12th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-12) 1-30 November 2008
http://www.usc.es/congresos/ecsoc/12/ECSOC12.htm & http://www.mdpi.org/ecsoc-12

[G0007]

Bond-Based 2D Quadratic Fingerprints in QSAR Studies. Virtual and *In Vitro* Tyrosinase Inhibitory Activity Elucidation

Yovani Marrero-Ponce,[1,2,3*] Gerardo M. Casañola-Martín,[1,4] Mahmud Tareq Hassan Khan,[5,6] Arjumand Ather,[7] Kalid M. Khan,[8] Vicente Romero Zaldivar,[9] Francisco Torrens,[2] Richard Rotondo,[10] and Facundo Pérez-Jiménez.[3]

[1]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.

[2]Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, Polígon la Coma s/n P. O. Box 22085, E-46071 Valencia, Spain.

[3]Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, València, Spain.

[4]Department of Biological Sciences, Faculty of Agricultural Sciences, University of Ciego de Avila, 69450, Ciego de Avila, Cuba.

[5]Pharmacology Research Lab., Faculty of Pharmaceutical Sciences, University of Science and Technology, Chittagong, Bangladesh.

[6]Department of Pharmacology, Institute of Medical Biology, University of Tromso, Tromso 9037, Norway.

[7]The Norwegian Structural Biology Centre (NorStruct), University of Tromso, Tromso 9037, Norway.

[8]HEJ Research Institute of Chemistry, University of Karachi, Pakistan.

[9]Faculty of Informatics, University of Cienfuegos, Cienfuegos 59430, Cuba.

[10]Advanced Medisyns, Inc. 601 Carlson Parkway, Suite 1050, Minnetonka, Minnesota 55305, USA.

[*]***Corresponding author.***

Fax: 53-42-281130 [or 53-42-281455] (Cuba) and 963543156 (València)

Phone: 53-42-281192 [or 53-42-281473] (Cuba) and 963543156 (València)

Cell: 610028990

***e*-mail**: ymarrero77@yahoo.es; yovani.marrero@uv.es; ymponce@gmail.com
or yovanimp@qf.uclv.edu.cu

URL: http://www.uv.es/yoma/

Abstract: In this report, we show the results of QSAR (quantitative structure-activity relationship) studies of tyrosinase inhibitory activity, by using the bond-based quadratic indices as molecular descriptors (MDs) and linear discriminant analysis (LDA), to generate discriminant functions to predict the anti-tyrosinase activity. The best two models (Eqs.**8** and **14**) of the total 12 QSAR models developed here show accuracies of 93.51% and 91.21%, as well as high Matthews correlation coefficients (***C***) of 0.86 and 0.82, respectively, in the training set. The validation external series depicts values of 90.00% and 89.44% for these best two equations **8** and **14**, correspondingly. Afterwards, a second external prediction data was used, to perform a virtual screening of compounds reported in the literature as active (tyrosinase inhibitors). In a final step, a series of lignans is analyzed by using the *in silico* developed models, and *in vitro* corroboration of the activity is carried out. An aspect of great importance to remark here, is that all compounds present greater inhibition values than Kojic Acid (standard tyrosinase inhibitor: IC_{50} =16.67μM). The current obtained results could be used as a method to increase the speed, in the *biosilico* discovery of leads for the treatment of skin disorders.

Keywords: *TOMOCOMD-CARDD* Software, Non-Stochastic and StochasticBond-Based Quadratic Indices, LDA-Based QSAR Model, Tyrosinase Inhibitor, Virtual Screening, Lignan.

Running head: *Bond-Based* TOMOCOMD-CARDD *MDs & Drug Discovery of Novel Tyrosinase Inhibitors*

INTRODUCTION

Tyrosinase or polyphenoloxidase (EC 1.14.18.1) is a bifunctional, copper-containing enzyme widely distributed on the phylogenetic tree. This enzyme uses molecular oxygen to catalyze the oxidation of monophenols to their corresponding *o*-diphenols (cresolase activity) as well as their subsequent oxidation to *o*-quinones (catecholase activity). The *o*-quinones thus generated polymerize to form melanin, through a series of subsequent enzymatic and non-enzymatic reactions [1-3]. This tyrosinase process is involved in abnormal accumulation of melanin pigments (hyperpigmentation). Therefore, tyrosinase inhibitors have been established as important constituents of cosmetic materials, as well as depigmenting agents for hyperpigmentation [4].

Tyrosinase may also play an important role in neuromelanin formation in the human brain, particulary in substantia nigra, and could be central to dopamine neurotoxicity, as well as contribute to the neurodegeneration associated with Parkinson's disease [5]. Melanoma specific anticarcinogenic activity is known to be linked with tyrosinase activity [6].

These broad applications and others have caused that tyrosinase inhibitors have been interested as molecular target of natural product and synthetic chemistry and, more recently, in other chemical fields such as computational and theoretical chemistry [7, 8]. In these areas, some QSAR studies from heterogeneous dataset of compounds have also been developed, finding models that permit predicting the anti-tyrosinase activity of organic-chemicals [9-13]. Such models could be used to increase the effectiveness of

HTS (High Throughput Screening) in several ways [14, 15], like the speeding up of lead identification or minimizing costs in drug discovery processes[16].

To this effect, our research group has applied extensively QSAR/QSPR investigations related to chemical, physicochemical and biological properties of different compounds and drugs [17-21], including analysis of nucleic acid-drug interactions [22, 23] and discovery of antimalarial compounds[24]. These studies and others have been made by using the 'in house' ***TOMOCOMD-CARDD*** software (acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign-***C***omputer ***A***ided '***R***ational' ***D***rug ***D***esign) [25], a novel computer-aided molecular design scheme.

Taking these considerations into account, we proposed here a new set of molecular descriptors, namely *non-stochastic and stochastic bond-based quadratic indices*; with the aim to prove the usefulness of these novel molecular descriptors, 1) their application to finding QSAR models in the prediction of tyrosinase inhibitory activity, 2) its use in a virtual screening of compounds and, as a final objective, 3) the virtual and experimental assays of a new series of lignans are shown. These new molecular parametrization and proposed algorithm could help to future successful identification of "real" or "virtual" organic-chemicals with such an activity.

MATERIALS AND METHODS

TOMOCOMD-CARDD Method

In an early publication, the theory of the bond, group and bond-type, as well as total non-stochastic and stochastic quadratic indices, computed from the *k*th non-stochastic and stochastic edge adjacency matrices, $\mathbf{E}^k$ and $\mathbf{ES}^k$, respectively; for small-to-medium sized organic compounds has been explained in some detail [26].

The total and local (group and bond-type) bond-based quadratic indices were calculated, by the interactive program for molecular design and bioinformatic research ***TOMOCOMD-CARDD*** [25]. The software was developed based on a user-friendly philosophy. Therefore, this graphics software shows a great efficiency of interaction with the user, without *prior* knowledge of programming skills (*e.g.* practicing pharmacist and organic chemist, teacher, university student, and so on). CARDD subprogram allows drawing the structures (drawing mode) and calculating 2D (topological), 3D-chiral (2.5D) and 3D (geometric and topographical) non-stochastic and stochastic molecular descriptors (calculation mode).

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

1. To draw of the molecular pseudographs for each molecule in the data set, by using the drawing mode.
2. To use appropriate weights in order to differentiate the molecular atoms. The weights used in this work are those previously proposed for the calculation of the DRAGON descriptors [27-29], i.e. atomic mass (M), atomic polarizability (P), atomic Mullinken electronegativity (K), van der Waals atomic volume (V) plus the atomic electronegativity in Pauling scale (G) [30]. The values of these atomic labels are shown in Table 1[27-30].
3. Computation of the non-stochastic and stochastic total and local bond-based quadratic indices can be carried out in the software calculation mode, where one can select the bond properties and the descriptor family before calculating the molecular indices. This software generates a table in which the rows correspond to the compounds, and

the columns correspond to the bond-based (both total and local) quadratic maps, or other MD family implemented in this program.

4. Development of a QSPR/QSAR equation by using several multivariate analytical techniques, for instance, linear discrimination analysis. Therefore, one can find a quantitative relationship between an activity **A** and the bond-based quadratic fingerprints having, for instance, the following appearance:

$$\mathbf{A} = a_0\boldsymbol{q}_0(w) + a_1\boldsymbol{q}_1(w) + a_2\boldsymbol{q}_2(w) + \ldots + a_k\boldsymbol{q}_k(w) + \mathrm{c} \qquad (\mathbf{1})$$

where **A** is the measured activity, $\boldsymbol{q}_k(w)$ are the k^{th} non-stochastic total bond-based quadratic indices, as well as the $\boldsymbol{a}_k$*'s* and c 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 both a test set and an external predicting set) validation techniques.

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

1) k^{th} (k = 15) total non-stochastic bond-based quadratic indices not considering and considering H-atoms in the molecular graph (G) [$\boldsymbol{q}_k(w)$ and $\boldsymbol{q}_k{}^{H}(w)$, respectively].

2) k^{th} (k = 15) total stochastic bond-based quadratic indices not considering and considering H-atoms in the molecular graph (G) [${}^{s}\boldsymbol{q}_k(w)$ and ${}^{s}\boldsymbol{q}_k{}^{H}(w)$, respectively].

3) k^{th} (k = 15) bond-type local (group = heteroatoms: S, N, O) non-stochastic quadratic indices not considering and considering H-atoms in the molecular graph (G) [$\boldsymbol{q}_{k\mathrm{L}}(w_E)$ and $\boldsymbol{q}_{k\mathrm{L}}{}^{H}(w_E)$, correspondingly]. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.

4) k^{th} (k = 15) bond-type local (group = heteroatoms: S, N, O) stochastic quadratic indices not considering and considering H-atoms in the molecular graph (G) [$^{s}\boldsymbol{q}_{k\mathrm{L}}(w_{\mathrm{E}})$, and $^{s}\boldsymbol{q}_{k\mathrm{L}}{}^{\mathrm{H}}(w_{\mathrm{E}})$, correspondingly]. These local descriptors are also putative molecular charge, dipole moment, and H-bonding acceptors.

Experimental Methods

Chemical Techniques

The isolation and structural characterization of the lignan compounds, their biological studies and cross references have been reported by one of our research teams [31].

In Vitro Assay

The tyrosinase inhibition assay was performed with Kojic Acid and L-Mimosine as standard inhibitors of the tyrosinase in a 96-well microplate format, by using a SpectraMax 340 micro-plate reader (Molecular Devices, CA, USA), according to the method developed by Hearing [32]. Briefly, the compounds first were screened for the *o*-diphenolase inhibitory activity of tyrosinase, by using L-DOPA as substrate. All the active inhibitors from the preliminary screening were subjected to IC_{50} studies. The compounds were dissolved in methanol to a concentration of 2.5%. Thirty units of mushroom tyrosinase (28 nM from Sigma Chemical Co., USA) were first preincubated with the test compounds in 50 nM Na-phosphate buffer (pH 6.8) for 10 min at 25 ^{o}C. Then the L-DOPA (0.5 mM) was added to the reaction mixture, and the enzymatic reaction was monitored by measuring the change in absorbance at 475 nm (at 37 ^{o}C), due to the formation of the DOPAchrome, for 10 min. The percentage of inhibition of the

enzyme was calculated as follows, by using a MS Excel®™ 2000 (Microsoft Corp., USA) computing sheet developed for this purpose:

Percent Inhibition = [B – S/B] x 100 (**2**)

Here the B and S are the absorbances for the blank and samples, respectively. After the screening of the compounds, median inhibitory concentrations (IC_{50}) were also calculated. All the studies have been carried out at least in triplicate and the result represents the mean ± SEM (standard error of the mean). Kojic Acid and L-Mimosine were used as standard inhibitors of the tyrosinase, and both of them were purchased from Sigma Chem. Co., USA.

RESULTS AND DISCUSSION

Assembling the Database

The main step to develop the QSAR models is to collect a database with compounds of several structural features, assuring an extrapolating power more accurate in the entire chemical space. A total of 658 compounds were selected with this purpose, as well as by considering these aspects mentioned above.

The subset of active chemicals in this data (246 tyrosinase inhibitors) was chosen from the literature, warranting sufficient structural diversity; it includes many representative tyrosinase inhibitor families such as: hydroxychalcones [33], kojic acid tripeptides [34], novel *N*-substituded *N*-nitrosohydroxylamines [35], vitamin B_6 compounds [36], steroids [37] and so on, which are illustrated in Figure 1, together with the reference tyrosinase inhibitors that are more used. The names of compounds in the dataset and their experimental values, taken from the literature, are given in Table 1 of

Supplementary Data. The molecular structures of all members of these families, and of the rest of tyrosinase inhibitors employed in this study, are depicted in Table 2 (Supplementary Data). The chosen active compounds present various types of inhibition, sources of the enzyme or substrates used to measure the activity, increasing the possibilities in the process of identification/selection of novel leads. Therefore, these data of tyrosinase inhibitors presented here could be successfully used for these scientists concerning the area of tyrosinase inhibitor researches.

Figure 1 comes about here (see end of the document)

In the case of inactive compounds we chose at random 412 drugs having a series of other pharmacological uses [38]. Here as in the active set we try to assure an adequate structural diversity.

Cluster Analysis Techniques to Perform Training and Test Sets

As we said above a crucial aspect is the structural diversity of these subsets. To verify this, we make use of different cluster analysis (CA) techniques, implemented in the STATISTICA software [39]. The dendrograms obtained through hierarchical CAs [40] of the active and inactive subsets are shown in Figures 2 and 3, respectively. We can visualize several groups in the output dendrograms of the performed *k*-NNCA, which permit us to prove the great quantity of structural families existing here.

Figure 2 and 3 comes about here (see end of the document)

Besides, the design of training and prediction series is carried out with the aim of finding the classification functions. To this effect partition of active and inactive sets is necessary to select the compounds in a 'rational' way. On one hand, a *k*-MCA I split the

active set (tyrosinase inhibitors) into 10 clusters. The data of inactive compounds was partitioned by a second *k*-MCA (*k*-MCA II) into 12 clusters. The used variables were the k^{th} non-stochastic bond-based quadratic indices. The analysis of variance for these k-MCAs are summarized in Table 2.

Table 2 comes about here (see end of the document)

Figure 4 comes about here (see end of the document)

In the diagram of Figure 4 are illustrated all the processes described above to perform the training (478 compounds) and prediction (180 compounds) sets. In the training series 183 compounds were actives, and 295, inactive substances. On the other hand, in the validation set 63 compounds were selected as tyrosinase inhibitors (actives), and the remaining 117, inactives (non-inhibitors of tyrosinase).

Finding QSAR Models

After the selection of the training series, the next step is to fit discriminant functions in order to classify the compounds as either active or inactive, in our case, we made use of the linear discriminant analysis (LDA),[41] implemented in the STATISTICA software [39], as statistical technique, due to its simplicity and extended use in drug design [21, 41-46]. In Table 3 are depicted all the obtained models, twelve in total. The first six models were performed using the non-stochastic bond-based quadratic indices (Eqs. **3-8**), and the last six, the stochastic molecular fingerprints (Eqs. **9-14**). The most common prediction values in the QSAR models are given for the training set in Table 4. Here we included Wilks' statistic (λ), the square of the Mahalanobis distance (D^2), and the Fisher

ratio (F), as well as always took into account p-level< 0.0001 for constructing the models.

Table 3 comes about here (see end of the document)

Table 4 comes about here (see end of the document)

On one hand, the first five LDA models, in both sets, were obtained by using each one of the five atomic properties used as atomic weight (atomic labels), proposed in ***TOMOCOMD-CARDD Method*** Section. On the other, the sixth model in both sets results from combining all the proposed weighting schemes.

The best two models were equations **8** and **14**, resulting of the combination of weighting schemes, for the non-stochastic and stochastic bond-based quadratic indices, correspondingly, which are shown in Table 4. These two equations had accuracies of 93.51% (***C*** = 0.86) and 91.21% (***C*** = 0.82). In Table 4 are also summarized the results of the statistical parameters, by using the non-stochastic and stochastic bond-based quadratic indices for the entire set of developed models, together with their respective values of accuracies, Matthews correlation coefficients (***C***), and some linear discriminant canonicals statistics, which measure the quality of the determined models. For instance, the good values of the canonical correlation coefficients of 0.85 and 0.72 for models **8** and **14**, respectively, are depicted in Table 4**.**

The molecular descriptors in the case of these best two models showed collinearity, after an analysis of the correlation between the variables (data not shown). For that reason we used the Randić's orthogonalization process, which has been described in detail in several publications [47-50], to eliminate interrelation between the variables.

The results of the orthogonalization processes, for the non-stochastic and stochastic bond-based quadratic indices, are showed in equations **15** and **16**, respectively.

$$\textbf{\textit{Class}} = -0.239 - 1.257^{1}O(^{P}\boldsymbol{q}_{1}{}^{H}(w)) + 1.733^{2}O(^{V}\boldsymbol{q}_{1}(w)) - 1.272^{3}O(^{K}\boldsymbol{q}_{3L}{}^{H}(w_E)) + 3.377^{4}O(^{G}\boldsymbol{q}_{0L}(w_E)) - 2.430^{5}O(^{V}\boldsymbol{q}_{0L}(w_E)) + 1.065^{6}O(^{M}\boldsymbol{q}_{5}(w)) - 2.027^{7}O(^{G}\boldsymbol{q}_{1L}(w_E)) - 1.452^{6}O(^{G}\boldsymbol{q}_{5}{}^{H}(w)) \quad (\mathbf{15})$$

N = 478 λ =0.41 D^2 = 5.99 F = 83.3 Canonical R = 0.77 $\chi 2$ = 417.1 Q_{Total}= 93.51% C = 0.86

$$\textbf{\textit{Class}} = -0.050 - 1.272\,{}^{1}O(^{P}\boldsymbol{q}_{2L}(w_E)) + 2.777^{2}O(^{K}\boldsymbol{q}_{0L}(w_E)) - 4.501^{3}O(^{G}\boldsymbol{q}_{3L}(w_E)) + 1.227^{4}O(^{G}\boldsymbol{q}_{14}(w)) - 3.773^{5}O(^{V}\boldsymbol{q}_{0}(w)) + 2.247^{6}O(^{P}\boldsymbol{q}_{0}(w)) - 0.881^{7}O(^{G}\boldsymbol{q}_{10L}{}^{H}(w_E)) + 1.265^{8}O(^{P}\boldsymbol{q}_{11L}{}^{H}(w_E)) + 0.862^{9}O(^{V}\boldsymbol{q}_{2L}{}^{H}(w_E)) - 5.689^{10}O(^{P}\boldsymbol{q}_{0L}{}^{H}(w_E)) - 6.091^{11}O(^{P}\boldsymbol{q}_{2L}{}^{H}(w_E)) \quad (\mathbf{16})$$

N = 478 λ = 0.44 D^2 = 5.31 F = 53.4 Canonical R = 0.75 $\chi 2$ = 383.7 Q_{Total}= 91.21% C = 0.82

Here, we used the symbols $^{m}O(\boldsymbol{q}_k(w))$, where the superscript m expresses the order of importance of the variable ($\boldsymbol{q}_k(w)$), after a preliminary forward stepwise analysis, and O means orthogonal. It is important to remark that all the statistical parameters are the same, whether we use either a set of non-orthogonal descriptors or the corresponding set of orthogonal indices.

First External Prediction Data

The external validation process is necessary to assess the real predictive power of any QSAR model [51, 52]. In this case, the predictive ability of the discriminant functions was evaluated through a test set. In Table 5 are shown the prediction performances in the test set for all the models; as can be seen, the best two models (Eqs. **8** and **14**) classify

correctly the 90.00% and 89.44%, respectively, in the prediction series. A high ***C*** value of 0.79 and 0.78 can be observed in the Eqs. **8** and **14**, correspondingly.

In Figures 5 and 6 we give a plot of the pictorial representations for the classification of all compounds, in both training and test sets from models **8** and **14**, respectively. These results encouraged us to use the models in a virtual screening study.

Table 5 comes about here (see end of the document)

Figure 5 comes about here (see end of the document)

Figure 6 comes about here (see end of the document)

All the posterior classification probabilities ($\Delta P\%$) and their respective canonical scores for all compounds (actives and inactive ones) in both training and prediction databases with all models are depicted in Tables 3-10 (Supplementary Data). By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})]\text{x}100$, or as inactive otherwise.

Second External Prediction Data

High Throughput Screening (HTS) has been proposed as one of the automated methods to resolve the arduous task of screening thousands of compounds, from large databases, due to expensive costs. In this sense another approach has emerged, the virtual HTS, to solve the major challenges in drug discovery: the massive cost of new drugs development [53]. Therefore, the *in silico* approaches and, among them, QSAR methods constitute a suitable alternative that can predict ahead of time the biological activity under study.

In order to test the ability of our QSAR models to detect new leads, we carried out a simulated virtual screening of tyrosinase inhibitors from the literature. To avoid the manipulation of a large database of chemicals, only 85 compounds, whose names are given in Table 6, were evaluated in the ligand-based virtual screening presented here. The collected organic-chemical are reported as active in the literature (see the last column of Table 6: **Ref**.). The molecular structures of these compounds are show in Table 11 (Supplementary Data).

Table 6 comes about here (see end of the document)

Figure 7 comes about here (see end of the document)

A cluster analysis (k-NNCA) was made to verify the molecular variability in the database of the virtual screening, aspect that can be visualized in the dendrogram of Figure 7. In the same Table 6, we also show the results of the classification of the compounds in the screening. Likewise, the results of ΔP% (including canonical scores) of the 85 compounds are depicted in Table 12 of Supplementary Data by using all the developed models.

In addition and for a better visualization of the results, we provide a plot with the ΔP% of these compounds by using Equations **8** and **14 (**Figs. 8 and 9, correspondingly). These external prediction data were classified by this *in silico* screening, with accuracies of 94.11% for both non-stochastic and stochastic molecular descriptors. The values of predictions were checked out from recent reports in the literature (see the last column of Table 6: **Ref**.).

Figure 8 comes about here (see end of the document)

The adequate results in this second external prediction series can be successfully applied in the find of higher promissory compounds with anti-tyrosinase activity, from a database of thousands or even million of chemicals. On the other hand, a database of commercially available compounds could be used to detect new bioactive compounds against enzyme tyrosinase, taking into account that those drugs selected from this ligand-based virtual screening could have well-established methods of synthesis and that their toxicological, pharmacodynamic and pharmaceutical properties be well known.

Third External Data Set: A Translation into the Real World of Tyrosinase Inhibitory Assays. In Silico and In Vitro Results.

The performance of the results obtained above encouraged us to carry out an *in silico* screening to search for novel lead compounds with tyrosinase inhibitory activity, as a way to show the applicability of the QSAR models obtained with the ***TOMOCOMD-CARDD*** approach, in the selection of hit or lead compounds.

In this sense, it was made an *in silico* screening of a pool of compounds, with structural features not presented in the database, to prove the extrapolative power of the QSAR models and find new tyrosinase-inhibitor chemicals.

A family of lignans, isolated from natural sources, was chosen to be evaluated with the LDA-based QSAR models, and posterior *in vitro* assays were done to corroborate the *in silico* predictions. The results of the classification of the compounds in this set for all the models are summarized in Table 7. This Table also shows the values of ΔP% for compounds in these series, as well as their canonical scores. As can be observed all the compounds presented activity against the tyrosinase enzyme.

These theoretical results showed an adequate correspondence with the experimental activity. An aspect of great importance to remark here is that all the compounds present greater inhibition values than Kojic Acid (standard tyrosinase inhibitor: IC_{50} =16.67μM). In addition, the Lignan 5 exhibited a more potent activity compared with L-mimosine (IC_{50} = 3.68μM). In Figure 10 are depicted the molecular structures of the compounds.

Table 7 comes about here (see end of the document)

Figure 10 comes about here (see end of the document)

It is important to stand out here, which the family employed in this study consists of structural core not present in the active database. These results exemplify how the current algorithm can be used for the selection/identification of novel lead tyrosinase-inhibitor compounds. Therefore, these compounds can be taken as *hits* and selected for a further chemistry optimization, improving the inhibitory activity and the rest of the ADMET properties. To this effect, this series of compounds with these structural features could be used to act as a potential lead molecule in the treatment of any disorders associated with tyrosinase inhibitory activity.

CONCLUSIONS

During the last decades, one of the researches involving the tyrosinase enzyme has been focused on the search for novel tyrosinase-inhibitor compounds. The wide distribution of this enzyme in the phylogenetic tree makes possible its wide use in pharmaceutical, food, agricultural chemistry and so on. For instance, in human beings this enzyme plays an important role in hyperpigmentation and melanogenesis disorders.

However, until nowadays the search of compounds with anti-tyrosinase activity has been based in trial-error methods [54].

Therefore, *in silico* screening can become a suitable tool in the development of drug discovery, capable of solving the ancient problem of large database query from a determined activity. Consequently, the QSAR models, together with HTS, can highlight the potential for new applications of *in silico* techniques, in the selective assessment by screening of a vast number of molecules against the tyrosinase inhibitory activity, reducing the *in vitro* tests.

Following this aim, in the present report we proposed a new set of molecular descriptors, namely *non-stochastic and stochastic bond-based quadratic indices*, as well as its application to the discrimination of the database of compounds, as either tyrosinase inhibitors or inactive substance. In addition, a virtual screening of compounds by using the QSAR models was carried out and showed adequate results. Finally, the *in silico* screening of a family of lignans and its experimental corroboration was developed.

This presented algorithm can effectively reduce the need to achieve a sufficient number of hits by increasing the speed in the selection/identification of new lead compounds. Hit finding, following these effective methods could be used to improve the search for novel depigmentation agents.

ACKNOWLEDGEMENTS: One of the authors (M-P. Y) thanks the program '*Estades Temporals per a Investigadors Convidats*' for a fellowship to work at Valencia University (2008). M-P. Y thanks are also given to the Generalitat Valenciana, (Spain) for partial financial support as well as support from Spanish MEC (Project Reference:

SAF2006-04698). MTHK is the recipient of a grant from MCBN-UNESCO (grant no. 1056). F. T. acknowledges financial support from the Spanish MEC DGI (Project No.CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173. Finally, but not less, M-P, Y thanks are given to the projects entitle “Strengthening postgraduate education and research in Pharmaceutical Sciences". This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium.

SUPPLEMENTARY DATA AVAILABLE: The complete list of compounds used in training and prediction sets, as well as their structures, posterior classification and scores according to LDA-based QSAR models, chemistry and data analysis of the obtained chemicals is available free of charge via Internet at…

REFERENCES AND NOTES

[1] G.W. Gadd, FEMS Microbiol. Lett., 9 (1980) 237.

[2] A.A. Bell, M.H. Wheeler, Annu. Rev. Phytopathol., 24 (1986) 411.

[3] W.C. Zimmerman, R.A. Blanchette, T.A. Burnes, R.L. Farrel, Mycologia, 87 (1995) 857.

[4] K. Ohguchi, T. Tanaka, T. Kido, K. Baba, M. Iinuma, K. Matsumoto, Y. Akao, Y. Nozawa, Biochem Biophys Res Commun, 307 (2003) 861.

[5] Y. Xu, A.H. Stokes, W.M. Freeman, S.C. Kumer, B.A. Vogt, K.E. Vrana, Mol Brain Res, 45 (1997) 159.

[6] Q.X. Chen, I. Kubo, J Agric Food Chem, 50 (2002) 4108.

[7] M.T.H. Khan, M.I. Choudhary, A. Ather, Atta-ur-Rahman., Minerva Biotechnologica, Accepted for publication (2006).

[8] W. Li, I. Kubo, Bioorg Med Chem, 12 (2004) 701.

[9] G.M. Casañola-Martin, M.T. Khan, Y. Marrero-Ponce, A. Ather, M.N. Sultankhodzhaev, F. Torrens, Bioorg Med Chem Lett, 16 (2006) 324.

[10] Y. Marrero-Ponce, M.T.H. Khan, G.M. Casañola-Martín, A. Ather, M.N. Sultankhodzhaev, F. Torrens, R. Rotondo, ChemMedChem (2006) Submitted for publication.

[11] Y. Marrero-Ponce, M.T.H. Khan, G.M. Casañola-Martín, A. Ather, M.N. Sultankhodzhaev, F. Torrens, QSAR Comb. Sci. (2006) Accepted for publication.

[12] G.M. Casañola-Martin, M.T.H. Khan, Y. Marrero-Ponce, A. Ather, S. Sultan, F. Torrens, R. Rotondo, Bioorg. Med. Chem. (2006) DOI:10.1016/j.bmc.2006.10.067.

[13] 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. Aran, A.R. Martinez-Fernandez, F. Torrens, R. Rotondo, F. Ibarra-Velarde, Y.J. Alvarado, Bioorg. Med. Chem., 14 (2006) 6502.

[14] J. Bajorath, Nat Rev Drug Discov, 1 (2002) 882.

[15] B. Pirard, Comb Chem High Throughput Screen, 7 (2004) 271.

[16] C.J. Manly, S. Louise-May, J.D. Hammer, Drug Discov Today, 6 (2001) 1101.

[17] Y. Marrero-Ponce, Molecules, 8 (2003) 687.

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

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

[20] M.C. Vega, A. Montero-Torres, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J.A. Escario, V.J. Aran, J.J. Nogal, A. Meneses-Marcel, F. Torrens, Bioorg Med Chem Lett, 16 (2006) 1898.

[21] Y. Marrero-Ponce, R. Medina-Marrero, Y. Martinez, F. Torrens, V. Romero-Zaldivar, E.A. Castro, J. Mol. Mod., 12 (2006) 255.

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

[23] Y. Marrero Ponce, J.A. Castillo Garit, D. Nodarse, Bioorg Med Chem, 13 (2005) 3397.

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

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

[26] Y. Marrero-Ponce, F. Torrens, J. Comp-Aided Mol. Des., 20 (2006) 685.

[27] L. Pauling, The Nature of Chemical Bond, Cornell University Press, Ithaca (New York), 1939.

[28] R. Todeschini, P. Gramatica, Perspect. Drug Dis. Des., 9-11 (1998) 355.

[29] V. Consonni, R. Todeschini, M. Pavan, J Chem Inf Comput Sci, 42 (2002) 682.

[30] L.B. Kier, L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Research Studies Press, Letchworth, U. K, 1986.

[31] H. Azhar Ul, A. Malik, M.T.H. Khan, Anwar-ul-Haq, S.B. Khan, A. Ahmad, M.I. Choudhary, Phytomedicine, 13 (2006) 255.

[32] V.J. Hearing, Methods in Enzymology, Academic Press, New York, 1987.

[33] O. Nerya, R. Musa, S. Khatib, S. Tamir, J. Vaya, Phytochemistry, 65 (2004) 1389.

[34] H. Kim, J. Choi, J.K. Cho, S.Y. Kim, Y.S. Lee, Bioorg Med Chem Lett, 14 (2004) 2843.

[35] M. Shiino, Y. Watanabe, K. Umezawa, Bioorg Chem, 31 (2003) 129.

[36] N. Yokochi, T. Morita, T. Yagi, J Agric Food Chem, 51 (2003) 2733.

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

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

[39] v.S.I. STATISTICA (data analysis software system), 2001. www.statsoft.com.

[40] J.W. Mc Farland, D.J. Gans, in: H. Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH Publishers, New York, 1995, p. 295.

[41] H. van de Waterbeemd, in: H. van de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH Publishers, Weinheim, 1995, p. 265.

[42] Y. Marrero-Ponce, H.G. Díaz, V. Romero, F. Torrens, E.A. Castro, Bioorg. Med. Chem. , 12 (2004) 5331.

[43] Y. Marrero-Ponce, M.A. Cabrera, V. Romero, D.H. González, F. Torrens, J. Pharm. Pharmaceut. Sci., 7 (2004) 186.

[44] Y. Marrero-Ponce, A. Montero-Torres, C.R. Zaldivar, M.I. Veitia, M.M. Perez, R.N. Sanchez, Bioorg Med Chem, 13 (2005) 1293.

[45] Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A.M. Sanchez, F. Torrens, E.A. Castro, Bioorg Med Chem, 13 (2005) 1005.

[46] E. Estrada, S. Vilar, E. Uriarte, Y. Gutierrez, J Chem Inf Comput Sci, 42 (2002) 1194.

[47] M. Randić, J Chem Inf Comput Sci, 31 (1991) 311.

[48] M. Randić, New J Chem, 15 (1991) 517.

[49] M. Randić, J Mol Struct (Theochem), 233 (1991) 45.

[50] B. Lučić, S. Nikolić, N. Trinajstić, D. Jurić, J Chem Inf Comput Sci (1995) 532.

[51] S. Wold, L. Erikson, in: H. van de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH Publishers, New York, 1995, p. 309.

[52] A. Golbraikh, A. Tropsha, J Mol Graph Model, 20 (2002) 269.

[53] D.F. Horrobin, J. R. Soc. Med, 93 (2000) 341.

[54] S. Okombi, D. Rival, S. Bonnet, A.M. Mariotte, E. Perrier, A. Boumendjel, J Med Chem, 49 (2006) 329.

ANNEXES
(Tables and Schemes to be Inserted in the Main Text)

Table 1. Values of the Atomic Weights Used for Quadratic Indices Calculation.[25-28]

ID	Atomic Mass g/mol	VdW[a] Volume Å^3	Mulliken Electronegativity	Polarizability Å^3	Pauling Electronegativity
H	1.01	6.709	2.592	0.667	2.2
B	10.81	17.875	2.275	3.030	2.04
C	12.01	22.449	2.746	1.760	2.55
N	14.01	15.599	3.194	1.100	3.04
O	16.00	11.494	3.654	0.802	3.44
F	19.00	9.203	4.000	0.557	3.98
Al	26.98	36.511	1.714	6.800	1.61
Si	28.09	31.976	2.138	5.380	1.9
P	30.97	26.522	2.515	3.630	2.19
S	32.07	24.429	2.957	2.900	2.58
Cl	35.45	23.228	3.475	2.180	3.16
Fe	55.85	41.052	2.000	8.400	1.83
Co	58.93	35.041	2.000	7.500	1.88
Ni	58.69	17.157	2.000	6.800	1.91
Cu	63.55	11.494	2.033	6.100	1.9
Zn	65.39	38.351	2.223	7.100	1.65
Br	79.90	31.059	3.219	3.050	2.96
Sn	118.71	45.830	2.298	7.700	1.96
I	126.90	38.792	2.778	5.350	2.66

[a]van der Waals

Table 2. Main Results of the *k*-MCAs for Tyrosinase Inhibitor and Inactive Drug-like Compounds.

Analysis of Variance				
Variables	**Between SS[a]**	**Within SS[b]**	**Fisher ratio (F)**	***p*-level[c]**
Tyrosinase Inhibitors Clusters (*k*-MCA I)				
${}^{\mathbf{M}}\boldsymbol{q}_5(x)$	45.82	11.05	109.23	0.00
${}^{\mathbf{V}}\boldsymbol{q}_1(x)$	265.71	42.40	165.02	0.00
${}^{\mathbf{V}}\boldsymbol{q}_{0\mathrm{L}}(x_{\mathrm{E}})$	209.08	35.80	153.77	0.00
${}^{\mathbf{P}}\boldsymbol{q}_1{}^{\mathrm{H}}(x)$	20.41	4.21	127.62	0.00
${}^{\mathbf{K}}\boldsymbol{q}_{3\mathrm{L}}{}^{\mathrm{H}}(x_{\mathrm{E}})$	204.94	42.53	126.88	0.00
${}^{\mathbf{G}}\boldsymbol{q}_5{}^{\mathrm{H}}(x)$	207.43	36.49	149.70	0.00
${}^{\mathbf{G}}\boldsymbol{q}_{0\mathrm{L}}(x_{\mathrm{E}})$	343.72	65.75	137.66	0.00
${}^{\mathbf{G}}\boldsymbol{q}_{1\mathrm{L}}(x_{\mathrm{E}})$	258.35	28.42	239.36	0.00
Inactives Clusters (*k*-MCA II)				
${}^{\mathbf{M}}\boldsymbol{q}_5(x)$	1357.05	96.66	510.53	0.00
${}^{\mathbf{V}}\boldsymbol{q}_1(x)$	265.87	82.04	117.84	0.00
${}^{\mathbf{V}}\boldsymbol{q}_{0\mathrm{L}}(x_{\mathrm{E}})$	495.51	139.01	129.62	0.00
${}^{\mathbf{P}}\boldsymbol{q}_1{}^{\mathrm{H}}(x)$	448.54	82.38	197.99	0.00
${}^{\mathbf{K}}\boldsymbol{q}_{3\mathrm{L}}{}^{\mathrm{H}}(x_{\mathrm{E}})$	487.84	69.64	254.72	0.00
${}^{\mathbf{G}}\boldsymbol{q}_5{}^{\mathrm{H}}(x)$	303.50	58.60	188.33	0.00
${}^{\mathbf{G}}\boldsymbol{q}_{0\mathrm{L}}(x_{\mathrm{E}})$	311.16	55.16	205.13	0.00
${}^{\mathbf{G}}\boldsymbol{q}_{1\mathrm{L}}(x_{\mathrm{E}})$	429.14	69.81	223.55	0.00

[a]Variability between groups.
[b]Variability within groups.
[c]Level of significance.

Table 3. Discriminant Models Obtained with Total and Local Non-Stochastic and Stochastic Bond-Based Quadratic Indices Used in This Study.

LDA-Based QSAR Models Obtained Using Non-Stochastic Bond-Based Quadratic Indices

$$\textit{Class} = -0.259 + 1.068\text{x}10^{-2}\,{}^{M}q_0^{H}(w) - 2.868\text{x}10^{-3}\,{}^{M}q_1^{H}(w) - 3.142\text{x}10^{-3}\,{}^{M}q_0(w) + 1.482\text{x}10^{-3}\,{}^{M}q_1(w) - 3.761\text{x}10^{-6}\,{}^{M}q_5(w) + 4.800\text{x}10^{-12}\,{}^{M}q_{15}(w) - 1.672\text{x}10^{-2}\,{}^{M}q_{0L}^{H}(w_E) + 7.041\text{x}10^{-3}\,{}^{M}q_{1L}^{H}(w_E) + 9.082\text{x}10^{-3}\,{}^{M}q_{0L}(w_E) - 6.268\text{x}10^{-3}\,{}^{M}q_{1L}(w_E) + 9.588\text{x}10^{-5}\,{}^{M}q_{3L}(w_E) \quad (3)$$

$$\textit{Class} = -0.902 - 2.631\text{x}10^{-3}\,{}^{V}q_0^{H}(w) + 2.141\text{x}10^{-3}\,{}^{V}q_1^{H}(w) - 1.076\text{x}10^{-4}\,{}^{V}q_3^{H}(w) - 9.834\text{x}10^{-7}\,{}^{V}q_6^{H}(w) + 6.289\text{x}10^{-13}\,{}^{V}q_{15}^{H}(w) - 6.570\text{x}10^{-4}\,{}^{V}q_0(w) + 1.088\text{x}10^{-3}\,{}^{V}q_1(w) + 2.237\text{x}10^{-6}\,{}^{V}q_{6L}^{H}(w_E) - 4.913\text{x}10^{-13}\,{}^{V}q_{15L}^{H}(w_E) + 1.881\text{x}10^{-3}\,{}^{V}q_{0L}(w_E) - 2.219\text{x}10^{-3}\,{}^{V}q_{1L}(w_E) - 5.055\text{x}10^{-8}\,{}^{V}q_{8L}(w_E) \quad (4)$$

$$\textit{Class} = -0.207 - 1.925\text{x}10^{-2}\,{}^{P}q_1^{H}(w) - 2.845\text{x}10^{-7}\,{}^{P}q_{10}^{H}(w) + 8.955\text{x}10^{-11}\,{}^{P}q_{15}^{H}(w) - 0.146\,{}^{P}q_2(w) + 0.108\,{}^{P}q_3(w) - 5.582\text{x}10^{-3}\,{}^{P}q_5(w) + 4.786\text{x}10^{-7}\,{}^{P}q_{11}(w) - 4.647\text{x}10^{-10}\,{}^{P}q_{15}(w) - 0.207\,{}^{P}q_{1L}(w_E) + 2.894\text{x}10^{-2}\,{}^{P}q_{2L}(w_E) \quad (5)$$

$$\textit{Class} = -0.743 + 4.671\text{x}10^{-5}\,{}^{K}q_5(w) - 0.185\,{}^{K}q_{0L}^{H}(w_E) + 0.128\,{}^{K}q_{1L}^{H}(w_E) + 5.546\text{x}10^{-2}\,{}^{K}q_{2L}^{H}(w_E) - 3.725\text{x}10^{-2}\,{}^{K}q_{3L}^{H}(w_E) + 4.781\text{x}10^{-3}\,{}^{K}q_{4L}^{H}(w_E) + 0.116\,{}^{K}q_{0L}(w_E) - 5.687\text{x}10^{-2}\,{}^{K}q_{1L}(w_E) - 1.800\text{x}10^{-3}\,{}^{K}q_{3L}(w_E) \quad (6)$$

$$\textit{Class} = -0.869 - 1.441\,{}^{G}q_5^{H}(w) + 4.098\text{x}10^{-9}\,{}^{G}q_{11}^{H}(w) + 1.723\text{x}10^{-2}\,{}^{G}q_1(w) + 8.038\text{x}10^{-5}\,{}^{G}q_5(w) - 0.217\,{}^{G}q_{0L}^{H}(w_E) + 0.145\,{}^{G}q_{1L}^{H}(w_E) + 5.870\text{x}10^{-2}\,{}^{G}q_{2L}^{H}(w_E) - 3.860\text{x}10^{-2}\,{}^{G}q_{3L}^{H}(w_E) + 5.675\text{x}10^{-3}\,{}^{G}q_{4L}^{H}(w_E) + 0.132\,{}^{G}q_{0L}(w_E) - 9.129\text{x}10^{-2}\,{}^{G}q_{1L}(w_E) - 8.556\text{x}10^{-4}\,{}^{G}q_{4L}(w_E) \quad (7)$$

$$\textit{Class} = -1.054 + 1.089\text{x}10^{-6}\,{}^{M}q_5(w) + 3.189\text{x}10^{-4}\,{}^{V}q_1(w) - 1.848\text{x}10^{-3}\,{}^{V}q_{0L}(w_E) - 1.832\text{x}10^{-2}\,{}^{P}q_1^{H}(w) - 2.050\text{x}10^{-3}\,{}^{K}q_{3L}^{H}(w_E) - 5.131\text{x}10^{-5}\,{}^{G}q_5^{H}(w) + 0.140\,{}^{G}q_{0L}(w_E) - 3.375\text{x}10^{-2}\,{}^{G}q_{1L}(w_E) \quad (8)$$

LDA-Based QSAR Models Obtained Using Stochastic Bond-Based Quadratic Indices

$$\textit{Class} = -0.554 - 4.436\text{x}10^{-2}\,{}^{M}q_5^{H}(w) + 4.499\text{x}10^{-2}\,{}^{M}q_8^{H}(w) - 4.467\text{x}10^{-3}\,{}^{M}q_1(w) + 5.007\text{x}10^{-3}\,{}^{M}q_6(w) - 6.296\text{x}10^{-3}\,{}^{M}q_{0L}^{H}(w_E) + 2.291\text{x}10^{-3}\,{}^{M}q_{1L}^{H}(w_E) + 5.328\text{x}10^{-2}\,{}^{M}q_{5L}^{H}(w_E) - 4.981\text{x}10^{-2}\,{}^{M}q_{8L}^{H}(w_E) + 6.024\text{x}10^{-3}\,{}^{M}q_{0L}(w_E) - 4.707\text{x}10^{-3}\,{}^{M}q_{8L}(w_E) - 1.118\text{x}10^{-2}\,{}^{M}q_{9L}(w_E) + 9.524\text{x}10^{-3}\,{}^{M}q_{11L}(w_E) \quad (9)$$

$$\textit{Class} = -0.321 - 4.726\text{x}10^{-3}\,{}^{V}q_0^{H}(w) - 3.297\text{x}10^{-2}\,{}^{V}q_2^{H}(w) + 3.921\text{x}10^{-2}\,{}^{V}q_4^{H}(w) - 9.564\text{x}10^{-4}\,{}^{V}q_0(w) + 2.259\text{x}10^{-3}\,{}^{V}q_{15}(w) - 4.954\text{x}10^{-3}\,{}^{V}q_{1L}^{H}(w_E) + 2.885\text{x}10^{-2}\,{}^{V}q_{2L}^{H}(w_E) - 8.300\text{x}10^{-2}\,{}^{V}q_{7L}^{H}(w_E) + 5.641\text{x}10^{-2}\,{}^{V}q_{8L}^{H}(w_E) + 2.230\text{x}10^{-3}\,{}^{V}q_{0L}(w_E) - 3.201\text{x}10^{-3}\,{}^{V}q_{2L}(w_E) \quad (10)$$

$$\textit{Class} = -0.689 + 0.645\,{}^{P}q_0^{H}(w) - 3.048\,{}^{P}q_2^{H}(w) + 1.997\,{}^{P}q_4^{H}(w) + 0.602\,{}^{P}q_{15}^{H}(w) - 0.231\,{}^{P}q_0(w) + 0.229\,{}^{P}q_{14}(w) - 0.924\,{}^{P}q_{0L}^{H}(w_E) + 2.302\,{}^{P}q_{2L}^{H}(w_E) - 1.538\,{}^{P}q_{11L}^{H}(w_E) + 0.519\,{}^{P}q_{0L}(w_E) - 0.234\,{}^{P}q_{1L}(w_E) - 0.378\,{}^{P}q_{2L}(w_E) \quad (11)$$

$$\textit{Class} = -1.278 - 0.205\,{}^{K}q_0^{H}(w) - 1.393\,{}^{K}q_3^{H}(w) + 1.164\,{}^{K}q_4^{H}(w) + 0.450\,{}^{K}q_{15}^{H}(w) + 4.829\text{x}10^{-2}\,{}^{K}q_{14}(w) + 0.618\,{}^{K}q_{2L}^{H}(w_E) - 0.687\,{}^{K}q_{10L}^{H}(w_E) + 9.772\text{x}10^{-2}\,{}^{K}q_{0L}(w_E) + 7.486\text{x}10^{-2}\,{}^{K}q_{1L}(w_E) - 0.229\,{}^{K}q_{3L}(w_E) \quad (12)$$

$$\textit{Class} = -1.235 - 0.236\,{}^{G}q_0^{H}(w) - 1.619\,{}^{G}q_3^{H}(w) + 1.346\,{}^{G}q_4^{H}(w) + 0.519\,{}^{G}q_{15}^{H}(w) + 5.781\text{x}10^{-2}\,{}^{G}q_{14}(w) + 0.719\,{}^{G}q_{2L}^{H}(w_E) - 0.783\,{}^{G}q_{10L}^{H}(w_E) + 0.113\,{}^{G}q_{0L}(w_E) + 9.073\text{x}10^{-2}\,{}^{G}q_{1L}(w_E) - 0.272\,{}^{G}q_{3L}(w_E) \quad (13)$$

$$\textit{Class} = -0.285 - 8.339\text{x}10^{-3}\,{}^{V}q_0(w) + 1.482\text{x}10^{-2}\,{}^{V}q_{2L}^{H}(w_E) + 1.225\,{}^{P}q_0(w) - 0.683\,{}^{P}q_{0L}^{H}(w_E) - 1.227\,{}^{P}q_{2L}^{H}(w_E) + 0.517\,{}^{P}q_{11L}^{H}(w_E) - 0.335\,{}^{P}q_{2L}(w_E) - 0.124\,{}^{K}q_{0L}(w_E) + 9.174\text{x}10^{-2}\,{}^{G}q_{14}(w) - 0.260\,{}^{G}q_{10L}^{H}(w_E) - 0.150\,{}^{G}q_{3L}(w_E) \quad (14)$$

Table 4. 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 Quadratic Indices										
Eq. 3 (12)	0.81	90.78	86.4	90.2	8.8	0.48	4.58	42.1	376.4	0.74
Eq. 4 (12)	0.73	87.24	86.3	79.2	7.8	0.55	3.46	31.8	281.6	0.67
Eq. 5 (10)	0.76	88.49	82.7	88.5	11.5	0.51	4.06	45.0	257.0	0.65
Eq. 6 (9)	0.76	88.49	83.7	86.9	10.5	0.49	4.35	53.7	334.6	0.71
Eq. 7 (12)	0.75	87.87	82.1	87.4	11.9	0.48	4.58	42.2	346.0	0.72
Eq. 8 (8)	**0.86**	**93.51**	**91.8**	**91.3**	**5.1**	**0.41**	**5.99**	**83.3**	**417.1**	**0.77**
LDA-Based QSAR Models Obtained Using Stochastic Quadratic Indices										
Eq. 9 (12)	0.71	85.98	79.6	85.2	13.6	0.55	3.45	31.7	281.2	0.67
Eq. 10 (11)	0.72	86.40	79.8	86.3	13.6	0.53	3.70	37.2	296.5	0.68
Eq. 11 (12)	0.69	85.56	82.0	79.8	10.8	0.59	2.98	27.4	251.5	0.64
Eq. 12 (10)	0.77	89.12	83.9	88.5	10.5	0.48	4.60	51.0	347.8	0.72
Eq. 13 (10)	0.78	89.33	84.0	89.1	10.5	0.48	4.62	51.2	348.6	0.72
Eq. 14 (11)	**0.82**	**91.21**	**87.3**	**90.2**	**8.1**	**0.44**	**5.31**	**53.4**	**383.7**	**0.75**

[a] Between parenthesis the quantity of variables of the models.
[b] Canonical correlation coefficient obtained from the linear discriminant **canonical** analysis

Table 5. 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 Bond-Based Quadratic Indices					
Eq. 3	0.81	91.11	85.1	90.5	8.5
Eq. 4	0.67	85.00	80.0	76.2	10.3
Eq. 5	0.72	86.67	76.7	88.9	14.5
Eq. 6	0.72	86.11	75.0	90.5	16.2
Eq. 7	0.66	83.30	72.0	85.7	17.9
Eq. 8	**0.79**	**90.00**	**81.7**	**92.1**	**11.1**
LDA-based QSAR Models Obtained Using Stochastic Bond-Based Quadratic Indices					
Eq. 9	0.71	86.10	76.4	87.3	14.5
Eq. 10	0.71	85.56	74.0	90.5	17.1
Eq. 11	0.61	82.22	75.4	73.0	12.8
Eq. 12	0.74	87.78	79.7	87.3	12.0
Eq. 13	0.70	85.56	75.3	87.3	15.4
Eq. 14	**0.78**	**89.44**	**80.6**	**92.1**	**12.0**

Table 6. Results for the Virtual Screening of Tyrosinase Inhibitors.

Compound[a]	Class[b]	Ref.[c]	Compound[a]	Class[b]	Ref.[c]
1 *p*-nitrophenol	+-++++	A	**27** Dithiothreitol	++-+++	O
	++++++	B		++++++	
2 3-(3,4-dihydroxyphenyl)-l-alanine	++++++	C	**28** Azelaic acid	+-++++	P
	++++++			-+-+++	
3 3-amino-4-hydroxybenzoic acid	++++++	C	**29** Undecandioic acid	+-++++	P
	++++++			-+++++	
4 4-amino-3-hydroxybenzoic acid	++++++	C	**30** Suberic acid	+-++++	P
	++++++			-+-+++	
5 3,4-diaminobenzoic acid	++++++	C	**31** Sebacic acid	+-++++	P
	++++++			-+++++	
6 3-aminobenzoic acid	**++++++**	C	**32** Dodecandioic acid	+-++++	P
	++++++			++++++	
7 4-aminobenzoic acid	++++++	C	**33** Tridecandioic acid	+-++++	P
	++++++			++++++	
8 4,6-O-hexahydroxy	+--+-+	D	**34** Traumatic acid	+-++++	P
diphenylglucose	++++++			++++++	
9 Tunicamycin	+--+++	E	**35** Pantothenic acid	++++++	K
	++++++			++++++	
10 methyl *p*-coumarate	++++++	F	**36** 5-(hydroxymethyl)-2-furfural	++++++	Q
	++++++			+--+++	R
11 o-phenylphenol	++++++	F	**37** Hinokitiol	++++++	S
	++++++			++-+++	
12 Phenylhydroquinone	++++++	F	**38** Penicillamine	++++++	T
	++++++			++++++	
13 Chamaecin	++++++	F	**39** Toluic acid	++++++	A
	++---+	G		++++++	
14 Stearyl glycyrrhetinate	++++++	H	**40**	++++++	U
	++++++			++++++	
15 2-(4-Methylphenyl)-	+++--+	I	**41**	++++++	U
1,3-selenazol-4-one	------	J		++++++	
16	-++--+	I	**42** 3,5-dihydroxy-	++++++	V
	------		4´-O-methoxystilbene	++++++	
17	-++--+	I	**43** p-hydroxybenzoic acid	++++++	W
	------			++++++	
18	+-+---	I	**44** o-hydroxybenzoic acid	++++++	W
	------			++++++	
19 3-fluorotyrosine	++++++	K	**45** Cysteine	++-+++	X
	++++++			++++++	
20 N-acetyltyrosine	--+--+	K	**46** Methimazole	+-----	X
	-+-+++			+----+	
21 N-formyltyrosine	+-++++	K	**47** BMY-28438	++++++	X
	++++++			++++++	
22 Gentisic acid	++++++	L	**48** Captopril	+--++-	Y
	++++++			+----+	
23 6-BH_4	-----+	M	**49** Yohimbine	+-++-+	Z
	+----+			++++++	
24 7-BH_4	-----+	M	**50** 4-(phenylazo)phenol	++++++	*a*
	+----+			++++++	
25 Propylparaben	++++++	N	**51** SACat	++++++	*a*
	++-+++			++++++	
26 Phenylalanine	++++++	K	**52** NPACat	++++++	*a*
	++++++			++++++	

Table 6. *Cont.*

Compound[a]	Class[b]	Ref.[c]	Compound[a]	Class[b]	Ref.[c]
53 DNPACat	++++++	*a*	**70**	++++++	*d*
	++++++			++-+++	
54 EDTA	-----+	*b*	**71**	++++++	*d*
	------			++++++	
55 Dodecyl gallate	++++++	*c*	**72**	++++++	*d*
	++++++			++-+++	
56 Gallic acid	++++++	*c*	**73**	++++++	*d*
	++++++			++++++	
57 (±)-flavanone	+++--+	*d*	**74**	++++++	*d*
	-++--+			++++++	
58 (-)-pinocembrin	++++++	*d*	**75**	++++++	*d*
	++++++			++++++	
59 (±)-naringenin	++++++	*d*	**76**	++++++	*d*
	++++++			++++++	
60 (+)-dihydromorin	++++++	*d*	**77**	++++++	*d*
	++++++			++++++	
61 Flavone	+++--+	*d*	**78**	++++++	*d*
	-++--+			++++++	
62 Myricetin	++++++	*d*	**79**	++++++	*d*
	++++++			++++++	
63 Artocarpin	++++++	*d*	**80**	++++++	*d*
	++++++			+--+++	
64 Artocarpesin	++++++	*d*	**81**	++++++	*d*
	++++++			++++++	
65 Isoartocarpesin	++++++	*d*	**82**	++++++	*d*
	++++++			++-+++	
66 (-)-Angolensin	++++++	*d*	**83**	++++++	*d*
	++++++			++-+++	
67 Pinosylvin	++++++	*d*	**84** 2´-O-feruloylaloesin	++++++	*e*
	++++++			++++++	
68 4-prenyloxyresveratrol	++++++	*d*	**85** Barbaloin	++++++	*e*
	++++++			++++++	
69	++++++	*d*			
	++++++				

[a]The molecular structures of these tyrosinase inhibitors is given as Supplementary Data (see Table 11). [b]Results of the classification of compounds in this set: (i) *Above*, classification of each compounds using the obtained models with non-stochastic bond-based quadratic indices in the following order: Eq. 3, 4, 5, 6, 7 and 8; and (ii) *Below*; classification of each compounds using the obtained models with stochastic bond-based quadratic indices in the following order Eq. 9, 10, 11, 12, 13, and 14.[c]References taken from the literature: [A]Bubacco, L.; van Gastel, M.; Groenen, E. J. J.; Vijgenboom, E.; Canters, G. W. *J. Biol. Chem.* **2003**, *278*, 7381–7389. [B]van Gastela, M.; Bubaccob, L.; Groenena, E. J. J.; Vijgenboomc, E.; Cantersc, G. W. *FEBS Lett.* **2000**, *474*, 228-232. [C]Gasowskaa, B.; Kafarskia, P.; Wojtasek, H. *Biochim. Biophys. Acta*. **2004**, *1673*, 170–177. [D] http://open.cacb.org.tw/index.php (2005-03-03 09:09:51). [E]Takahashi, H.; Parsons, P. G. *J. Invest. Dermatol.* **1992**, *98*, 481-487. [F]Kubo, I.; Niheia, K.; Tsujimoto, K. *Bioorg. Med. Chem.* **2004**, *12*, 5349–5354. [G]Nihei, K-I.; Yamagiwa, Y.; Kamikawab, T.; Kubo, I. *Bioorg. Med. Chem Lett.* **2004**, *14*, 681–683. [H]Um, S-J.; Park, M-S.; Park, S-H.; Han, H-S.; Kwonb, Y-J.; Sin, H-S. *Bioorg. Med. Chem.* **2003**, *11*, 5345–5352. [I]Barlocco, D.; Barrett, D.; Edwards, P.; Langston, S.; Pérez-Pérez, M. J.; Walker, M.; Weidner, J.; Westwell, A. *Drug Disc. Today.* **2003**, *8*, 372-373. [J]Koketsu, M.; Choi, S.Y.; Ishihara, H.; Lim B. O.; Kim, H.; Kim, S., Y. *Chem. Pharm. Bull.* (*Tokyo*). **2002**, *12*, 1594-1596. http://www.thecosmeticsite.com/formulating/959621.htlm (April-00). [L]Curto, E. V.; Kwong, C.; Hermersdorfer, H.; Glatt, H.; Santis, C.; Virador, V.; Hearing, V. J.; Dooley, T. P. *Biochem. Pharmacol.* **1999**, *57*, 663–672. [M]Wood, J. M.; Schallreuter-Wood, K. U.; Lindsey, N. J.; Callaghan, S.; Gardner, M. L. G *Biochem. Biophys. Res. Commun*. **1995**, *206*, 480–485. [N]Hori, I.; Nihei, K-I.; Kubo, I. *Phytother. Res*. **2004**, *18*, 475–479. [O]Naish-Byfield, S.; Cooksey, C. J.; Riley, P. A. *Biochem. J*. **1994**, *304*, 155–162. [P]Nazzaro-Porro, M.; Passi, S. *J. Invest. Dermatol.* **1978**, *71*, 205-208. [Q]Sharma, V. K.; Choi, J.; Sharma, N.; Choi, M.; Seo, S-Y. *Phytotherapy Res.* **2004**, *18*, 841-844. [R]Kang, H. S.; Choi, J. H.; Cho, W. K.; Park, J. C.; Choi, J. S. *Arch Pharm Res*. **2004**, *7*, 742-50. [S]Sakuma, K.; Ogawa, M.; Sugibayashi, K.; Yamada, K.; Yamamoto, K. *Arch Pharm Res*. **1999**, *4*, 335-339. [T]Lovstad, R. A. *Biochem. Pharmacol*. **1976**, *25*, 533-535. [U]Kubo, I.; Kinst-Hori, I.; Yokokawa, Y. *J. Nat. Prod*. **1994**, *57*, 545-551. [V]Regev-Shoshani, G.; Shoseyov, O.; Bilkis, I.; Kerem, Z. *Biochem. J*. **2003**, *374*, 157–163. [W]Bernard, P.; Berthon, J-Y. *Int. J. Cosmetic Sci*. **2000**, *22*, 219-226. [X]Imada, C.; Sugimoto, Y.; Makimura, T.; Kobayashi, T.; Hamada, N.; Watanabe, E. *Fish. Sci*. **2001,** *67*, 1151–1156. [Y]Espín, J. C.; Wichers, H. J. *Biochim. Biophys. Acta*. **2001**, *1544*, 289-300. [Z]Fuller, B. B.; Drake, M. A.; Spaulding, D. T.; Chaudry, F. *J. Invest. Dermatol*. **2000**, *114*, 268-276. [a]Borojerdi, S. S.; Haghbeen, K.; Karkhane, A. A.; Fazli, M.; Sabouryc, A. A. *Biochem. Biophys. Res. Commun*. **2004**,

314, 925–930. [b]Kong, K-H.; Hong, M-P.; Choi, S-S.; Kim, Y-T.; Cho, S-H. *Biotechnol. Appl. Biochem*. **2000**, *31*, 113–118. [c]Kubo, I.; Chen, Q-X.; Nihei, K-I. *Food Chem*. **2003**, *81*, 241-247. [d]Shimizu, K.; Kondo, R.; Sakai, K. *Planta Medica*. **2000**, *66*, 11-15. [e]Yagi, A.; Kanbara, T.; Morinobu, N. *Planta Medica*. **1987**, 515-517.

Table 7. Results for Ligand-Based *in silico* Screening and Tyrosinase Inhibitory Activities of New Lignans.

Compound*	ΔP%[a]	Scores[a]	ΔP%[b]	Scores[b]	ΔP%[c]	Scores[c]	ΔP%[d]	Scores[d]	ΔP%[e]	Scores[e]	ΔP%[f]	Scores[f]	IC_{50} ±SEM[g] (μM)
Lignan 1	**93,98**	**-1,25**	**85,01**	**1,83**	**64,55**	**1,29**	**95,62**	**2,05**	**94,85**	**1,91**	**96,22**	**-1,82**	10.06±1.064
	94,41	*2,11*	*93,31*	*1,95*	*85,01*	*-1,93*	*94,93*	*-1,91*	*94,61*	*-1,86*	*87,17*	*1,36*	
Lignan 2	**99,88**	**-2,24**	**57,74**	**1,18**	**53,30**	**1,13**	**99,92**	**3,97**	**99,87**	**3,64**	**99,91**	**-3,34**	6.72±0.652
	99,78	*3,86*	*97,81*	*2,55*	*96,19*	*-2,76*	*99,94*	*-3,97*	*99,94*	*-3,97*	*99,71*	*3,03*	
Lignan 3	**99,38**	**-1,83**	**50,46**	**1,07**	**54,37**	**1,10**	**99,42**	**3,02**	**99,05**	**2,71**	**99,62**	**-2,76**	7.81±1.0971
	98,92	*3,00*	*97,72*	*2,52*	*95,27*	*-2,63*	*99,64*	*-3,16*	*99,64*	*-3,14*	*98,93*	*2,47*	
Lignan 4	**97,28**	**-1,55**	**92,40**	**2,21**	**78,82**	**1,40**	**94,87**	**1,97**	**95,09**	**1,93**	**97,78**	**-2,04**	9.76±1.1024
	95,94	*2,28*	*97,04*	*2,39*	*91,79*	*-2,30*	*98,19*	*-2,40*	*98,10*	*-2,36*	*91,70*	*1,56*	
Lignan 5	**99,92**	**-2,26**	**58,48**	**1,19**	**23,16**	**0,93**	**99,95**	**4,19**	**99,90**	**3,77**	**99,94**	**-3,52**	3.21±0.1654
	99,85	*4,05*	*97,13*	*2,40*	*96,05*	*-2,74*	*99,86*	*-3,61*	*99,87*	*-3,61*	*99,76*	*3,12*	
Lignan 6	**93,29**	**-1,00**	**59,50**	**1,21**	**37,78**	**0,75**	**70,89**	**1,07**	**61,81**	**0,89**	**90,45**	**-1,43**	15.13±1.9521
	82,08	*1,44*	*64,25*	*1,00*	*70,31*	*-1,49*	*43,19*	*-0,64*	*42,81*	*-0,62*	*94,37*	*1,74*	
Lignan 7	**97,46**	**-1,28**	**76,11**	**1,55**	**-34,37**	**0,49**	**87,48**	**1,52**	**79,52**	**1,23**	**95,18**	**-1,72**	5.61±0.3551
	94,26	*2,09*	*47,36*	*0,74*	*75,06*	*-1,61*	*45,55*	*-0,67*	*44,86*	*-0,65*	*95,59*	*1,85*	

*The molecular structures of these chemicals are shown in Figure 10. [a,b,c,d,e,f]ΔP% = [P(Active) - P(Inactive)]x100 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. 3, 4, 5, 6, 7, and 8; and (ii) Below in *italic*; classification of each compound using the obtained models with stochastic linear indices in the following order Eqs. 9, 10, 11, 12, 13, and 14. [g]IC_{50} are the 50% inhibitory concentrations against the enzyme tyrosinase and SEM is the standard error of the mean.

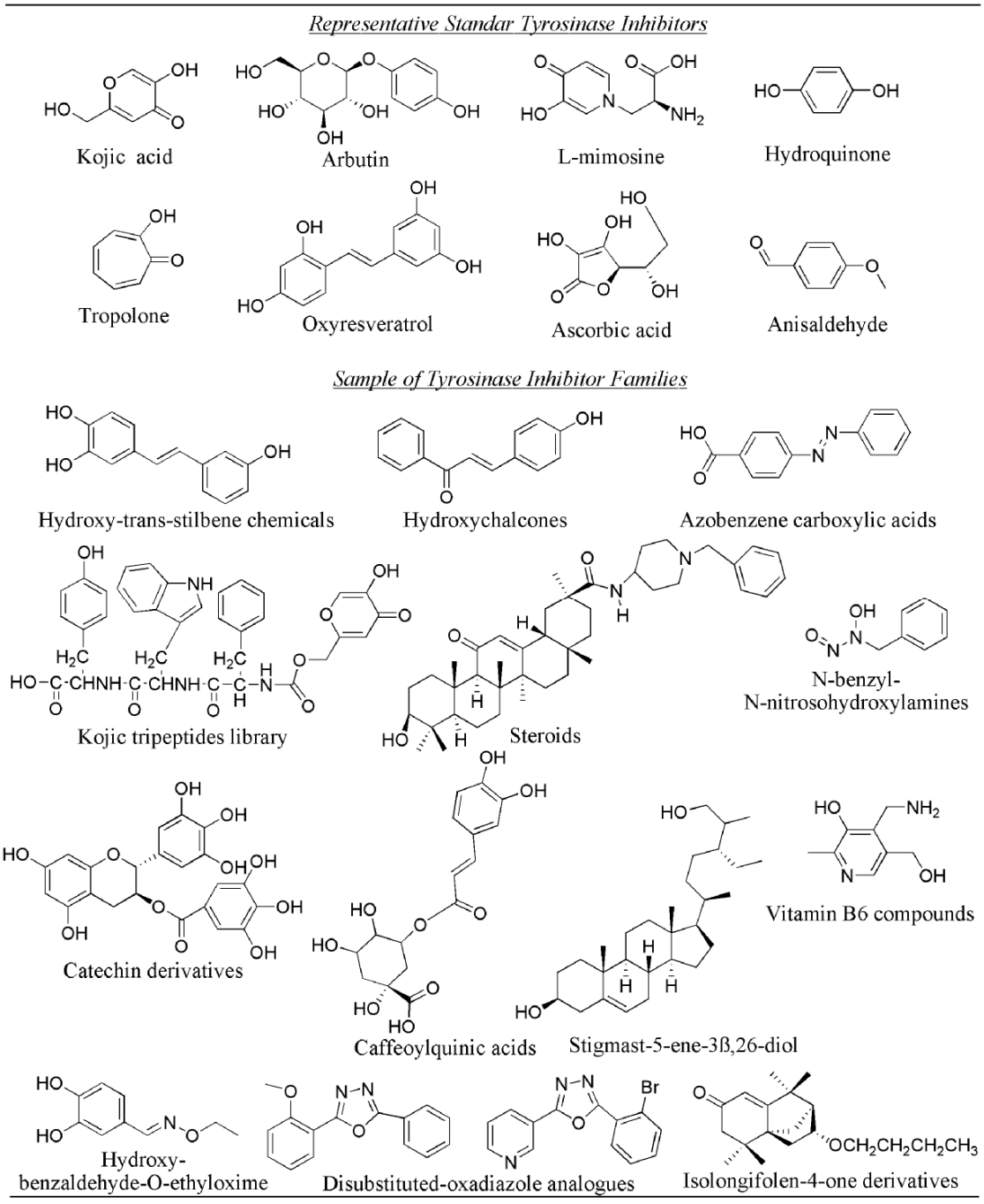

Figure 1. Random, but not exhaustive sample of the molecular families of tyrosinase inhibitors studied here and some reference drugs.

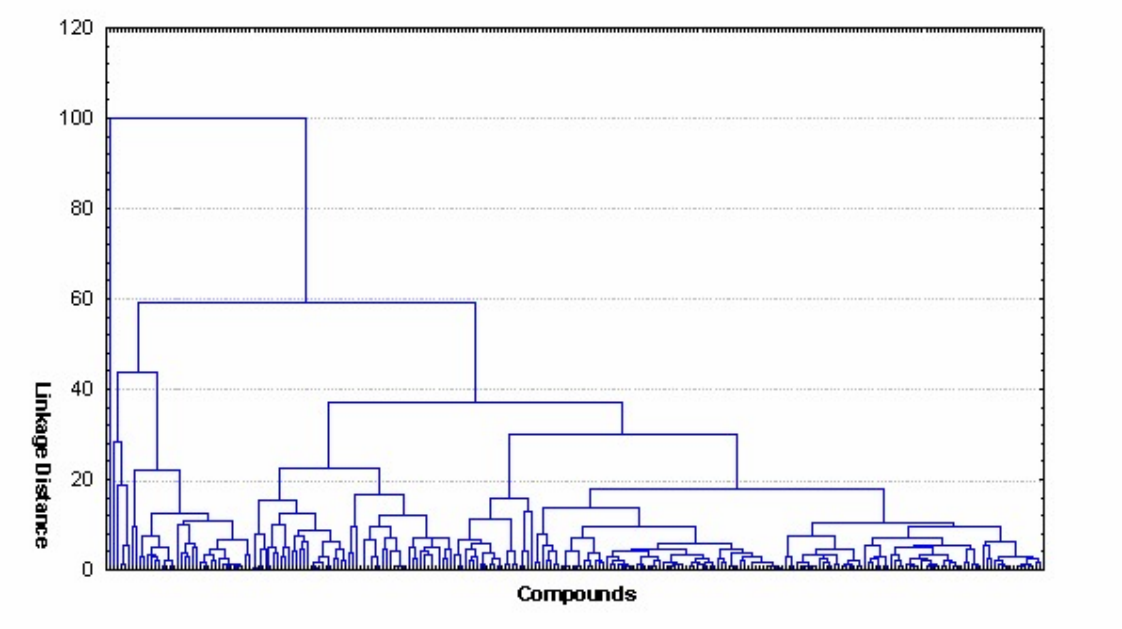

Figure 2. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of tyrosinase inhibitors used in the training and prediction set of the present work.

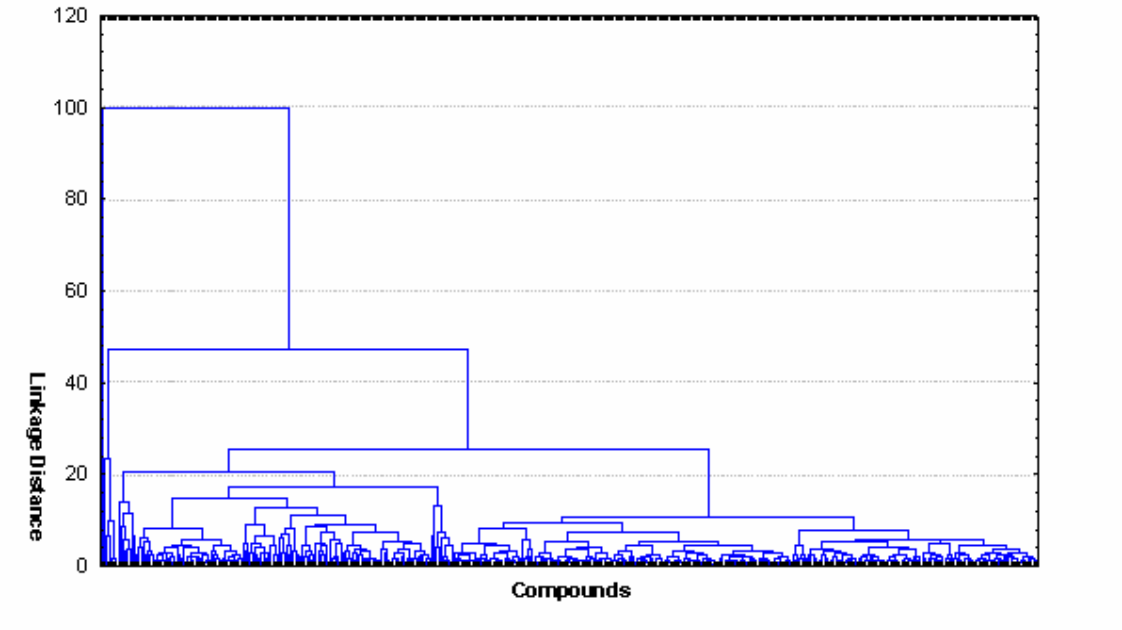

Figure 3. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of inactive compounds (non-tyrosinase inhibitors) used in the training and prediction set of the present work.

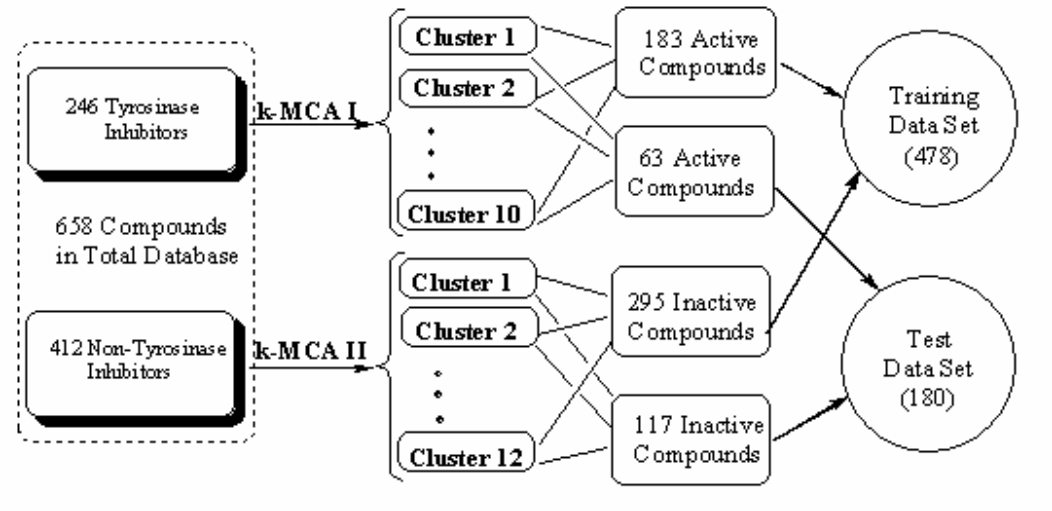

Figure 4. General algorithm used to design training and test sets through *k*-MCA.

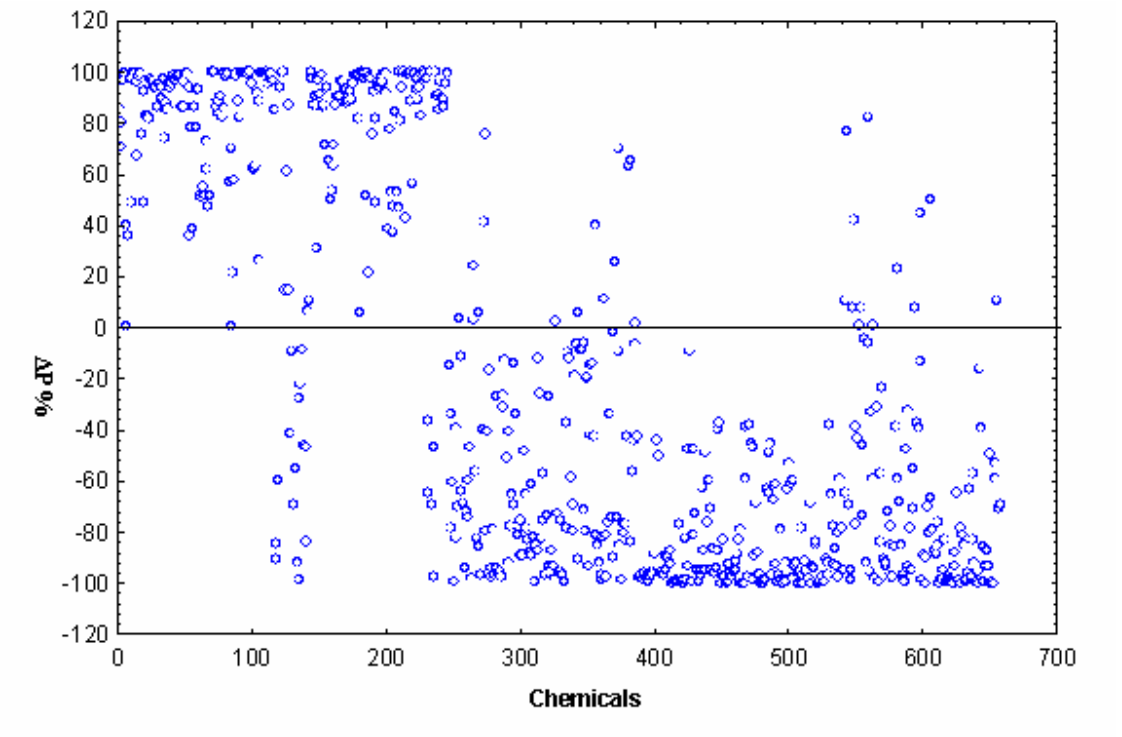

Figure 5. Plot of the ΔP% from Eq. **8** (using non-stochastic quadratic indices) for each compound in the training and test sets. Compounds 1-183 and 184-246 are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247-541 and 542-658 are inactive (non-tyrosinase inhibitors) in training and test sets, correspondingly.

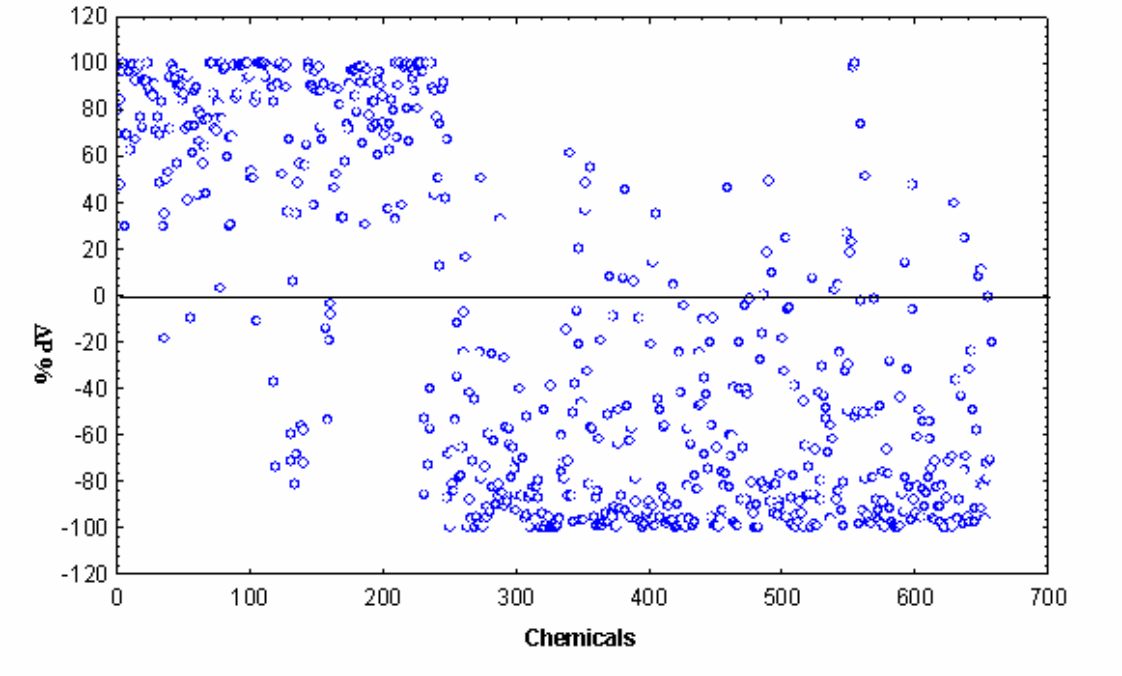

Figure 6. Plot of the ΔP% from Eq. **14** (using stochastic quadratic indices) for each compound in the training and test sets. Compounds 1-183 and 184-246 are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247-541 and 542-658 are inactive (non-tyrosinase inhibitors) in training and test sets, correspondingly.

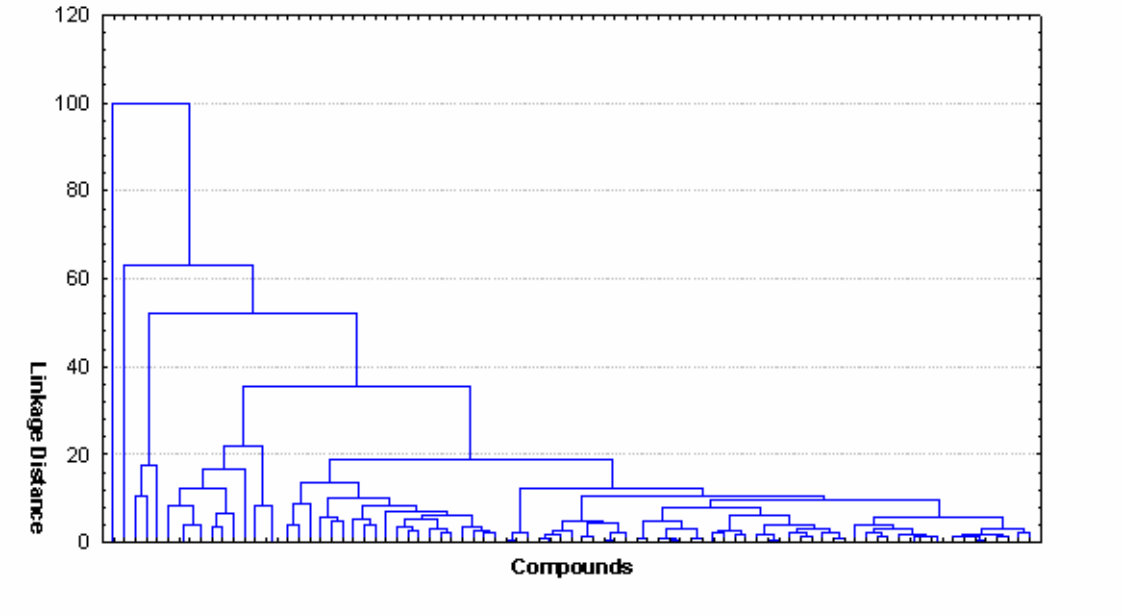

Figure 7. A dendrogram illustrating the results for the hierarchical *k*-NNCA of the set of active chemicals used for evaluating the predictive ability of the QSAR models for ligand-based virtual screening.

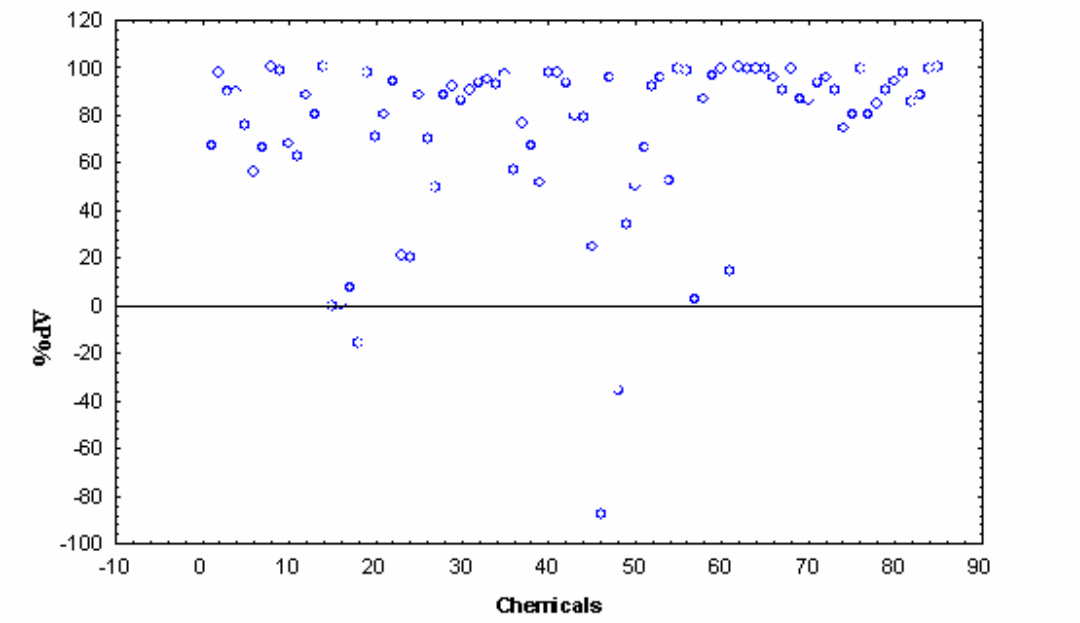

Figure 8. Plot of the ΔP% from Eq. **8** (using non-stochastic bond-based quadratic indices) for each compound selected in virtual screening protocols.

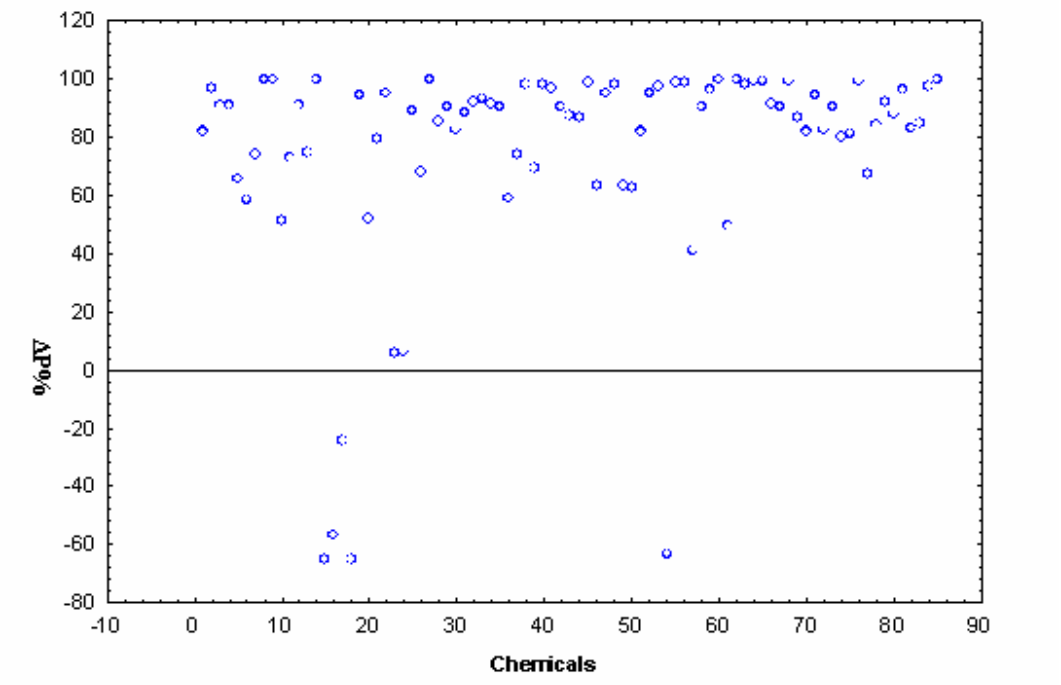

Figure 9. Plot of the ΔP% from Eq. **14** (using stochastic bond-based quadratic indices) for each compound selected in virtual screening protocols.

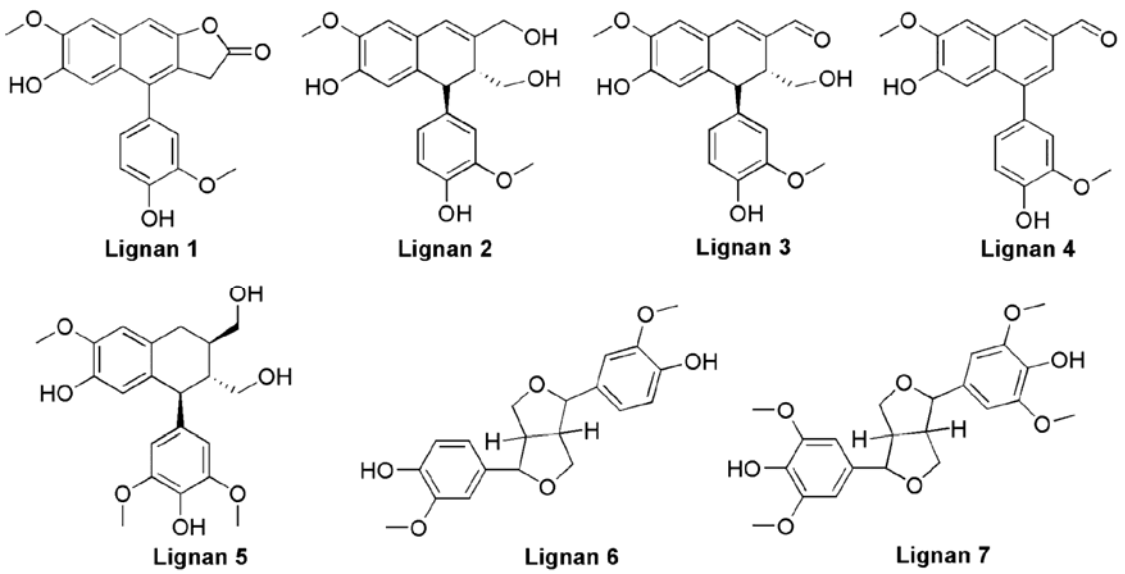

Figure 10. Molecular structure of the new lignans identified as hit via virtual screening.