Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 324–330

New tyrosinase inhibitors selected by atomic linear indices-based classification models

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

[a]*Department of Pharmacy, Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba*
[b]*Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba*
[c]*Department of Biological Sciences, Faculty of Agricultural Sciences, University of Ciego de Avila, 69450, Ciego de Avila, Cuba*
[d]*Pharmacology Research Laboratory, Faculty of Pharmaceutical Sciences, University of Science and Technology Chittagong, Chittagong, Bangladesh*
[e]*Department of Biochemistry and Molecular Biology, Center for Biotechnology, University of Ferrara, Via L. Borsari, 46, FE 44100 Ferrara, Italy*
[f]*Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (València), Spain*
[g]*S. Yunusov Institute of Chemistry of Plant Substances, Academy of Sciences, Tashkent, Uzbekistan*

Received 9 September 2005; revised 28 September 2005; accepted 29 September 2005
Available online 3 November 2005

Abstract—In the present report, the use of the atom-based linear indices for finding functions that discriminate between the tyrosinase inhibitor compounds and inactive ones is presented. In this sense, discriminant models were applied and globally good classifications of 93.51% and 92.46% were observed for non-stochastic and stochastic linear indices best models, respectively, in the training set. The external prediction sets had accuracies of 91.67% and 89.44%. In addition, these fitted models were used in the screening of new cycloartane compounds isolated from herbal plants. A good behavior is shown between the theoretical and experimental results. These results provide a tool that can be used in the identification of new tyrosinase inhibitor compounds.

Tyrosinase is the key enzyme in melanin biosynthesis, catalyzing the first two steps of this pathway: the hydroxylation in the ortho position of tyrosine (monophenolase or cresolase activity) and the oxidation of L-DOPA (L-3,4-dihydroxyphenylalanine) to *o*-dopaquinone (diphenolase or catecholase activity), both in the presence of molecular oxygen. It is a copper protein widely distributed in nature, which shows similar structural and functional characteristics when purified from different biological sources.[1–3]

Because of its central role in melanogenesis, tyrosinase is a key target for screening and discovery of new inhibitory compounds is underway in the hope of preventing the occurrence of these hyperpigmentation disorders.[4,5]

Compounds such as hydroquinone,[6] ascorbic acid derivates,[7] kojic acid,[8] azelaic acid,[9] corticosteroids,[10] retinoids,[11] arbutin,[12] and others have been reported to show the inhibitory efficacy. Although a large number of naturally occurring tyrosinase inhibitors have already been described,[13] their individual activities either are not potent enough to be considered for practical use or safety regulation concerning food additives limits their in vivo use. There is, therefore, a constant search for tyrosinase inhibitors that can be obtained by either laboratory synthesis[14] or extraction from plants.[15,16]

On the other hand, for pharmaceutical research and development, chemoinformatics provides, at present, the tools for 'rational' selection/identification and/or design/optimization of new chemical entities (NCE), reducing the number of tested compounds, compared with conventional trial-and-error methods.[17]

Recently, a novel scheme for the rational-in silico-molecular design (or selection/identification of chemi-

* Corresponding author. Tel.: +53 42 281192; fax: +53 42 281130; e-mail addresses: yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es
URL: http://www.uv.es/yoma/.

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.09.085

cals) and QSAR/QSPR studies has been introduced by one of our research groups. It is the so-called topological molecular computer design (TOMOCOMD).[18] This method has been developed to generate molecular descriptors based on the linear algebra theory. This approach has been successfully employed in QSPR[19,20] and QSAR[21,22] studies, including investigations related to nucleic acid–drug interactions[23] and the fast-track experimental discovery of novel antimalarial compounds.[24]

The main objective of this research was to find various statistical linear discriminant analysis (LDA) models, using the non-stochastic (and stochastic) total and atom-type linear indices in order to separate the tyrosinase inhibitor compounds (actives) from inactive ones, with the aim to power the early identification of potential tyrosinase inhibitors, isolated and characterized from herbal plants.

To assure an adequate extrapolation power for the LDA models, a data set with a great molecular diversity was chosen. We have selected 658 compounds for making up the data set, 246 with tyrosinase inhibitor activity, considering different modes of inhibition, and the rest, 412, having a series of other pharmacological uses[25] (inactives).

The molecular descriptors, non-stochastic and stochastic atom-based linear indices, were calculated using the 'in house' TOMOCOMD-CARDD (acronym of the Computed-Aided Rational Drug Design) software. The total and local linear indices for small-to-medium sized organic compounds have been explained in some detail in the literature.[26–29] However, the fundamental remarks of this approach are summarized as Supplementary Data.

The main steps for the application of this method in QSAR can be briefly summarized as follows: (1) draw the molecular pseudographs for each molecule of the data set, using the software drawing mode. This procedure is carried out by a selection of the active atomic symbols belonging to different groups in the periodic table; (2) use appropriate atom weights in order to differentiate the molecular atoms. In the present report, we characterized each atomic nucleus with the following parameters: atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), van der Waals atomic volume (V) and the atomic electronegativity in Pauling scale (G).[30] The values of these atomic labels are shown in Table 1 of Supplementary Data.[30–33] This weighting scheme was previously proposed for the calculation of the DRAGON descriptors;[31–33] (3) compute the total and local (atom-type or hybrid group) linear indices. They can be carried out in the software calculation mode, from which one can select the atomic properties and the family descriptor before calculating the molecular indices. This software generates a table in which the rows correspond to the compounds and columns correspond to the total and local linear indices or any other family of molecular descriptors implemented in this program; (4) find a QSAR equation by using statistical or artificial intelligent techniques, and so on. That is to say, one can find a quantitative multilinear regression analysis (MRA), neural networks (NN), linear discrimination analysis (LDA), and relation between **A** (activity) values and the linear indices having, for instance, the following appearance: $\mathbf{A} = a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + \cdots + a_k f_k(x) + c$, where $f_k(x)$ [or $f_{kL}(x)$] is the kth total [or local] linear indices, and the $\mathbf{a}_k$'s are the coefficients obtained by the linear regression analysis; (5) test the robustness and predictive power of the QSAR equation by using internal and external cross-validation techniques.

The atom-based *TOMOCOMD-CARDD* molecular descriptors computed in this study were the following: (i) kth ($k = 15$) total (global) non-stochastic linear indices not considering and considering H-atoms in the molecule [${}^{w}f_k(x)$ and ${}^{w}f_k{}^{H}(x)$, respectively], (ii) kth ($k = 15$) local (atom-type = heteroatoms: S, N, O) linear indices not considering and considering H atoms in the molecule, ${}^{w}f_{kL}(x_E)$ and ${}^{w}f_{kL}{}^{H}(x_E)$, correspondingly. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors, (iii) kth ($k = 15$) local (atom-type = H atoms bonding to heteroatoms: S, N, O) linear indices considering H atoms in the molecule. These local descriptors are putative H-bonding donors (hydrogen bonding capacity), lipophilicity, and so on, and (iv) The kth total [${}^{ws}f_k(x)$ and ${}^{ws}f_k{}^{H}(x)$] and atom-type [${}^{ws}f_k(x_E)$, ${}^{ws}f_k{}^{H}(x_E)$ and ${}^{ws}f_k{}^{H}(x_{E\text{-}H})$] stochastic linear indices were also computed. Here, we used the symbols ${}^{w}f_k(x)$ and ${}^{ws}f_k(x)$ for non-stochastic and stochastic linear indices, respectively, where the superscript w expresses the atomic property (atomic label) used to differential each atom in the molecule and computing the molecular descriptors, namely M, P, K, V, G,[30] and $f_k(x)$ means kth linear indices.

The names of tyrosinase inhibitor compounds in the database together with their experimental data taken from the literature are also shown in Table 2 of Supplementary Data. The molecular structures are given as Table 3 of Supplementary Data. This data set can be considered as a helpful tool for all the researchers in this field.

The chemicals in the database were divided into training and test sets with 478 and 180 compounds, respectively. The training set was used to develop the discriminant functions, and these were obtained by using the forward stepwise Linear Discriminant Analysis (LDA) as implemented in the statistic package STATISTICA.[34] The kth ($k \leqslant 15$) total and atom-type non-stochastic and stochastic linear indices were used as independent variables.

In this sense, there were obtained 12 LDA-based QSAR models. The first six models used the non-stochastic total and local linear indices (Eqs. 1–6) and the last six, stochastic molecular descriptors (Eqs. 7–12). The equations of the models are given in Table 1.

On one hand, the first five LDA models in both sets were obtained using each of the five atomic properties used as

Table 1. Discriminant models obtained with total and local non-stochastic and stochastic linear indices used in this study

LDA-based QSAR models obtained using non-stochastic linear indices

$$\textbf{Class} = -0.135 - 1.077\times10^{-3}\ {}^{M}f_3^{H}(x) + 9.710\times10^{-4}\ {}^{M}f_4(x) - 6.199\times10^{-8}\ {}^{M}f_{12}(x) + 7.719\times10^{-10}\ {}^{M}f_{15}(x) - 2.899\times10^{-2}\ {}^{M}f_{0L}{}^{H}(x_E) - 2.250\times10^{-10}\ {}^{M}f_{15L}(x_E) - 4.857\ {}^{M}f_{0L}{}^{H}(x_{E\text{-}H}) + 0456\ {}^{M}f_{1L}{}^{H}(x_{E\text{-}H}) - 1.715\ {}^{M}f_{3L}{}^{H}(x_{E\text{-}H}) \quad (1)$$

$$\textbf{Class} = 0.357 - 3.076\times10^{-2}\ {}^{V}f_2^{H}(x) + 1.400\times10^{-2}\ {}^{V}f_1(x) - 5.257\times10^{-5}\ {}^{V}f_6(x) - 1.895\times10^{-7}\ {}^{V}f_{11}(x) + 4.891\times10^{-10}\ {}^{V}f_{15}(x) - 5.797\times10^{-2}\ {}^{V}f_{0L}{}^{H}(x_E) + 3.446\times10^{-10}\ {}^{V}f_{15L}(x_E) 0.658\ {}^{V}f_{0L}{}^{H}(x_{E\text{-}H}) - 0.189\ {}^{V}f_{1L}{}^{H}(x_{E\text{-}H}) - 3.317\times10^{-2}\ {}^{V}f_{2L}{}^{H}(x_{E\text{-}H}) \quad (2)$$

$$\textbf{Class} = -6.428\times10^{-2} - 3.781\times10^{-4}\ {}^{P}f_6^{H}(x) + 5.920\times10^{-2}\ {}^{P}f_3(x) - 1.062\times10^{-2}\ {}^{P}f_4(x) - 0.498\ {}^{P}f_{0L}{}^{H}(x_E) - 0.143\ {}^{P}f_{3L}{}^{H}(x_E) + 1.589\times10^{-7}\ {}^{P}f_{13L}{}^{H}(x_E) + 0.139\ {}^{P}f_{3L}(x_E) - 2.188\times10^{-6}\ {}^{P}f_{11L}(x_E) + 3.230\ {}^{P}f_{0L}{}^{H}(x_{E\text{-}H}) - 0.158\ {}^{P}f_{3L}{}^{H}(x_{E\text{-}H}) \quad (3)$$

$$\textbf{Class} = -0.914 + 3.594\times10^{-4K}f_6(x) - 1.104\times10^{-7}\ {}^{K}f_{14}(x) + 2.619\times10^{-8}\ {}^{K}f_{15}(x) - 1.320\times10^{-2}\ {}^{K}f_{4L}{}^{H}(x_E) + 1.506\times10^{-2}\ {}^{K}f_{3L}(x_E) + 9.713\times10^{-4}\ {}^{K}f_{6L}(x_E) - 2.808\ {}^{K}f_{0L}{}^{H}(x_{E\text{-}H}) + 2.370\ {}^{K}f_{1L}{}^{H}(x_{E\text{-}H}) \quad (4)$$

$$\textbf{Class} = -0.929 + 2.203\times10^{-4}\ {}^{G}f_5^{H}(x) - 2.819\times10^{-6}\ {}^{G}f_9(x) + 0.486\ {}^{G}f_{1L}{}^{H}(x_E) - 0.140\ {}^{G}f_{3L}{}^{H}(x_E) - 1.753\times10^{-3}\ {}^{G}f_{6L}{}^{H}(x_E) + 8.720\times10^{-4}\ {}^{G}f_{7L}{}^{H}(x_E) + 2.226\times10^{-2}\ {}^{G}f_{4L}(x_E) + 3.753\times10^{-3}\ {}^{G}f_{5L}(x_E) - 1.481\times10^{-4}\ {}^{G}f_{8L}(x_E) - 3.417\ {}^{G}f_{0L}{}^{H}(x_{E\text{-}H}) + 2.557\ {}^{G}f_{1L}{}^{H}(x_{E\text{-}H}) \quad (5)$$

$$\textbf{Class} = 0.260 - 1.991\times10^{-3}\ {}^{V}f_2^{H}(x) + 0.115^{K}f_1(x) - 3.052\times10^{-6}\ {}^{G}f_9(x) - 6.243\times10^{-2}\ {}^{V}f_{0L}{}^{H}(x_E) - 0.156\ {}^{P}f_{3L}{}^{H}(x_E) + 1.103\times10^{-7}\ {}^{M}f_{13L}{}^{H}(x_E) + 0.158\ {}^{P}f_{3L}(x_E) + 2.251\times10^{-3}\ {}^{G}f_{5L}(x_E) - 7.716\times10^{-5}\ {}^{G}f_{8L}(x_E) + 0.225\ {}^{M}f_{1L}{}^{H}(x_{E\text{-}H}) - 0.152\ {}^{V}f_{1L}{}^{H}(x_{E\text{-}H}) \quad (6)$$

LDA-based QSAR models obtained using stochastic linear indices

$$\textbf{Class} = 0.344 - 0.198\ {}^{M}f_{1L}{}^{H},(x_E) + 6.805\times10^{-2}\ {}^{M}f_{5L}{}^{H}(x_E) + 0.587\ {}^{M}f_{6L}{}^{H}(x_E) - 0.358\ {}^{M}f_{8L}{}^{H}(x_E) + 0.132\ {}^{M}f_{1L}(x_E) - 0.219\ {}^{M}f_{4L}(x_E) + 0.310\ {}^{M}f_{1L}{}^{H}(x_{E\text{-}H}) + 0.128\ {}^{M}f_{5L}{}^{H}(x_{E\text{-}H}) - 0.554\ {}^{M}f_{13L}{}^{H}(x_{E\text{-}H}) \quad (7)$$

$$\textbf{Class} = 0.272 + 0.209\ {}^{V}f_{6L}{}^{H}(x_E) - 0.453\ {}^{V}f_{2L}(x_E) + 0.438\ {}^{V}f_{4L}(x_E) - 8.030\times10^{-2}\ {}^{V}f_{5L}(x_E) - 1.173\ {}^{V}f_{12L}(x_E) + 1.580\ {}^{V}f_{14L}(x_E) + 1.059\ {}^{V}f_{0L}{}^{H}(x_{E\text{-}H}) - 0.298\ {}^{V}f_{1L}{}^{H}(x_{E\text{-}H}) + 0.956\ {}^{V}f_{9L}{}^{H}(x_{E\text{-}H}) + 0.884\ {}^{V}f_{12L}{}^{H}(x_{E\text{-}H}) - 1.035\ {}^{V}f_{13L}{}^{H}(x_{E\text{-}H}) - 0.917\ {}^{V}f_{14L}{}^{H}(x_{E\text{-}H}) \quad (8)$$

$$\textbf{Class} = -0.631 - 0.785^{P}f_0^{H}(x) + 0.308^{P}f_4^{H}(x) + 0.404^{P}f_5^{H}(x) - 0.214\ {}^{P}f_{15}{}^{H}(x) + 0.260\ {}^{P}f_1(x) - 0.889\ {}^{P}f_{1L}{}^{H}(x_E) - 0.843\ {}^{P}f_{2L}{}^{H}(x_E) + 1.648\ {}^{P}f_{14L}{}^{H}(x_E) - 6.891^{P}f_{5L}{}^{H}(x_{E\text{-}H}) + 12.195\ {}^{P}f_{7L}{}^{H}(x_{E\text{-}H}) - 5.203\ {}^{P}f_{15L}{}^{H}(x_{E\text{-}H}) \quad (9)$$

$$\textbf{Class} = 0.202 + 1.253\ {}^{K}f_{2L}{}^{H}(x_E) + 53.854\ {}^{K}f_{13L}{}^{H}(x_E) - 53.804\ {}^{K}f_{15L}{}^{H}(x_E) + 1.454\ {}^{K}f_{2L}(x_E) - 2.117\ {}^{K}f_{6L}(x_E) - 10.146\ {}^{K}f_{9L}(x_E) + 9.247\ {}^{K}f_{15L}(x_E) - 7.244\ {}^{K}f_{0L}{}^{H}(x_{E\text{-}H}) + 2.376\ {}^{K}f_{1L}{}^{H}(x_{E\text{-}H}) + 4.160\ {}^{K}f_{4L}{}^{H}(x_{E\text{-}H}) \quad (10)$$

$$\textbf{Class} = -3.556\times l10^{-2} + 2.022\ {}^{G}f_{2L}{}^{H}(x_E) + 37.249\ {}^{G}f_{13L}{}^{H}(x_E) - 37.959\ {}^{G}f_{15L}{}^{H}(x_E) - 2.272\ {}^{G}f_{6L}(x_E) - 3.588\ {}^{G}f_{7L}(x_E) + 4.411\ {}^{G}f_{13L}(x_E) - 2.798\ {}^{G}f_{0L}{}^{H}(x_{E\text{-}H}) + 2.073\ {}^{G}f_{1L}{}^{H}(x_{E\text{-}H}) \quad (11)$$

$$\textbf{Class} = 0.175 + 0.311\ {}^{V}f_{6L}{}^{H}(x_E) + 32.906\ {}^{G}f_{13L}{}^{H}(x_E) - 31.996\ {}^{G}f_{15L}{}^{H}(x_E) - 0.294\ {}^{V}f_{2L}(x_E) - 1.103\ {}^{G}f_{5L}(x_E) + 0.474\ {}^{M}f_{1L}{}^{H}(x_{E\text{-}H}) - 0.122\ {}^{V}f_{1L}{}^{H}(x_{E\text{-}H}) + 2.268\ {}^{P}f_{7L}{}^{H}(x_{E\text{-}H}) - 0.254\ {}^{M}f_{13L}{}^{H}\ (x_{E\text{-}H}) - 0.257\ {}^{V}f_{13L}{}^{H}(x_{E\text{-}H}) \quad (12)$$

atomic weights (atomic labels) proposed above. On the other, the sixth model in both sets results from combining all the proposed weighting schemes.

The Wilks' λ parameter (U-statistic), square Mahalanobis distance (D^2), and Fisher ratio (F) for the training set are shown in Table 2. These statistical parameters together with the linear discriminant canonical statistics canonical regression coefficient (R_{can}) and chi-squared (χ^2) measure the quality of the determined models. The equations were shown to be statistically significant at p-level ($p < 0.0001$).

As it can be observed in Table 2, the fitted models with the combination of the weighted schemes exhibit the best results (Eqs. **6** and **12**, respectively). These best two models correctly classified the 93.51% and 92.46% (accuracy) of the training set. The equations showed high Matthews correlation coefficients (***C***) of 0.86 and 0.84. Table 2 also depicts the values of specificity, sensitivity, and false positive rate (also known as 'false alarm rate'), statistical parameters much used in QSAR studies.[35]

Although the statistical parameters had a good behavior, it is not enough to assure the predictive power of the models. For that reason, we carried out an external validation process using a test set[36,37] and the results are given in Table 3. In this sense, the TOMOCOMD-CARRD models (Eqs. **6** and **12**) show globally good classifications of 91.67% and 89.44%, respectively, in the prediction set. Furthermore, a high value of ***C*** can be observed in Eqs. **6** and **12** (see Table 3).

The classification of cases was performed by means of the posterior classification probabilities. By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ or as inactive otherwise. P(Active) and P(Inactive) are the probabilities with which the equations classify a compound as active and inactive, respectively. The classification results (including the canonical scores) for the database (active and inactive ones) with the models **6** and **12** are given as Tables 4–11 of Supplementary Data. In addition, we provide a plot with the $\Delta P\%$ for the actives and inactives using the non-stochastic and stochastic linear indices (Figs. 1 and 2).

Table 2. Prediction performances and statistical parameters for LDA-based QSAR models in the training set

Models[a]	Matthews corr. coefficient (C)	Accuracy 'Q_{Total}' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False positive rate (%)	Wilks' λ	D^2	F	Chi-Sqr (χ2)	Canonical R (R_{can})[b]
LDA-based QSAR models obtained using non-stochastic linear indices										
Eq. 1 (9)	0.80	90.59	86.3	89.6	8.8	0.49	4.46	55.0	340.3	0.72
Eq. 2 (10)	0.76	88.49	83.3	87.4	10.9	0.48	4.59	50.8	346.9	0.72
Eq. 3 (10)	0.79	89.75	84.5	89.6	10.2	0.49	4.47	49.5	340.3	0.72
Eq. 4 (8)	0.79	89.96	85.0	89.6	9.8	0.47	4.70	65.3	353.5	0.73
Eq. 5 (11)	0.81	91.00	86.5	90.7	8.8	0.45	5.12	51.5	374.1	0.74
Eq. 6 (11)	**0.86**	**93.51**	**90**	**93.4**	**6.4**	**0.43**	**5.68**	**57.1**	**401.7**	**0.76**
LDA-based QSAR models obtained using stochastic linear indices										
Eq. 7 (9)	0.76	88.70	83.4	88.0	10.9	0.47	4.77	58.8	356.8	0.73
Eq. 8 (12)	0.79	90.17	85.4	89.6	9.5	0.48	4.60	42.3	347.1	0.72
Eq. 9 (11)	0.77	88.91	83.9	88.0	10.5	0.55	3.48	34.9	282.9	0.67
Eq. 10 (10)	0.75	88.08	82.5	87.4	11.5	0.48	4.52	50.1	343.1	0.72
Eq. 11 (8)	0.72	86.61	80.5	87.8	12.9	0.52	3.86	53.7	306.8	0.69
Eq. 12 (10)	**0.84**	**92.46**	**88.5**	**92.4**	**7.5**	**0.40**	**6.33**	**70.1**	**431.8**	**0.77**

[a] The quantity of variables of the models is shown in parentheses.
[b] Canonical correlation coefficient obtained from the linear discriminant **canonical** analysis.

Table 3. Prediction performances for LDA-based QSAR models in the test set

Models	Matthews corr. coefficient (C)	Accuracy 'Q_{Total}' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False positive rate (%)
LDA-based QSAR models obtained using non-stochastic linear indices					
Eq. 1	0.64	83.33	74.63	79.37	14.53
Eq. 2	0.65	83.33	72.00	85.71	17.95
Eq. 3	0.73	86.67	75.32	92.06	16.24
Eq. 4	0.71	86.11	75.68	88.89	15.38
Eq. 5	0.77	88.89	78.67	93.65	13.68
Eq. 6	**0.82**	**91.67**	**86.36**	**90.48**	**7.69**
LDA-based QSAR models obtained using stochastic linear indices					
Eq. 7	0.70	85.56	74.67	88.89	16.24
Eq. 8	0.82	91.67	90.00	85.71	5.13
Eq. 9	0.71	86.11	76.39	87.30	14.53
Eq. 10	0.80	90.56	83.82	90.48	9.40
Eq. 11	0.76	88.89	82.09	87.30	10.26
Eq. 12	**0.77**	**89.44**	**82.35**	**88.89**	**10.26**

Table 4. Results of ligand-based in silico screening and tyrosinase inhibitory activities of new cycloartane compounds

Compound*	$\Delta P\%$[a]	Scores[a]	$\Delta P\%$[b]	Scores[b]	$\Delta P\%$[c]	Scores[c]	$\Delta P\%$[d]	Scores[d]	$\Delta P\%$[e]	Scores[e]	$\Delta P\%$[f]	Scores[f]	$IC_{50} \pm SEM$[g] (μM)
C1	**99.79**	**3.26**	**99.95**	**−3.92**	**96.97**	**2.19**	**99.98**	**−4.32**	**99.43**	**2.71**	**99.97**	**3.74**	102.4 ± 0.3
	99.80	*3.28*	*99.97*	*4.29*	*98.28*	*2.74*	*99.66*	*3.02*	*99.55*	*3.19*	*99.98*	*3.67*	
C2	**99.94**	**3.85**	**99.98**	**−4.37**	**97.95**	**2.37**	**99.99**	**−4.68**	**99.81**	**3.19**	**99.99**	**4.10**	95.3 ± 0.2
	99.87	*3.46*	*99.99*	*4.75*	*99.02*	*3.05*	*99.84*	*3.37*	*99.78*	*3.56*	*99.99*	*4.08*	
C3	**99.95**	**3.96**	**99.99**	**−4.67**	**96.46**	**2.11**	**99.98**	**−4.47**	**99.40**	**2.69**	**99.99**	**4.34**	48.92 ± 0.08
	99.74	*3.16*	*99.99*	*5.01*	*99.21*	*3.16*	*99.80*	*3.26*	*99.68*	*3.36*	*99.99*	*3.87*	
C4	**99.87**	**3.51**	**99.98**	**−4.23**	**92.83**	**1.77**	**99.94**	**−3.84**	**96.94**	**1.96**	**99.97**	**3.79**	13.95 ± 0.6
	98.61	*2.38*	*99.96*	*4.08*	*97.98*	*2.66*	*98.64*	*2.36*	*97.74*	*2.36*	*99.86*	*2.97*	
C5	**99.92**	**3.72**	**99.98**	**−4.28**	**98.70**	**2.59**	**99.98**	**−4.44**	**99.55**	**2.82**	**99.98**	**3.93**	54.6 ± 0.3
	99.61	*2.96*	*99.97*	*4.22*	*97.77*	*2.60*	*99.67*	*3.03*	*99.01*	*2.79*	*99.95*	*3.41*	
C6	**99.92**	**3.71**	**99.97**	**−4.06**	**99.85**	**3.61**	**100.00**	**−5.09**	**99.95**	**3.75**	**99.99**	**4.34**	85.01 ± 0.08
	99.74	*3.15*	*99.91*	*3.73*	*98.43*	*2.79*	*99.51*	*2.85*	*99.81*	*3.63*	*99.98*	*3.79*	

*The molecular structures of these chemicals are shown in Figure 3. [a,b,c,d,e,f]$\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ as well as canonical scores of each compound in this set: (i) Above in **bold**, classification of each compound using the obtained models with non-stochastic linear indices in the following order: Eqs. 1, 2, 3, 4, 5, and 6; and (ii) Below in *italic*; classification of each compound using the obtained models with stochastic linear indices in the following order Eqs. 7, 8, 9, 10, 11, and 12. [g]IC_{50} are the 50% inhibitory concentrations against the enzyme tyrosinase and SEM is the standard error of the mean.

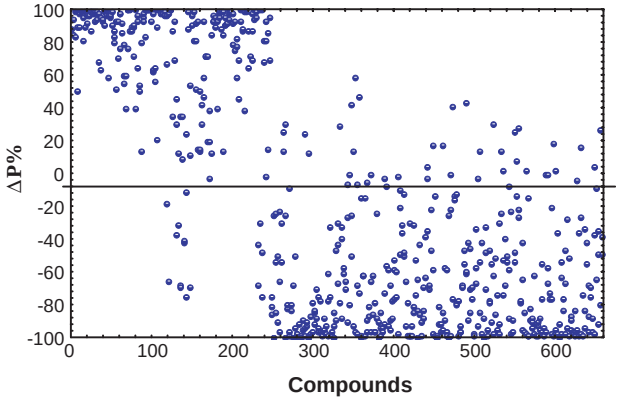

Figure 1. Plot of the $\Delta P\%$ from Eq. 6 (using non-stochastic linear indices) for each compound in the training and test sets. Compounds **1**–**183** and **184**–**246** are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247–541 and 542–658 are inactive (non-inhibitors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

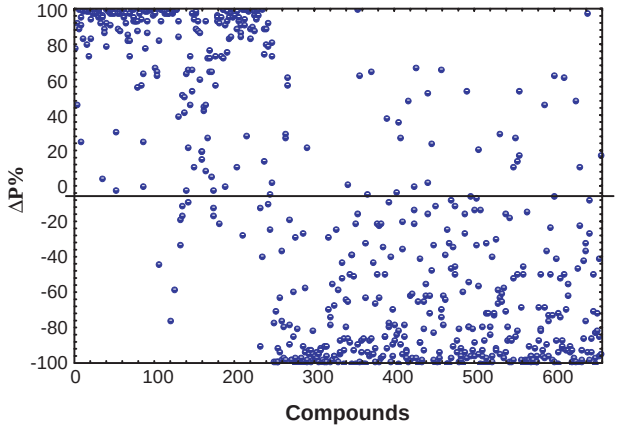

Figure 2. Plot of the $\Delta P\%$ from Eq. 12 (using stochastic linear indices) for each compound in the training and test sets. Compounds **1**–**183** and **184**–**246** are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247–541 and 542–658 are inactive (non-inhibitors of tyrosinase) in both training and test sets, correspondingly.

On the other hand, the techniques for assaying new compounds on virtual screening can predict, ahead of time, the likely result of a many-year biological-property study. Taking this consideration into account, we evaluated 75 compounds using the models of TOMOCOMD-CARDD approach. The names and structures from these chemicals are given in Tables 12 and 13, respectively, of Supplementary Data. The selected compounds are reported in the literature as active/inactive compounds (see the last column of Table 12, Ref. of Supplementary Data). Together with these, we show the results of posterior classification probabilities (and canonical scores) depicted in Table 14 of Supplementary Data. The obtained models, Eqs. **6** and **12**, showed an overall accuracy of 90.66% and 85.33%, correspondingly. The results validate the models for use in the ligand-based virtual screening.[38]

The mayor impact on drug discovery is always the identification of novel lead compounds. In this sense, another of our research teams has been focused on searching for new tyrosinase inhibitors based on trial-and-error methods.[39–45] Besides, in this case we used the LDA models developed with TOMOCOMD-CARDD molecular descriptors in the virtual screening of a cycloartane family isolated from herbal plants.

As can be seen in Table 4, all the discriminant functions classified as actives (tyrosinase inhibitors) the new six compounds. To corroborate the predictive ability of our QSAR models, the chemicals were isolated and an in vitro assay was carried out.[46]

The experimental information of in vitro pharmacology test concerning these compounds has been described in a previous report,[46] and fundamental remarks were summarized as Supplementary Data (see page 58). As it can be observed, the theoretical results obtained are in correspondence with the evaluated activity (see Table 4). Also the $\Delta P\%$ values from each obtained model and the canonical scores are reported in this table.

All the chemical structures had activity and one of them **C4** (IC_{50} = 13.95 μM) showed activity higher than that

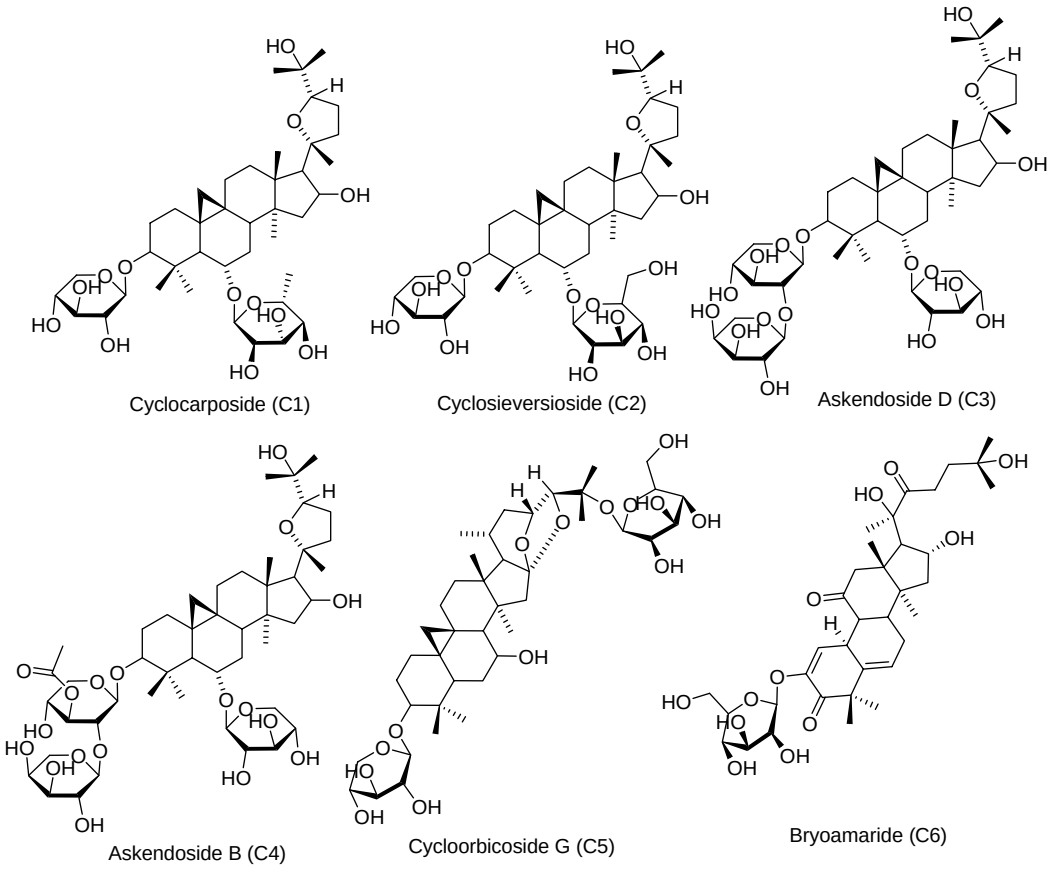

Figure 3. Molecular structure of the cycloartane compounds.

of kojic acid (IC_{50} = 16.67 μM), the drug used as tyrosinase inhibitor reference. The remaining compounds, **C1** (IC_{50} = 102.39 μM), **C2** (IC_{50} = 92.25 μM), **C3** (IC_{50} = 48.92 μM), **C5** (IC_{50} = 54.64 μM), and **C6** (IC_{50} = 85.01 μM), exhibited a mild effect on inhibitory activity against the enzyme. The structures of the compounds are depicted in Figure 3.

The research on tyrosinase inhibitors has become an important area to study the role in hyperpigmentation and melanogenesis disorders.[5] In this case, a new approach for the rational selection of new active compounds against the enzyme has been described. These models based on TOMOCOMD-CARDD descriptors and pattern recognition techniques can identify new chemical structures with tyrosinase activity. A new method is proposed for increasing the speed of discovering new lead-like compounds, as a suitable alternative to the screening and in vitro assay. This was proved experimentally through the isolation and characterization of six new compounds with the corresponding tyrosinase inhibitory assay. In this sense, it can be said that the accumulation of this kind of knowledge will provide a useful clue for the design of effective and selective tyrosinase inhibitors.[47]

Acknowledgments

One of the authors (M.-P.Y) thanks the program 'Estades Temporals per a Investigadors Convidats' for a fellowship to work at Valencia University. M. T. H. K. is the recipient of a grant from MCBN-UNESCO (Grant No. 1056), and fellowships from CIB (Italy) and Associasione Veneta per la Lotta alla Talassemia (AVTL, Italy). F. T. acknowledges financial support from the Spanish MEC DGI (Project No. CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173).

Supplementary data

The complete list of compounds used in training and prediction sets, as well as their structures, posterior classification, and scores according to LDA-based QSAR models, and in vitro pharmacology test of the isolated chemicals, are available in the supplementary data associated with this article, which can be found, in the online version, at doi:10.1016/j.bmcl.2005.09.085.

References and notes

1. Lerch, K. In *Metal Ions in Biological Systems*; Siegel, H., Ed.; Marcel Dekker: Zürich, 1981; p 143.
2. Robb, D. A. In *Biochemistry of Fruits and Vegetables*; Friend, J., Rhodes, M., Eds.; Academic Press: London, 1981; pp 181–192.
3. Solomon, E. I.; Sundaram, U. M.; Machonkin, T. E. *Chem. Rev.* **1996**, *96*, 2563.
4. Jones, K.; Hughes, J.; Hong, M.; Jia, Q.; Orndorff, S. *Pigment Cell Res.* **2002**, *15*, 335.

5. No, J. K.; Soung, D. Y.; Kim, Y. J.; Shim, K. H.; Jun, Y. S.; Rhee, S. H.; Yokozawa, T.; Chung, H. Y. *Life Sci.* **1999**, *65*, 241.
6. Garcia, A.; Fulrton, J. E. *Dermatol. Surg.* **1996**, *22*, 443.
7. Kojima, S.; Yamaguch, H.; Morita, K.; Ueno, Y.; Paolo, R. *Biol. Pharm. Bull.* **1995**, *18*, 1076.
8. Cabanes, J.; Chazaarra, S.; Garcia-Camona, F. *J. Pharm. Pharmacol.* **1994**, *46*, 982.
9. Verallo-Rowell, V. M.; Verallo, V.; Graupe, K.; Lopez-villafuerte, L.; Garcia-Lopez, M. *Acta. Derm-Vereol.* **1989**, *143*, 58.
10. Takiwake, H.; Shirai, S.; Kohono, H.; Soh, H.; Arase, S. *J. Invest. Dermatol.* **1994**, *103*, 642.
11. Kimbrough-Green, C. K. *Arch. Dermatol.* **1994**, *130*, 727.
12. Maeda, K.; Fukuda, M.; Griffith, C. E.; Finkel, L. J.; Hamilton, T. A.; Bulengo-Ransby, S. M. *J. Pharmacol. Exp. Ther.* **1996**, *276*, 765.
13. Kubo, I. In *Phytochemicals for Pest Control*; Hedin, P., Hollingworth, R., Masler, E., Miyamoto, J., Thompson, D., Eds.; ACS Symposium Series 658; American Chemical Society: Washington, DC, 1997; pp 310–326.
14. Kubo, I.; Kinst-Hori, I. *J. Agric. Food Chem.* **2000**, *48*, 1393.
15. Kubo, I.; Kinst-Hori, I. *J. Agric. Food Chem.* **1998**, *46*, 1268.
16. Kubo, I.; Kinst-Hori, I. *J. Agric. Food Chem.* **1998**, *46*, 5338.
17. Browne, L. J.; Taylor, L. L. DDW (*Drug Discovery Word*), Fall **2002**, 72.
18. Marrero-Ponce, Y.; Romero, V. TOMOCOMD software, Central University of Las Villas, 2002. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future, a professional version will be obtained upon request to Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es.
19. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687.
20. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010.
21. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R.; García Sánchez, R. N. *Bioorg. Med. Chem. Lett.* **2005**, *17*, 3838.
22. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. *J. Pharm. Pharmaceut. Sci.* **2004**, *7*, 186.
23. Marrero-Ponce, Y.; Castillo Garit, J. A.; Nodarse, D. *Bioorg. Med. Chem.* **2005**, *13*, 3397.
24. Marrero-Ponce, Y.; Iyarreta-Veitía, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Ávila, P. E.; Kirchgatter, K. *J. Chem. Inf. Model* **2005**, *45*, 1082.
25. Negwer, M. *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag: Berlin, 1987.
26. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. *Molecules* **2004**, *9*, 1100.
27. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; Sanchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 1005.
28. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón-Peréz, M.; García Sánchez, R. *Bioorg. Med. Chem.* **2005**, *13*, 1293.
29. Marrero-Ponce, Y.; Medina-Marrero, R.; Martinez, Y.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. A. *J. Mol. Mod.* DOI1007/s00894-005-002-8.
30. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, UK, 1986.
31. Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: New York, 1939, pp 2–60.
32. Todeschini, R.; Gramatica, P. *Perspect. Drug Disc. Des.* **1998**, *9–11*, 355–380.
33. Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.
34. STATISTICA (data analysis software system), vs 6.0; StatSoft Inc., 2001.;
35. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. *Bioinformatics* **2000**, *16*, 412.
36. Wold, S.; Erikson, L. In *Chemometric Methods in Molecular*; van de Waterbeemed, H., Ed.; VCH Publishers: New York, 1995; pp 309–318.
37. Golbraikh, A.; Tropsha, J. *Mol. Graph. Mod.* **2002**, *20*, 269.
38. Gálvez, J.; García, R.; Salabert, M. T.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520.
39. Khan, S. B.; Azhar-Ul-Haq; Afza, N.; Malik, A.; Khan, M. T. H.; Shah, M. R.; Choudhary, M. I. *Chem. Pharm. Bull.* **2005**, *53*, 86.
40. Choudhary, M. I.; Sultan, S.; Khan, M. T. H.; Yasin, A.; Shaheen, F.; Atta-ur-Rahman *Nat. Prod. Res.* **2004**, *18*, 529.
41. Ahmad, V. U.; Ullah, F.; Hussain, J.; Farooq, U.; Zubair, M.; Khan, M. T. H.; Choudhary, M. I. *Chem. Pharm. Bull.* **2004**, *52*, 1458.
42. Khan, K. M.; Maharvi, G. M.; Abbaskhan, A.; Hayat, S.; Khan, M. T. H.; Makhmoor, T.; Choudhary, M. I.; Shaheen, F.; Atta-ur-Rahman *Helv. Chim. Acta* **2003**, *86*, 457.
43. Khan, M. T. H.; Choudhary, M. I.; Khan, K. M.; Ranib, M.; Atta-ur-Rahman *Bioorg. Med. Chem.* **2005**, *13*, 3385.
44. Choudhary, M. I.; Musharraf, S. G.; Khan, M. T. H.; Abdelrahman, D.; Parvez, M.; Shaheen, F.; Atta-ur-Rahman *Helv. Chim. Acta* **2003**, *86*, 3450.
45. Şabudak, T.; Khan, M. T. H.; Choudhary, M. I.; Oksuz, S. *Nat. Prod. Res.* (in press).
46. Hearing, V. J., Ed.; *Methods in Enzymology*, Academic Press: New York, 1987; Vol. 142, pp 154.
47. Kubo, I. *J. Food Chem.* **2003**, *81*, 241.