\ {}^{VP}b_{6L}{}^H(x_E,y_E) - 3.527 \times 10^{-2\ VP}b_{0L}(x_E,y_E) - 4.570 \times 10^{-2\ VP}b_{1L}(x_E,y_E) + 0.757\ {}^{VP}b_{0L}{}^H(x_{E-H},y_{E-H}) - 1.582\ {}^{VP}b_{3L}{}^H(x_{E-H},y_{E-H}) + 1.006\ {}^{VP}b_{5L}{}^H(x_{E-H},y_{E-H})$ ... (31)

$Class = -0.390 + 0.130\ {}^{VK}b_8(x,y) - 0.132\ {}^{VK}b_{15}(x) + 6.646 \times 10^{-2\ VK}b_3(x,y) - 6.330 \times 10^{-2\ VK}b_6(x,y) + 4.988 \times 10^{-2\ VK}b_{1L}{}^H(x_E,y_E) - 0.125\ {}^{VK}b_{12L}{}^H(x_E,y_E) + 8.020 \times 10^{-2\ VK}b_{15L}{}^H(x_E,y_E) - 1.929 \times 10^{-2\ VK}b_{0L}(x_E,y_E) - 3.926 \times 10^{-2\ VK}b_{3L}(x_E,y_E) + 5.045 \times 10^{-2\ VK}b_{10L}(x_E,y_E) - 0.613\ {}^{VK}b_{0L}{}^H(x_{E-H},y_{E-H}) + 0.594\ {}^{VK}b_{1L}{}^H(x_{E-H},y_{E-H})$ ... (32)

$Class = 0.377 + 1.033 \times 10^{-2\ PK}b_1(x,y) - 0.159\ {}^{PK}b_{0L}{}^H(x_E,y_E) + 0.596\ {}^{PK}b_{1L}{}^H(x_E,y_E) - 0.555\ {}^{PK}b_{14L}(x_E,y_E) - 0.205\ {}^{PK}b_{1L}(x_E,y_E) + 0.294\ {}^{PK}b_{10L}(x_E,y_E) + 2.051\ {}^{PK}b_{1L}{}^H(x_{E-H},y_{E-H}) - 4.711\ {}^{PK}b_{2L}{}^H(x_{E-H},y_{E-H}) + 13.854\ {}^{PK}b_{3L}{}^H(x_{E-H},y_{E-H}) - 14.693\ {}^{PK}b_{5L}{}^H(x_{E-H},y_{E-H}) + 3.363\ {}^{PK}b_{14L}{}^H(x_{E-H},y_{E-H})$ ... (33)

4.2.2. Orthogonalization and external validation

A close inspection of the molecular descriptors included in the two best LDA-based QSAR models showed that several of these molecular fingerprints are strongly interrelated to each other (data not shown). To overcome this difficulty, we used the Randić orthogonalization process of the molecular descriptors. This process is an approach in which molecular descriptors are transformed in such a way that they do not mutually correlate (see subsection 3.3.3). Both the nonorthogonal (original) descriptors and the derived orthogonal descriptors contain the same information. Therefore, the

Table 7. Prediction performance and statistical parameters for LDA-based QSAR models in the training set.

Models[a]	C[b]	Accuracy "Q_{total}" [%]	Specificity [%]	Sensitivity "hit rate" [%]	False positive rate [%]	Wilks' λ	D^2	F	(χ^2)	R_{can}[c]
LDA-based QSAR models obtained with nonstochastic bilinear indices										
Eq. (22) (9)	0.83	91.84	88.7	90.2	7.1	0.47	4.82	59.5	359.6	0.73
Eq. (23) (10)	0.82	91.63	88.2	90.2	7.5	0.45	5.18	57.4	377.7	0.74
Eq. (24) (11)	0.87	93.93	93.3	90.7	4.1	0.43	5.50	61.0	393.2	0.75
Eq. (25) (11)	0.81	91.21	87.7	89.6	7.8	0.47	4.77	47.9	377.7	0.74
Eq. (26) (4)	**0.99**	**99.58**	**100**	**98.9**	**0**	**0.24**	**13.06**	**366.3**	**668.6**	**0.87**
Eq. (27) (11)	0.82	91.63	88.2	90.2	7.5	0.44	5.32	53.5	383.3	0.75
LDA-based QSAR models obtained with stochastic bilinear indices										
Eq. (28) (11)	0.72	86.82	81.6	84.7	11.9	0.47	4.66	46.8	350.4	0.72
Eq. (29) (11)	0.77	88.91	85.7	85.2	8.8	0.49	4.45	44.7	339.0	0.72
Eq. (30) (11)	0.75	88.02	85.4	83.1	8.8	0.47	4.75	47.8	355.3	0.73
Eq. (31) (12)	0.72	86.82	80.6	86.3	12.9	0.54	3.53	32.4	285.8	0.68
Eq. (32) (12)	**0.79**	**89.96**	**85.0**	**89.6**	**9.8**	**0.48**	**4.57**	**42.1**	**345.5**	**0.72**
Eq. (33) (11)	0.76	88.70	83.1	88.5	11.2	0.47	4.72	47.5	353.7	0.73

[a] Values in parentheses indicate the quantity of variables of the models. [b] Matthews correlation coefficient. [c] Canonical correlation coefficient obtained from the linear discriminant canonical analysis.

same statistical parameters of the QSAR models are obtained.[102–108] In Equations (34) and (35) we show the results of the orthogonalization of the nonstochastic and stochastic bilinear indices included in both models, respectively:

$$Class = -2.251 + 9.300 {}^{1}O({}^{VK}b_0(x,y)) - 3.599 {}^{2}O({}^{VK}b_0{}^{H}(x,y))$$
$$-1.329 {}^{3}O({}^{VK}b_{4L}{}^{H}(x_E,y_E)) + 3.472 {}^{4}O({}^{VK}b_{5L}{}^{H}(x_E,y_E))$$
$$N = 478, \lambda = 0.24, D^2 = 13.06, F = 366.3, R = 0.87,$$
$$\chi 2 = 668.6, Q = 99.5, C = 0.99 \tag{34}$$

$$Class = -0.245 - 29.379 {}^{1}O({}^{VK}b_{0L}{}^{H}(x_{E-H},y_{E-H}))$$
$$+0.991 {}^{2}O({}^{VK}b_{1L}{}^{H}(x_{E-H},y_{E-H})) - 1.257 {}^{3}O({}^{VK}b_{15L}{}^{H}(x_E,y_E))$$
$$-3.018 {}^{4}O({}^{VK}b_{0L}(x_E,y_E)) + 4.566 {}^{5}O({}^{VK}b_{10L}(x_E,y_E))$$
$$-10.390 {}^{6}O({}^{VK}b_{12L}{}^{H}(x_E,y_E)) + 2.176 {}^{7}O({}^{VK}b_{1L}{}^{H}(x_E,y_E))$$
$$+0.500 {}^{8}O({}^{VK}b_3(x)) + 75.488 {}^{9}O({}^{VK}b_8{}^{H}(x,y)) \tag{35}$$
$$-36.108 {}^{10}O({}^{VK}b_6(x,y)) - 2.136 {}^{11}O({}^{VK}b_3(x_E,y_E))$$
$$-3.446 {}^{12}O({}^{VK}b_{15}{}^{H}(x,y)$$
$$N = 478, \lambda = 0.48, D^2 = 4.57, F = 42.1, R_{can} = 0.72,$$
$$\chi 2 = 345.5, Q = 89.9, C = 0.79$$

Here, we use the symbols $^{m}O(b_k(x,y))$, where the superscript m expresses the order of importance of the variable $(b_k(x,y))$ after a preliminary forward stepwise analysis and O means orthogonal.

It must be highlighted here that the orthogonal descriptor-based models coincide with the collinear (ordinary) TOMO-COMD–CARDD-descriptors-based models in all the statistical parameters. The statistical coefficients of LDA QSARs λ, F, D^2, C, and accuracy are the same whether we use either a set of non-orthogonal descriptors or the corresponding set of orthogonal indices. This is not surprising, because the latter models are derived as a linear combination of the former and cannot have more information content than them.[102–108]

The classification of all compounds in the complete training dataset provides some assessment of the goodness of the fit of the models, but it does not provide a thorough criterion of how the models can predict the biological properties for new compounds. To assess such predictive power, the use of an external test set is essential.[112, 113]

In this sense, the activity of the compounds in the test set was predicted with the obtained discrimination functions. The overall accuracy for the two best models of the nonstochastic and stochastic bilinear fingerprints [Eqs. (26) and (32)] was 100% ($C = 1$) and 87.78% ($C = 0.73$), respectively, in the prediction series. The accuracy and other statistical parameters (sensitivity, specificity, and false positive rate) of the test set are depicted in Table 8. Furthermore, in Figures 5 and 6 a plot of %ΔP (see section 3.3.2) can be observed, from the classification obtained by models [Eq. (26)] and [Eq. (32)], for each compound in the training and test sets. The results obtained with the approach used herein are adequate; they validate these models for use in ligand-based virtual screening.

The complete set of organic compounds in the training data set as well as their classification (including canonical scores)

Table 8. Prediction performance for LDA-based QSAR models in the test set.

Models	$C^{[a]}$	Accuracy "Q_{total}" [%]	Specificity [%]	Sensitivity "hit rate" [%]	False positive rate [%]
LDA-based QSAR models obtained with nonstochastic bilinear indices					
Eq. (22)	0.78	90.00	83.6	88.9	9.4
Eq. (23)	0.80	91.00	83.8	90.5	9.4
Eq. (24)	0.78	90.00	84.6	87.3	8.5
Eq. (25)	0.83	91.67	83.3	95.2	10.3
Eq. (26)	**1**	**100**	**100**	**100**	**0**
Eq. (27)	0.79	90.00	80.8	93.7	12.0
LDA-based QSAR models obtained with stochastic bilinear indices					
Eq. (28)	0.68	84.44	73.3	87.3	17.1
Eq. (29)	0.74	87.78	79.7	87.3	12.0
Eq. (30)	0.71	86.67	79.1	84.1	12.0
Eq. (31)	0.68	85.00	76.5	82.5	13.7
Eq. (32)	**0.73**	**87.78**	**82.5**	**82.5**	**9.4**
Eq. (33)	0.76	88.89	84.1	84.1	8.5
[a] Matthews correlation coefficient.					

using all the models are shown in Tables 2–5 of the Supporting Information. In addition, the results of classification using all developed equations for active and inactive compounds in the test set are shown in Tables 6–9 of the Supporting Information.

4.3. Database virtual screening reveals new classes of tyrosinase inhibitors

The massive cost of developing new drugs and the small economic size of the market for most of the drugs make this development slow. For these reasons, pharmaceutical industries require, now more than ever, a new approach that is able to answer the challenge of discovering new lead drugs at minimal cost.[114] In this way, the pharmaceutical industry has based its paradigm in high-throughput screening (HTS) as a technology for cutting time and costs, even though no guarantees for success are provided.[115] Despite great advances in the biological screening of large numbers of compounds by HTS, the process of drug discovery is still an arduous task. In this sense, virtual screening (based on QSAR techniques) has emerged as an interesting alternative to HTS and as an important drug-discovery tool.[18, 19]

One of the most important features of any QSAR model is its ability to predict the desired activity for new compounds from compound databases. In this context the development of computational methods that permit in silico assays of tyrosinase inhibitor activity for virtual compound libraries before these compounds are synthesized in the laboratory can be considered as an alternative to HTS in the drug-discovery process.

With the aim of testing the feasibility of our models and to show the potential of the current approach to detect new lead compounds with "unknown" structures, we carried out a simulated ligand-based virtual screen of tyrosinase inhibitors. Table 9 lists the names of the 75 compounds that were chosen for this along with their reported activity or inactivity and cor-
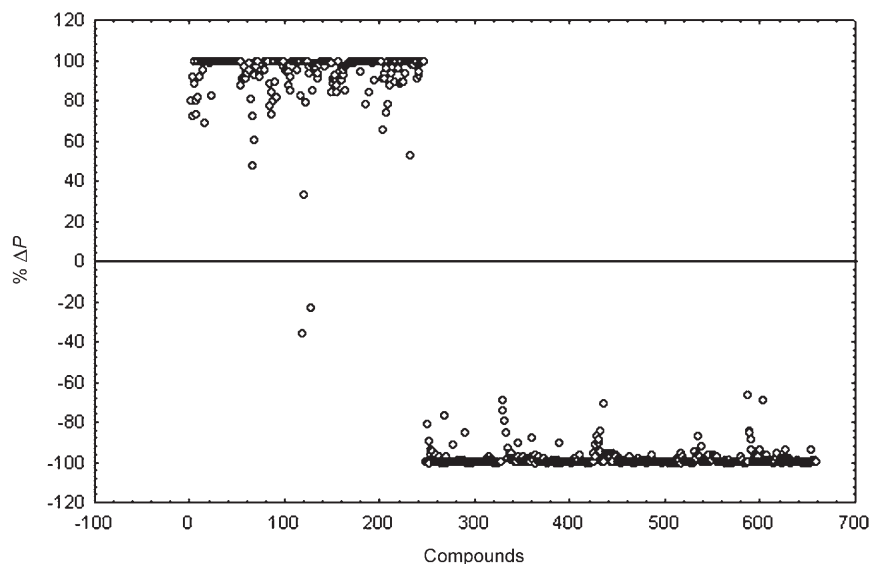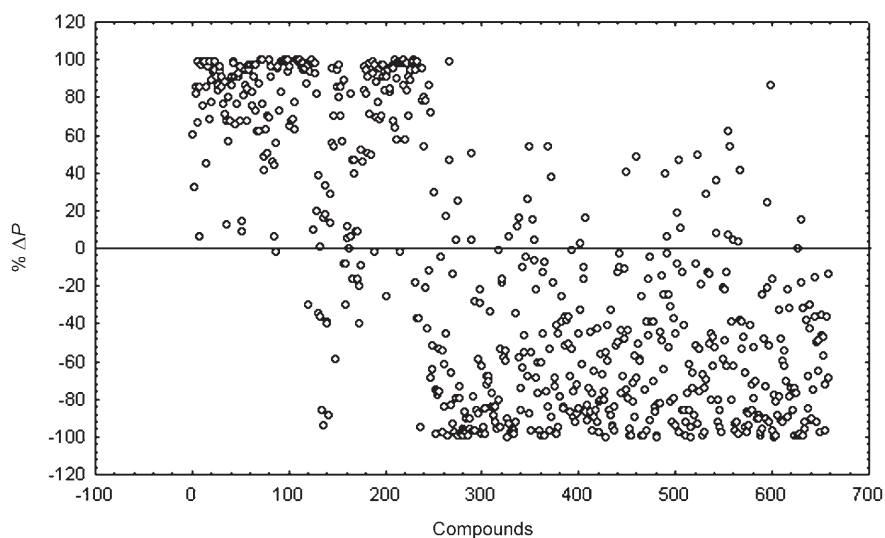
Figure 5. Plot of the predicted %ΔP from [Eq. (26)] (using nonstochastic bilinear indices) for each compound in the training and test sets. Compounds **1**–**183** and **184**–**246** are active (tyrosinase inhibitors) in the training and test sets, respectively; compounds **247**–**541** and **542**–**658** are inactive (non-tyrosinase inhibitors) in both training and test sets, correspondingly.

Figure 6. Plot of the %ΔP from [Eq. (32)] (using stochastic bilinear indices) for each compound in the training and test sets. Compounds **1**–**183** and **184**–**246** are active (tyrosinase inhibitors) in the training and test sets, respectively; compounds **247**–**541** and **542**–**658** are inactive (non-tyrosinase inhibitors) in both training and test sets, correspondingly.

are shown in Table 11 of the Supporting Information. A plot of %ΔP values of the good classification for the models [Eq. (26)] and [Eq. (32)] are given in Figures 8 and 9. These figures are pictorial representations of the accuracy of the best two LDA-based QSAR models, which classified most of the compounds included in this "simulated" virtual screening experiment well. For instance, 90.66 and 82.66% of the screened compounds were well classified by Equations (26) and (32), respectively. Verification of the predictions carried out by all the models obtained comes from recently published reports from which these compounds were selected (see 'Ref.' column of Table 9).

After making this type of analysis, the strategy is to include in the training the new compounds discovered through the virtual screening set, and to carry out new models to find novel discrimination functions. However, a new model could have some variability with respect to the previous one, due to the inclusion of new kinds of organic compounds, but this can increase the available spectrum of compounds and permit the recognition of other different structural patterns manifested in a better selection of compounds from the database as tyrosinase inhibitors. Therefore, a constant improvement of the dataset is one of the major impacts and is considered as an iterative process, because it con-

responding literature reference. The structures of these compounds are shown in Table 10 of the Supporting Information.

In the first instance, a *k*-NNCA was carried out to show the molecular diversity in the data set of the virtual screen. Many different subsystems can be observed in the dendrogram of Figure 7, proving the great molecular variability of the compounds.

The results for the classification of the compounds in the virtual screen (external set) are given in Table 9. In addition, the values of %ΔP (posterior classification probabilities) and canonical scores of the compounds with all the models obtained

tributes toward getting better classification models in which a great quantity of compounds with novel structural features are evaluated against the activity of the enzyme.

Several drugs were identified by the models as possible tyrosinase inhibitors. There is great variability in the functions of these compounds. Such drugs included, for example, one anti-rheumatic compound (penicillamine; also used as a copper-chelating agent in the treatment of Wilson's hepatolenticular degeneration), one anti-hyperthyroid compound (methimazole), one anti-hypertensive agent (captopril), one mydriatic lead (yohimbine; also used as pharmacological probe for α_{2A}-

Table 9. Results of ligand-based virtual screening.

Compound[a]	Class[b]	Ref.	Compound[a]	Class[b]	Ref.
			Active Compounds (Tyrosinase Inhibitors)		
1 *p*-Nitrophenol	1) + + + + +	[163]	**27** Dithiothreitol	1) + + + − + +	[150]
	2) + + + + +	[132]		2) + + + − + +	
2 3-(3,4-Dihydroxyphenyl)-l-alanine	1) + + + + +	[140]	**28** Azelaic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + + +	
3 3-Amino-4-hydroxybenzoic acid	1) + + + + +	[140]	**29** Undecandioic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + − + +	
4 4-Amino-3-hydroxybenzoic acid	1) + + + + +	[140]	**30** Suberic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + + +	
5 3,4-Diaminobenzoic acid	1) + + + + +	[140]	**31** Sebacic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + + +	
6 3-Aminobenzoic acid	1) + + + + +	[140]	**32** Dodecandioic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + − + +	
7 4-Aminobenzoic acid	1) + + + + +	[140]	**33** Tridecandioic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + − + +	
8 4,6-*O*-Hexahydroxydiphenylglucose	1) + + + + +	[164]	**34** Traumatic acid	1) + + + + +	[167]
	2) + + + + +			2) + + + − + +	
9 Tunicamycin	1) + + + + +	[149]	**35** Pantothenic acid	1) + + + + +	[148]
	2) + + + + +			2) + + + + +	
10 Methyl *p*-coumarate	1) + + + + +	[130]	**36** 5-(Hydroxymethyl)-2-furfural	1) + + + + +	[168]
	2) + + + + +			2) + + + + +	[159]
11 *o*-Phenylphenol	1) + + + + +	[130]	**37** Hinokitiol	1) + + + + +	[169]
	2) + + + + +			2) + + + + +	
12 Phenylhydroquinone	1) + + + + +	[130]	**38** Penicillamine	1) + + + − + +	[170]
	2) + + + + +			2) + + + − + +	
13 Chamaecin	1) + + + + +	[130]	**39** Toluic acid	1) + + + + +	[163]
	2) + + + + +	[151]		2) + + + + +	
14 Stearyl glycyrrhetinate	1) + + + + +	[80]	**40**	1) + + + + +	[171]
	2) + + + − + +			2) + + + + +	
15 2-(4-Methylphenyl)-1,3-selenazol-4-one	1) + + + + +	[118]	**41**	1) + + + + +	[171]
	2) − − + + − +	[165]		2) + + + + +	
16	1) + + + + +	[118]	**42** 3,5-Dihydroxy-4′-*O*-methoxystilbene	1) + + + + +	[172]
	2) − − − + − −			2) + + + + +	
17	1) + + + + +	[118]	**43** *p*-Hydroxybenzoic acid	1) + + + + +	[125]
	2) − − − + − −			2) + + + + +	
18	1) + + + + +	[118]	**44** *o*-Hydroxybenzoic acid	1) + + + + +	[125]
	2) − − + + − −			2) + + + + +	
19 3-Fluorotyrosine	1) + + + + +	[148]	**45** Cysteine	1) + + + − + +	[173]
	2) + + + + +			2) + + + + +	
20 *N*-Acetyltyrosine	1) + + + + +	[148]	**46** Methimazole	1) + + + − − +	[173]
	2) + + + + +			2) + + + + +	
21 *N*-Formyltyrosine	1) + + + + +	[148]	**47** BMY-28438	1) + + + + +	[173]
	2) + + + + +			2) + + + + +	
22 Gentisic acid	1) + + + + +	[10]	**48** Captopril	1) + + + − + +	[174]
	2) + + + + +			2) + + + − + +	
23 6-BH₄	1) + + − − + +	[166]	**49** Yohimbine	1) + + + − + +	[175]
	2) + + + − −			2) + + + − + +	
24 7-BH₄	1) + + − − + +	[166]	**50** 4-(Phenylazo)phenol	1) + + + + +	[158]
	2) + + + + +			2) + + + + +	
25 Propylparaben	1) + + + + +	[131]	**51** SACat[c]	1) + + + − + +	[158]
	2) + + + + +			2) − + + + + +	
26 Phenylalanine	1) + + + + +	[148]	**52** NPACat[d]	1) + + + + +	[158]
	2) + + + + +			2) + + + + +	
53 DNPACat[e]	1) + + + + +	[158]	**55** Dodecyl gallate	1) + + + + +	[176]
	2) + + + + +			2) + + + + +	
54 EDTA[f]	1) − − + + +	[129]	**56** Gallic acid	1) + + + + +	[176]
	2) − + + + +			2) + + + + +	
			Inactive Compounds		
57	1) − − + − + −	[81]	**67** 5-Methyluracil	1) − − − + + +	[122]
	2) − − + − + +			2) − − − − −	
58	1) − − + − + −	[81]	**68** Uracil	1) − − − − + +	[122]
	2) − − + − + +			2) − − − − −	
59	1) − − − − + −	[81]	**69** Thiourea	1) − − − − − −	[146]
	2) + − − + + +			2) − − − − −	

Table 9. (Continued)

Compound[a]	Class[b]	Ref.	Compound[a]	Class[b]	Ref.
60 Caffeine	1) − − − − + − 2) − − − − − −	[83]	**70** Veratric acid methyl ester	1) − − − − + − 2) − − − − − −	[154]
61 Trimethylresveratrol	1) − − − − + − 2) − − − − − −	[138] [84]	**71** 6-Nitroquipazine	1) − − − − + − 2) − − − − − −	[177]
62 4-Aminoazobenzene-4′-sulfonic acid	1) + + + − + + 2) + − + − + +	[78]	**72** 4-Methoxybenzaldehyde-*O*-ethyloxime	1) − − − − + − 2) − − − − − −	[85]
63 2-Methoxy-4-isopropyl benzaldehyde	1) − − − − + − 2) − − − + − −	[151]	**73**	1) + + + + + + 2) + + + + + +	[85]
64 Petroselinic acid	1) + + + − + + 2) − + + − + +	[167]	**74**	1) − + − + + + 2) + + + + + +	[85]
65 Crocusatins F	1) − − − + + − 2) − − − + − −	[122]	**75** Phytyl-1-hexanoate	1) + + + + + + 2) − − + − + +	[92]
66 2-Formyl-5-methoxyfuran	1) − − − − + − 2) − − − − − −	[122]			

[a] Structures of these tyrosinase inhibitors are given in Table 10 of the Supporting Information. [b] Results for the classification of compounds in this set: 1) classification of each compound using the obtained models with nonstochastic bilinear indices in the following order: Eqs. (22), (23), (24), (25), (26), and (27); 2) classification of each compound using the obtained models with stochastic bilinear indices in the following order: Eqs. (28), (29), (30), (31), (32), and (33). [c] SACat = 4-[(4-sulfonamido)azo]-1,2-benzenediol. [d] NPACat = 4-[(4-nitrophenyl)azo]-1,2-benzenediol. [e] DNPACat = 4-[(2,4-dinitrophenyl)azo]-1,2-benzenediol. [f] ethylenediaminetetraacetic acid.
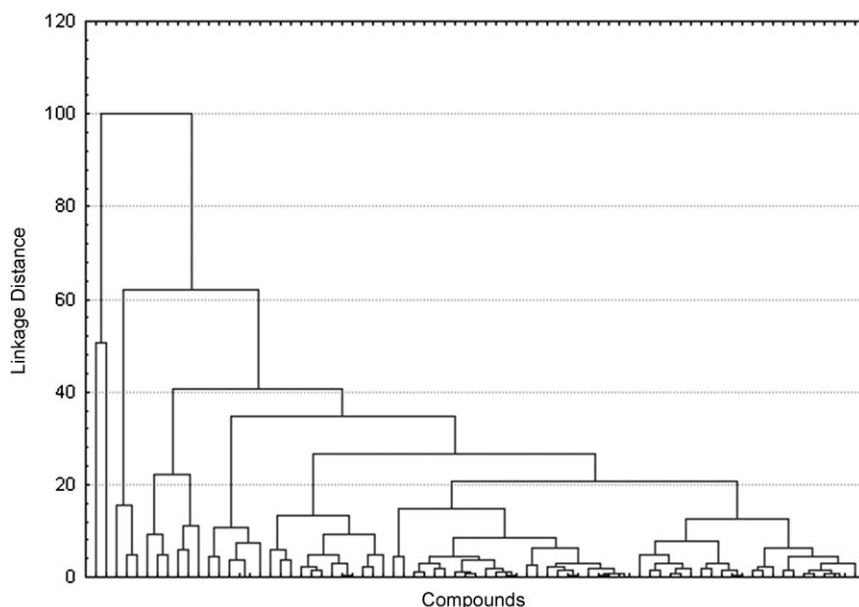
Figure 7. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of active/inactive compounds used for evaluating the predictive ability of the QSAR models for ligand-based virtual screening.

4.4. Computational drug discovery: in silico biological activity modeling and experimental results

As shown, we explored the ability of our classification models to find new subsystems carrying out an experiment of lead generation for the case of tyrosinase inhibitors. These results encouraged us to develop a search for novel active compounds not yet described in the literature as tyrosinase inhibitors. In this context, one of our research teams has been focused mainly on trial-and-error searching for new tyrosinase inhibitors.[14–16] At the same time, we are also identifying new drug candidates using computational screening (based on QSAR techniques). Herein, we perform in silico assays for a cycloartane family isolated and characterized from natural sources (herbal plants), searching novel tyrosinase inhibitors by using the discriminant functions obtained through the TOMOCOMD–CARDD method and LDA technique.

adrenoreceptor studies as well as in the treatment of impotence), one antibacterial molecular entity (BMY-28438), and one analgesic and anti-inflammatory compound (gentisic acid). As can be observed there is also great variability in their molecular structures. This result is the most important validation for the models developed herein because we have demonstrated that they are able to detect a series of drugs as active, and these compounds have shown the predicted activity. This approach is very interesting because the drugs with some pharmacological use that were selected as new lead tyrosinase inhibitors have well-established methods of synthesis, and their toxicological, pharmacodynamic, and pharmaceutical behaviors are also well known.

The LDA-based QSAR models were used to evaluate seven compounds, and in order to corroborate the predictions, they were isolated, and an in vitro assay against the enzyme was carried out. Table 10 lists the %ΔP values of the compounds in the data, as well as their canonical scores using all the devel-
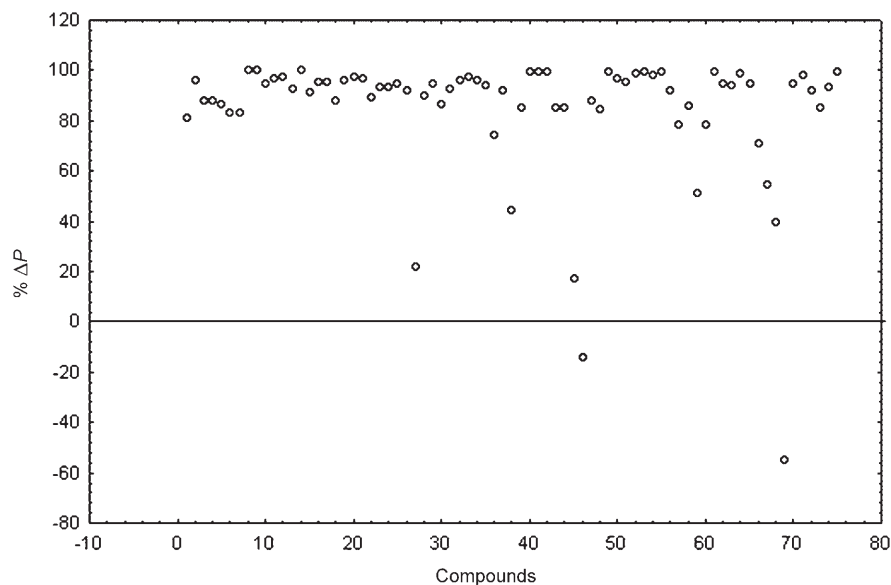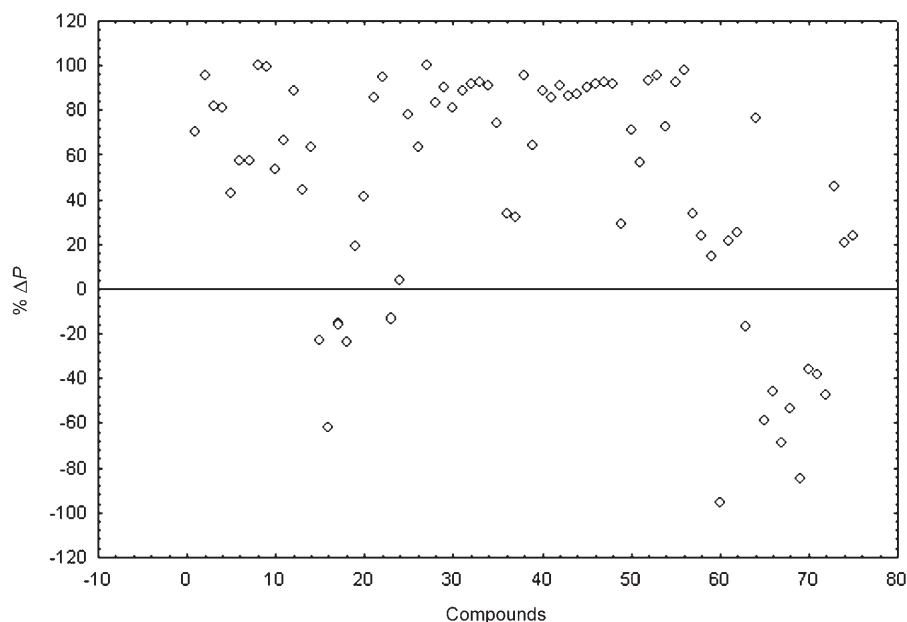
Figure 8. Plot of the % ΔP from [Eq. (26)] (using nonstochastic bilinear indices) for each compound selected in virtual screening protocols. Compounds **1**–**56** and **57**–**75** are active and inactive, respectively.

Figure 9. Plot of the % ΔP from [Eq. (32)] (using stochastic bilinear indices) for each compound selected in virtual screening protocols. Compounds **1**–**56** and **57**–**75** are active and inactive, respectively.

8.32 μM), **CA4** (IC$_{50}$ = 12.09 μM), and **CA7** (IC$_{50}$ = 4.93 μM) exhibited pronounced activities when compared with standard tyrosinase inhibitors like kojic acid (IC$_{50}$ = 16.67 μM) or L-mimosine (IC$_{50}$ = 3.68 μM). In addition, it is important underscore the case of compound **CA6** (IC$_{50}$ = 1.32 μM), which has very potent activity against the enzyme, even compared with the reference drugs. The structures of the compounds are depicted in Figure 10.

Taking into account this result we also made a *k*-NNCA for all the active compounds included in the training, test, and virtual screening sets and the novel compounds. This hierarchical CA analysis was realized with the objective of comparing structural similarities between the newly discovered active compounds and all the active dataset. The resulting dendrogram shows the great degree of structural variability in the subsets of the complete data under investigation (Figure 11). An exhaustive analysis carried out for each cluster showed that these new compounds were included in a cluster together with some steroids such as **68**, **69**, **70**, and **244**, all compounds of the training set. This result is reasonable because the novel organic compounds are methyl steroids, a derived family of steroids. In this way, it is also important to highlight that **CA6** (IC$_{50}$ = 1.32 μM) is very close to compound **244** (IC$_{50}$ = 2.93 μM) in the cluster in correspondence to the similarity

oped models. These results exemplify how the present approach could be used for the selection and identification (lead generation) of novel tyrosinase inhibitors, which may be used to prevent or treat pigmentation disorders.

As can be observed, the theoretical predictions have a generally very good coincidence with the observed activity for all the compounds. All the chemical structures have inhibitory activity toward mushroom tyrosinase. Only one of them, compound **CA5** (IC$_{50}$ = 22.21 μM) showed less activity than kojic acid, a tyrosinase inhibitor reference. The remaining compounds, **CA1** (IC$_{50}$ = 7.92 μM), **CA2** (IC$_{50}$ = 15.94 μM), **CA3** (IC$_{50}$ =

of their structural features and potent tyrosinase inhibitory activity (see Figure 1 of the Supporting Information and Figure 10).

The result is a very promising starting point for the future design optimization for new compounds with higher tyrosinase activity. In this sense, compound **CA6** presented more potent effects in the inhibition of tyrosinase than the reference drug L-mimosine, and this result opens the door to a virtual study of this structural pattern in order to improve it in the search for druglike compounds with tyrosinase inhibitory activity. Besides, this structure can be selected as *hit*, and make a

Table 10. Results of ligand-based in silico screening and tyrosinase inhibitory activities for new cycloartane compounds.[a]

Compound[b]	%ΔP[c]	Scores[d]	%ΔP[c]	Scores[d]	%ΔP[c]	Scores[d]	%ΔP[c]	Scores[d]	%ΔP[c]	Scores[d]	%ΔP[c]	Scores[d]	IC$_{50}$ [μM][e]
CA1	99.79	3.19	99.83	−3.21	99.74	2.96	99.90	−3.73	−100.0	2.57	99.62	2.85	7.92 ± 0.387
	98.01	*−2.25*	*97.74*	*2.17*	*96.24*	*1.88*	*97.53*	*−2.51*	*97.28*	*2.08*	*97.02*	*2.00*	
CA2	99.69	3.03	99.70	−2.95	99.43	2.62	99.77	−3.36	−100.0	2.74	99.36	2.62	15.94 ± 1.93
	99.30	*−2.74*	*98.00*	*2.23*	*95.74*	*1.83*	*94.64*	*−2.09*	*96.29*	*1.94*	*97.22*	*2.03*	
CA3	99.33	2.67	99.20	−2.51	98.62	2.24	99.55	−3.10	−100.0	2.70	98.49	2.25	8.32 ± 0.097
	88.25	*−1.41*	*85.68*	*1.27*	*85.07*	*1.22*	*88.23*	*−1.65*	*90.49*	*1.48*	*90.68*	*1.46*	
CA4	99.22	2.59	98.93	−2.38	98.38	2.17	99.49	−3.03	−100.0	2.70	98.05	2.13	12.09 ± 1.03
	81.39	*−1.18*	*81.00*	*1.12*	*79.65*	*1.07*	*87.72*	*−1.63*	*88.86*	*1.40*	*87.40*	*1.31*	
CA5	99.47	2.77	99.41	−2.65	99.07	2.41	99.64	−3.21	−100.0	2.68	98.83	2.35	22.21 ± 1.94
	95.79	*−1.90*	*93.39*	*1.66*	*82.75*	*1.15*	*88.21*	*−1.65*	*89.82*	*1.45*	*84.31*	*1.20*	
CA6	99.91	3.59	99.95	−3.71	99.91	3.39	99.94	−3.93	−100.0	2.71	99.78	3.09	1.32 ± 0.373
	99.28	*−2.72*	*99.29*	*2.73*	*98.70*	*2.38*	*98.54*	*−2.79*	*99.30*	*2.72*	*98.99*	*2.50*	
CA7	99.71	3.05	99.76	−3.04	99.57	2.73	99.80	−3.40	−100.0	2.72	99.17	2.51	4.93 ± 0.197
	98.39	*−2.35*	*97.48*	*2.12*	*96.02*	*1.86*	*97.12*	*−2.43*	*97.28*	*2.08*	*98.25*	*2.24*	

[a] For each compound two rows of data are listed; upper row: classification of each compound using the obtained models with nonstochastic bilinear indices in the following order: Eqs. (22), (23), (24), (25), (26), and (27); lower row (in *italic*); classification of each compound using the obtained models with stochastic bilinear indices in the following order: Eqs. (28), (29), (30), (31), (32), and (33). [b] Molecular structures of these compounds are shown in Figure 10. [c] %ΔP = [P(Active)−P(Inactive)] 100. [d] Canonical scores. [e] 50% Inhibitory concentration (±SEM) against tyrosinase.
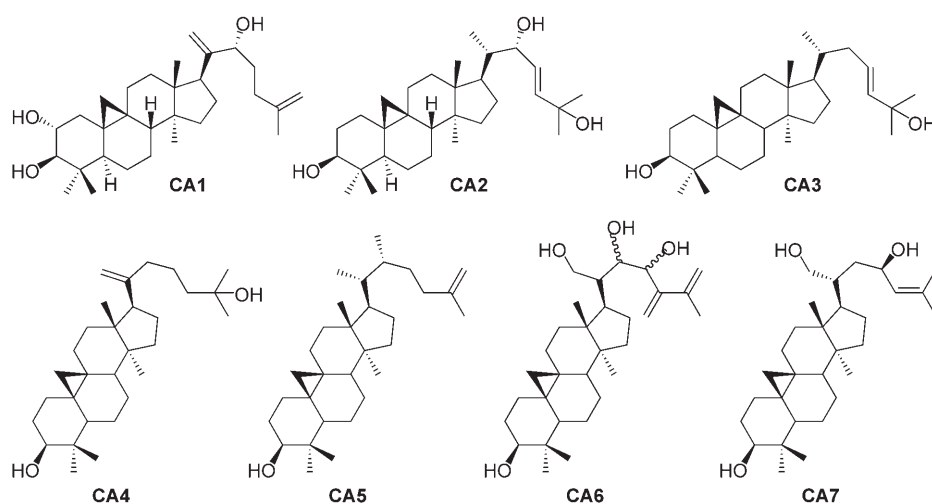

Figure 10. Molecular structures of the new cycloartane compounds.

novel candidates to be assayed as druglike compounds.[18,19] In this sense, the introduction and use of graph-theoretical MDs for rational drug design has become an attractive tool for medicinal chemists. The fusion of HTS and classification-based QSAR models in an attempt to minimize costs in terms of time, finances, and human and animal resources is becoming a viable alternative to massive screening. In this way and taking into account that most of the known tyrosinase inhibitor compounds have been discovered by empirical ways (trial-and-error methods), we make use of a QSAR approach for the biological evaluation of new compounds isolated and characterized from herbal plants.

Herein we present a new set of molecular descriptors, namely nonstochastic and stochastic bilinear indices, implemented in TOMOCOMD–CARDD, and their application in the rational selection of new active compounds against tyrosinase. In addition, we show that the new molecular descriptors implemented in the TOMOCOMD–CARDD approach can be applied to generate useful discriminant models for the classification of compounds as tyrosinase inhibitors. This method permits the prediction of the biological property under consideration, thus increasing the likelihood of an in silico discovery of new candidate lead compounds and minimizing the use of resources. Considering a training data set of compounds with considerable structural variation, we decrease the degree of uncertainty for this process. These collected data of the active compounds can be used by all researchers as an important
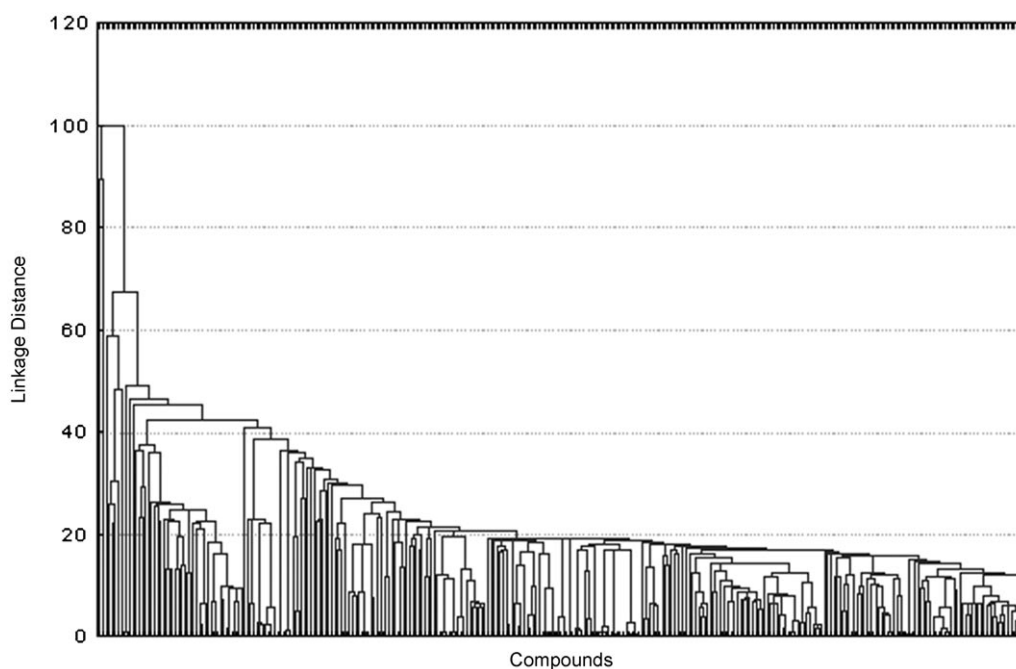
chemistry optimization to find the appropriate combination of lead activity, pharmacological properties, toxicity, and good behavior in clinical animal assays. It is important to recall that the aim of this study is not only to validate the model but also to provide an experimental example of how to use the model for potential drug discovery.

5. Conclusion

Tyrosinase attracts scientific interest in the search for new compounds that inhibit its activity, owing to its important role in hyperpigmentation and melanogenesis disorders.[5–8] The discovery and characterization of these novel tyrosinase inhibitors have become an important area of study for their potential applications. In recognition of the complexity and cost of the process of drug discovery, the use of "rational" search methodologies is recommended. Consequently, medicinal chemists are called to develop more efficient strategies for the search of

Figure 11. A dendrogram illustrating the results of the hierarchical *k*-NNCA of the set of all active compounds (tyrosinase inhibitors) included in the training, test, virtual screening, and discovery of new active cycloartanes in the work presented herein.

tool not only for theoretical research but also for general scientific work in this area. The simulated virtual screen of compounds for tyrosinase inhibitor activity has proven the ability of our models to adequately discriminate new active compounds from inactives and the possibilities of in silico identification of novel tyrosinase inhibitors.

The novel MDs together with pattern-recognition techniques can be used for increasing the rate at which new lead-like compounds are discovered on the way to experimental screening and in vitro pharmacological assays. An experimental corroboration was made through the isolation, characterization, and assay for tyrosinase inhibition of the seven reported new organic compounds. Research in the field of natural product chemistry has not only formed the scientific basis for the traditional use of medicinal plants, but also plays an important role in the discovery of pharmaceutical, nutraceutical, and cosmetic bioactive ingredients.[116] For that reason, a symbiosis between methods in natural product chemistry and in silico approaches can provide new clues to the process of rational drug design, as well as a starting point in the search for not only increased efficacy, but also for more potent tyrosinase inhibitors, which may be used to treat the disorders of hyperpigmentation and melanogenesis.

Acknowledgements

Keywords: atom-based bilinear indices · computer chemistry · cycloartanes · ligand-based virtual screening · tyrosinase inhibitors

[1] J. C. Garcia-Borron, F. Solano, *Pigm. Cell Res.* **2002**, *15*, 162–173.
[2] P. A. Riley, *Pigm. Cell Res.* **1993**, *6*, 182–185.
[3] O. Nerya, J. Vaya, R. Musa, S. Izrael, R. Ben-Arie, S. Tamir, *J. Agric. Food Chem.* **2003**, *51*, 1201–1207.
[4] W. S. Oetting, R. A. King, *Hum. Mutat.* **1999**, *13*, 99–115.
[5] I. L. Guevara, A. G. Pandya, *Int. J. Dermatol.* **2001**, *40*, 212–215.
[6] Q. H. Nguyen, T. P. Bui, *Int. J. Dermatol.* **1995**, *34*, 75–84.
[7] G. Okan, C. Baykal, *J. Eur. Acad. Dermatol. Venereol.* **1999**, *13*, 218–220.
[8] K. Maeda, M. Fukuda, *J. Pharmacol. Exp. Ther.* **1996**, *276*, 765–769.
[9] J. Abramowitz, W. Chavin, *Chem.-Biol. Interact.* **1980**, *32*, 195–208.
[10] E. V. Curto, C. Kwong, H. Hermersdorfer, H. Glatt, C. Santis, V. Virador, V. J. Hearing, Jr., T. P. Dooley, *Biochem. Pharmacol.* **1999**, *57*, 663–672.
[11] K. B. Penney, C. J. Smith, J. C. Allen, *J. Invest. Dermatol.* **1984**, *82*, 308–310.
[12] K. Jones, J. Hughes, M. Hong, Q. Jia, S. Orndorff, *Pigm. Cell Res.* **2002**, *15*, 335–340.
[13] K. T. Lee, B. J. Kim, J. H. Kim, *Int. J. Cosmet. Sci.* **1997**, *19*, 291–298.
[14] M. T. Khan, M. I. Choudhary, K. M. Khan, M. Rani, Atta-ur-Rahman, *Bioorg. Med. Chem.* **2005**, *13*, 3385–3395.
[15] M. I. Choudhary, S. Sultan, M. T. H. Khan, A. Yasin, F. Shaheen, Atta-ur-Rahman, *Nat. Prod. Res.* **2004**, *18*, 529–535.

[16] V. U. Ahmad, F. Ullah, J. Hussain, U. Farooq, M. Zubair, M. T. Khan, M. I. Choudhary, *Chem. Pharm. Bull.* **2004**, *52*, 1458–1461.

[17] Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E. A. Castro, *Bioorg. Med. Chem.* **2005**, *13*, 2881–2899.

[18] J. Xu, A. Hagler, *Molecules* **2002**, *7*, 566–700.

[19] H. J. M. Seifert, K. Wolf, D. Vitt, *BioSilico* **2003**, *1*, 143–149.

[20] K. S. Dixit, S. N. Mitra, *CRIPS* **2002**, *3*, 1–7.

[21] E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana, E. De Clercq, *J. Med. Chem.* **2000**, *43*, 1975–1985.

[22] H. Gonzáles-Díaz, Y. Marrero-Ponce, I. Hernãndez, I. Bastida, E. Tenorio, O. Nasco, E. Uriarte, N. Castañedo, M. A. Cabrera, E. Aguila, O. Marrero, A. Morales, M. Perez, *Chem. Res. Toxicol.* **2003**, *16*, 1318–1327.

[23] M. Lajiness in *Computational Chemical Graph Theory* (Ed.: D. H. Rouvray), Nova Science, New York, **1990**.

[24] E. Estrada, A. Peña, R. García-Domenech, *J. Comput.-Aided Mol. Des.* **1998**, *12*, 583–595.

[25] H. González-Díaz, E. Olazábal, N. Castañedo, I. Hernández, A. Morales, H. S. Serrano, J. González, R. Ramos de Armas, *J. Mol. Model.* **2002**, *8*, 237–245.

[26] Y. Marrero-Ponce, V. Romero, *TOMOCOMD software*, **2002**, Central University of Las Villas. TOMOCOMD version 1.0, for Windows, is a preliminary experimental version; in future a professional version can be obtained upon request to Y. Marrero-Ponce: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es.

[27] Y. Marrero-Ponce, *Molecules* **2003**, *8*, 687–726.

[28] Y. Marrero-Ponce, *Bioorg. Med. Chem.* **2004**, *12*, 6351–6369.

[29] Y. Marrero-Ponce, J. A. Castillo-Garit, F. Torrens, V. Romero-Zaldivar, E. Castro, *Molecules* **2004**, *9*, 1100–1123.

[30] Y. Marrero-Ponce, M. A. Cabrera, V. Romero, E. Ofori, L. A. Montero, *Int. J. Mol. Sci.* **2003**, *4*, 512–536.

[31] Y. Marrero-Ponce, M. A. Cabrera, V. Romero, D. H. González, F. Torrens, *J. Pharm. Pharm. Sci.* **2004**, *7*, 186–199.

[32] Y. Marrero-Ponce, M. A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio, F. Torrens, *Internet Electron. J. Mol. Des.* **2005**, *4* 124–150.

[33] Y. Marrero-Ponce, H. G. Díaz, V. Romero, F. Torrens, E. A. Castro, *Bioorg. Med. Chem.* **2004**, *12*, 5331–5342.

[34] Y. Marrero-Ponce, R. Medina, E. A. Castro, R. de Armas, H. González, V. Romero, F. Torrens, *Molecules* **2004**, *9*, 1124–1147.

[35] Y. Marrero-Ponce, R. Medina-Marrero, J. A. Castillo-Garit, V. Romero-Zaldivar, F. Torrens, E. A. Castro, *Bioorg. Med. Chem.* **2005**, *13*, 3003–3015.

[36] Y. Marrero-Ponce, D. Nodarse, H. D. González, R. Ramos de Armas, V. Romero-Zaldivar, F. Torrens, E. Castro, *Int. J. Mol. Sci.* **2004**, *5*, 276–293.

[37] Y. Marrero-Ponce, J. A. Castillo-Garit, D. Nodarse, *Bioorg. Med. Chem.* **2005**, *13*, 3397–3404.

[38] Y. Marrero-Ponce, M. Iyarreta-Veitia, A. Montero-Torres, C. Romero-Zaldivar, C. A. Brandt, P. E. Avila, K. Kirchgatter, Y. Machado, *J. Chem. Inf. Model.* **2005**, *45*, 1082–1100.

[39] J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, R. Rotondo, *J. Mol. Graphics Modell.* **2006**, DOI: 10.1016/j.jmgm.2006.09.007.

[40] Y. Marrero-Ponce, A. Meneses-Marcel, J. A. Castillo-Garit, Y. Machado-Tugores, J. A. Escario, A. G. Barrio, D. M. Pereira, J. J. Nogal-Ruiz, V. J. Arán, A. R. Martínez-Fernández, F. Torrens, R. Rotondo, F. Ibarra-Velarde, Y. J. Alvarado, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 6502–6524.

[41] Y. Marrero-Ponce, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010–2026.

[42] Y. Marrero-Ponce, A. Huesca-Guillen, F. Ibarra-Velarde, *J. Mol. Struct. (Theochem)* **2005**, *717*, 67–79.

[43] Y. Marrero-Ponce, J. A. Castillo-Garit, *J. Comput.-Aided Mol. Des.* **2005**, *19*, 369–383.

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

[45] Y. Marrero-Ponce, A. Montero-Torres, C. R. Zaldivar, M. I. Veitia, M. M. Perez, R. N. Sanchez, *Bioorg. Med. Chem.* **2005**, *13*, 1293–1304.

[46] Y. Marrero-Ponce, R. Medina-Marrero, Y. Martinez, F. Torrens, V. Romero-Zaldivar, E. A. Castro, *J. Mol. Model.* **2006**, *12*, 255–271.

[47] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazábal, H. S. Serrano, A. Morales, N. Castañedo, F. Ibarra-Velarde, A. Huesca-Guillen, A. M. Sanchez, F. Torrens, E. A. Castro, *Bioorg. Med. Chem.* **2005**, *13*, 1005–1020.

[48] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazábal, H. S. Serrano, A. Morales, N. Castañedo, F. Ibarra-Velarde, A. Huesca-Guillen, E. Jorge, A. del Valle, F. Torrens, E. A. Castro, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 615–634.

[49] A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado-Tugores, A. Montero-Torres, D. M. Pereira, J. A. Escario, J. J. Nogal-Ruiz, C. Ochoa, V. J. Aran, A. R. Martinez-Fernandez, R. N. Garcia Sanchez, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3838–3843.

[50] R. Wang, Y. Gao, L. Lai, *Perspect. Drug Discovery Des.* **2000**, *19*, 47–66.

[51] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **2000**, *43*, 3714–3717.

[52] A. K. Ghose, G. M. Crippen, *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35.

[53] K. J. Miller, *J. Am. Chem. Soc.* **1990**, *112*, 8533–8542.

[54] J. Gasteiger, M. Marsili, *Tetrahedron Lett.* **1978**, *19*, 3181–3184.

[55] L. B. Kier, L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, UK, **1986**.

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

[57] N. Trinajstić, *Chemical Graph Theory*, 2nd ed., CRC, Boca Raton, **1992**, 32–33.

[58] I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, **1986**.

[59] D. J. Klein, *Internet Electron. J. Mol. Des.* **2003**, *2*, 814–834.

[60] N. Jacobson, *Basic Algebra I*, 2nd ed., W. H. Freeman and Company, New York, **1985**, pp. 343–361.

[61] K. F. Riley, M. P. Hobson, S. J. Vence, *Mathematical Methods for Physics and Engineering*, Cambridge University Press, Cambridge, **1998**, pp. 228–236.

[62] E. Hernández, *Álgebra y Geometría*, Universidad Autónoma de Madrid, Madrid, **1987**, pp. 521–544.

[63] J. de Burgos-Román, *Álgebra y Geometría Cartesíana*, 2da ed., McGraw-Hill Interamericana de España, Madrid, **2000**, pp. 208–246.

[64] J. de Burgos-Román, *Curso de Álgebra y Geometría*, Alambra Longman, Madrid, **1994**, pp. 638–684.

[65] G. Werner, *Linear Algebra*, 4th ed., Springer, New York, **1981**, pp. 261–288.

[66] Y. Marrero-Ponce, F. Torrens, *J. Chem. Inf. Model.* **2006**, submitted for publication.

[67] M. Randić, *J. Math. Chem.* **1991**, *7*, 155–168.

[68] P. D. Walker, P. G. Mezey, *J. Am. Chem. Soc.* **1993**, *115*, 12423–12430.

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

[70] L. H. Hall, C. T. Story, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004–1014.

[71] J. D. Gough, L. H. Hall, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 356–361.

[72] V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.

[73] R. Todeschini, P. Gramatica, *Perspect. Drug Discovery Des.* **1998**, *9–11*, 355–380.

[74] L. Pauling, *The Nature of Chemical Bond*, Cornell University Press, Ithaca, **1939**, pp. 2–60.

[75] O. Nerya, R. Musa, S. Khatib, S. Tamir, J. Vaya, *Phytochemistry* **2004**, *65*, 1389–1395.

[76] S. Khatib, O. Nerya, R. Musa, M. Shmuel, S. Tamir, J. Vaya, *Bioorg. Med. Chem.* **2005**, *13*, 433–441.

[77] K. M. Khan, G. M. Maharvi, A. Abbaskhan, S. Hayat, M. T. H. Khan, T. Makhmoor, M. I. Choudhary, F. Shaheen, Atta-ur-Rahman, *Helv. Chim. Acta* **2003**, *86*, 457–464.

[78] K. Komori, K. Yatagai, T. Tatsuma, *J. Biotechnol.* **2004**, *108*, 11–16.

[79] H. Kim, J. Choi, J. K. Cho, S. Y. Kim, Y. S. Lee, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2843–2846.

[80] S. J. Um, M. S. Park, S. H. Park, H. S. Han, Y. J. Kwon, H. S. Sin, *Bioorg. Med. Chem.* **2003**, *11*, 5345–5352.

[81] M. Shiino, Y. Watanabe, K. Umezawa, *Bioorg. Med. Chem.* **2001**, *9*, 1233–1240.

[82] M. Shiino, Y. Watanabe, K. Umezawa, *Bioorg. Chem.* **2003**, *32*, 129–135.

[83] J. K. No, D. Y. Soung, Y. J. Kim, K. H. Shim, Y. S. Jun, S. H. Rhee, T. Yokozawa, H. Y. Chung, *Life Sci.* **1999**, *65*, PL241–246.

[84] Y. M. Kim, J. Yun, C. K. Lee, H. Lee, K. R. Min, Y. Kim, *J. Biol. Chem.* **2002**, *277*, 16340–16344.

[85] J. P. Ley, H. J. Bertram, *Bioorg. Med. Chem.* **2001**, *9*, 1879–1885.

[86] H. S. Lee, *J. Agric. Food Chem.* **2002**, *50*, 1400–1403.

[87] K. Iwai, N. Kishimoto, Y. Kakino, K. Mochida, T. Fujita, *J. Agric. Food Chem.* **2004**, *52*, 4893–4898.

[88] N. Yokochi, T. Morita, T. Yagi, *J. Agric. Food Chem.* **2003**, *51*, 2733–2736.

[89] L. P. Xie, Q. X. Chen, H. Huang, H. Z. Wang, R. Q. Zhang, *Biochemistry* **2003**, *68*, 487–491.

[90] H. S. Kang, H. R. Kim, D. S. Byun, B. W. Son, T. J. Nam, J. S. Choi, *Arch. Pharmacal Res.* **2004**, *27*, 1226–1232.

[91] M. I. Choudhary, S. G. Musharraf, M. T. H. Khan, D. Abdelrahman, M. Parvez, F. Shaheen, Atta-ur-Rahman, *Helv. Chim. Acta* **2003**, *86*, 3450–3460.

[92] T. Şabudak, M. T. H. Khan, M. I. Choudhary, S. Oksuz, *Nat. Prod. Res.* **2006**, DOI: 10.1080/14786410500196821.

[93] M. Negwer, *Organic-Chemical Drugs and their Synonyms*, Akademie-Verlag, Berlin, **1987**.

[94] E. Estrada, A. Peña, *Bioorg. Med. Chem.* **2000**, *8*, 2755–2770.

[95] J. W. McFarland, D. J. Gans in *Chemometric Methods in Molecular Design* (Ed.: H. van de Waterbeemd), VCH, Weinheim, **1995**, pp. 295–307.

[96] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey, **1988**.

[97] STATISTICA (data analysis software system) version 6.0, StatSoft Inc., **2001**.

[98] M. J. Duart, R. García-Domenech, G. M. Anton-Fos, J. Galvez, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 561–572.

[99] H. van de Waterbeemd in *Chemometric Methods in Molecular Design* (Ed.: H. van de Waterbeemd), VCH, Weinheim, **1995**, pp. 265–288.

[100] J. V. de Julian-Ortiz, C. G. deAlapont, I. Ríos-Santamarina, R. García-Domenech, E. Galvez, *J. Mol. Graphics Modell.* **1998**, *16*, 14–18.

[101] J. Gálvez, R. García, M. T. Salabert, R. Soler, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.

[102] E. Estrada, S. Vilar, E. Uriarte, Y. Gutierrez, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1194–1203.

[103] M. Randić, *J. Mol. Struct. (Theochem)* **1991**, *233*, 45–59.

[104] M. Randić, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311–320.

[105] M. Randić, *New J. Chem.* **1991**, *15*, 517–525.

[106] B. Lučić, S. Nikolić, N. Trinajstić, D. Jurić, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532–538.

[107] D. J. Klein, M. Randić, D. Babić, B. Lučić, S. Nikolić, N. Trinajstić, *Int. J. Quantum Chem.* **1997**, *63*, 215–222.

[108] E. Estrada, E. Uriarte, *Curr. Med. Chem.* **2001**, *8*, 1573–1588.

[109] M. T. H. Khan, M. I. Choudhary, Atta-ur-Rahman, R. P. Mamedova, M. A. Agzamova, M. N. Sultankhodzhaev, M. I. Isaev, *Bioorg. Med. Chem.* **2006**, *14*, 6085–6088.

[110] V. J. Hearing, *Methods in Enzymology*, Academic Press, New York, **1987**, pp. 154.

[111] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, H. Nielsen, *Bioinformatics* **2000**, *16*, 412–424.

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

[113] A. Golbraikh, A. Tropsha, *Mol. Diversity* **2002**, *5*, 231–243.

[114] J. V. de Julian-Ortiz, *Comb. Chem. High Throughput Screening* **2001**, *4*, 295–310.

[115] D. F. Horrobin, *J. R. Soc. Med.* **2000**, *93*, 341–345.

[116] Q. Jia, *Aloe Plant* **2003**, http://www.aloecorp.com/email/0203GuestDrQiJia.pdf.

[117] K. Ohguchi, T. Tanaka, T. Kido, K. Baba, M. Iinuma, K. Matsumoto, Y. Akao, Y. Nozawa, *Biochem. Biophys. Res. Commun.* **2003**, *307*, 861–863.

[118] D. Barlocco, D. Barrett, P. Edwards, S. Langston, M. J. Pérez-Pérez, M. Walker, J. Weidner, A. Westwell, *Drug Discovery Today* **2003**, *8*, 372–373.

[119] T. J. Ha, M. S. Yang, D. S. Jang, S. U. Choi, K. H. Park, *Bull. Korean Chem. Soc.* **2001**, *22*, 97.

[120] K. Nihei, I. Kubo, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2409–2412.

[121] K. Likhitwitayawuid, B. Sritularak, *J. Nat. Prod.* **2001**, *64*, 1457–1459.

[122] C. Y. Li, T. S. Wu, *J. Nat. Prod.* **2002**, *65*, 1452–1456.

[123] L. G. Fenoll, P. A. Garcia-Ruiz, R. Varon, F. Garcia-Canovas, *J. Agric. Food Chem.* **2003**, *51*, 7781–7787.

[124] A. Poma, G. Pacioni, S. Colafarina, M. Miranda, *FEMS Microbiol. Lett.* **1999**, *180*, 69–75.

[125] P. Bernard, J.-Y. Berthon, *Int. J. Cosmet. Sci.* **2000**, *22*, 219–226.

[126] L. C. Wu, Y. C. Chen, J. A. Ho, C. S. Yang, *J. Agric. Food Chem.* **2003**, *51*, 4240–4246.

[127] B. C. Behera, U. Makhija, *Curr. Sci.* **2002**, *82*, 61–66.

[128] H. S. VKang, H. R. Kim, D. S. Byun, B. W. Son, T. J. Nam, J. S. Choi, *Arch. Pharmacal Res.* **2004**, *12*, 1226–1232.

[129] K. H. Kong, M. P. Hong, S. S. Choi, Y. T. Kim, S. H. Cho, *Biotechnol. Appl. Biochem.* **2000**, *31(Pt 2)*, 113–118.

[130] I. Kubo, K. Niheia, K. Tsujimoto, *Bioorg. Med. Chem.* **2004**, *12*, 5349–5354.

[131] I. Hori, K. Nihei, I. Kubo, *Phytother. Res.* **2004**, *18*, 475–479.

[132] M. van Gastel, L. Bubacco, E. J. Groenen, E. Vijgenboom, G. W. Canters, *FEBS Lett.* **2000**, *474*, 228–232.

[133] A. Palumbo, M. d'Ischia, G. Misuraca, G. Prota, *Biochim. Biophys. Acta* **1991**, *1073*, 85–90.

[134] J. C. Espín, R. H. Veltmanb, H. J. Wichersb, *Physiol. Plant.* **2000**, *109*, 1–6.

[135] L. A. Mueller, U. Hinz, J.-P. Zryd, *Phytochemistry* **1996**, *42*, 1511–1515.

[136] M. Jimenez, F. Garcia-Carmona, *Arch. Biochem. Biophys.* **2000**, *373*, 255–260.

[137] P. D. Thomas, H. Kishi, H. Cao, M. Ota, T. Yamashita, S. Singh, K. Jimbow, *J. Invest. Dermatol.* **1999**, *113*, 928–934.

[138] N.-H. Shin, S. Y. Ryu, E. J. Choi, S.-H. Kang, I.-M. Chang, K. R. Min, Y. Kim, *Biochem. Biophys. Res. Commun.* **1998**, *243*, 801–803.

[139] Q. X. Chen, I. Kubo, *J. Agric. Food Chem.* **2002**, *50*, 4108–4112.

[140] B. Gasowska, P. Kafarski, H. Wojtasek, *Biochim. Biophys. Acta* **2004**, *1673*, 170–177.

[141] S. E. Stanca, I. C. Popescu, *J. Mol. Catal. B* **2004**, *27*, 221–225.

[142] S. Gutteridge, D. Robb, *Eur. J. Biochem.* **1975**, *54*, 107–116.

[143] S. Menon, R. W. Fleck, G. Yong, K. G. Strothkamp, *Arch. Biochem. Biophys.* **1990**, *280*, 27–32.

[144] M. Jimenez, S. Chazarra, J. Escribano, J. Cabanes, F. Garcia-Carmona, *J. Agric. Food Chem.* **2001**, *49*, 4060–4063.

[145] L. P. Xie, Q. X. Chen, H. Huang, X. D. Liu, H. T. Chen, R. Q. Zhang, *Int. J. Biochem. Cell Biol.* **2003**, *35*, 1658–1666.

[146] R. Gilly, D. Mara, S. Oded, K. Zohar, *J. Agric. Food Chem.* **2001**, *49*, 1479–1485.

[147] I. Kubo, I. Kinst-Hori, *J. Agric. Food Chem.* **1998**, *46*, 1268–1271.

[148] http://www.thecosmeticsite.com/formulating/959621.html.

[149] H. Takahashi, P. G. Parsons, *J. Invest. Dermatol.* **1992**, *98*, 481–487.

[150] S. Naish-Byfield, C. J. Cooksey, P. A. Riley, *Biochem. J.* **1994**, *304 (Pt. 1)*, 155–162.

[151] K. Nihei, Y. Yamagiwa, T. Kamikawa, I. Kubo, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 681–683.

[152] I. Kubo, Q. X. Chen, K. Nihei, J. S. Calderon, C. L. Cespedes, *Z. Naturforsch. C* **2003**, *58*, 713–718.

[153] I. Kubo, I. Kinst-Hori, *J. Agric. Food Chem.* **1999**, *47*, 4121–4125.

[154] M. Miyazawa, T. Oshima, K. Koshio, Y. Itsuzaki, J. Anzai, *J. Agric. Food Chem.* **2003**, *51*, 6953–6956.

[155] J. Cabanes, F. Garcia-Carmona, F. Garcia-Canovas, J. L. Iborra, J. A. Lozano, *Biochim. Biophys. Acta* **1984**, *790*, 101–107.

[156] B. Gasowska, H. Wojtasek, J. Hurek, M. Drag, K. Nowak, P. Kafarski, *Eur. J. Biochem.* **2002**, *269*, 4098–4104.

[157] F. Karbassi, A. A. Saboury, M. T. Khan, M. I. Choudhary, Z. S. Saifi, *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 349–353.

[158] S. Shareefi Borojerdi, K. Haghbeen, A. A. Karkhane, M. Fazli, A. A. Sabo[u]ryc, *Biochem. Biophys. Res. Commun.* **2004**, *314*, 925–930.

[159] H. S. Kang, J. H. Choi, W. K. Cho, J. C. Park, J. S. Choi, *Arch. Pharmacal Res.* **2004**, *27*, 742–750.

[160] S. B. Khan, S. Azhar Ul, N. Afza, A. Malik, M. T. Khan, M. R. Shah, M. I. Choudhary, *Chem. Pharm. Bull.* **2005**, *53*, 86–89.

[161] Y. Masamoto, H. Ando, Y. Murata, Y. Shimoishi, M. Tada, K. Takahata, *Biosci. Biotechnol. Biochem.* **2003**, *67*, 631–634.

[162] F. Shaheen, M. Ahmad, M. T. Khan, S. Jalil, A. Ejaz, M. N. Sultankhodjaev, M. Arfan, M. I. Choudhary, Atta-ur-Rahman, *Phytochemistry* **2005**, *66*, 935–940.

[163] L. Bubacco, M. Van Gastel, E. J. Groenen, E. Vijgenboom, G. W. Canters, *J. Biol. Chem.* **2003**, *278*, 7381–7389.

[164] http://open.cacb.org.tw/index.php.

[165] M. Koketsu, S. Y. Choi, H. Ishihara, B. O. Lim, H. Kim, S. Y. Kim, *Chem. Pharm. Bull.* **2002**, *50*, 1594–1596.

[166] J. M. Wood, K. U. Schallreuter-Wood, N. J. Lindsey, S. Callaghan, M. L. Gardner, *Biochem. Biophys. Res. Commun.* **1995**, *206*, 480–485.

[167] M. Nazzaro-Porro, S. Passi, *J. Invest. Dermatol.* **1978**, *71*, 205–208.

[168] V. K. Sharma, J. Choi, N. Sharma, M. Choi, S. Y. Seo, *Phytother. Res.* **2004**, *18*, 841–844.

[169] K. Sakuma, M. Ogawa, K. Sugibayashi, K. Yamada, K. Yamamoto, *Arch. Pharmacal Res.* **1999**, *22*, 335–339.

[170] R. A. Løvstad, *Biochem. Pharmacol.* **1976**, *25*, 533–535.

[171] I. Kubo, I. Kinst-Hori, Y. Yokokawa, *J. Nat. Prod.* **1994**, *57*, 545–551.

[172] G. Regev-Shoshani, O. Shoseyov, I. Bilkis, Z. Kerem, *Biochem. J.* **2003**, *374*, 157–163.

[173] C. Imada, Y. Sugimoto, T. Makimura, T. Kobayashi, N. Hamada, E. Watanabe, *Fish. Sci.* **2001**, *67*, 1151–1156.

[174] J. C. Espín, H. J. Wichers, *Biochim. Biophys. Acta* **2001**, *1544*, 289–300.

[175] B. B. Fuller, M. A. Drake, D. T. Spaulding, F. Chaudhry, *J. Invest. Dermatol.* **2000**, *114*, 268–276.

[176] I. Kubo, Q. Chen, K. Nihei, *Food Chem.* **2003**, *81*, 241–247.

[177] M. McEwan, P. G. Garsons, *J. Invest. Dermatol.* **1990**, *89*, 82–86.