3D MI-DRAGON: New Model for the Reconstruction of US FDA Drug-Target Network and Theoretical-Experimental Studies of Inhibitors of Rasagiline Derivatives for AChE

Francisco Prado-Prado[1,4,*], Xerardo García-Mera[1], Manuel Escobar[1], Nerea Alonso[1], Olga Caamaño[1], Matilde Yañez[2] and Humberto González-Díaz[3]

[1]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela (USC), 15782, Spain;* [2]*Department of Pharmacology, Faculty of Pharmacy, USC, 15782, Spain;* [3]*Department of Microbiology and Parasitology, Faculty of Pharmacy, USC, 15782, Spain;* [4]*Biomedical Sciences Department, Health Science Division, University of Quintana Roo, 77039, Chetumal, Mexico*

Abstract: The number of neurodegenerative diseases has been increasing in recent years. Many of the drug candidates to be used in the treatment of neurodegenerative diseases present specific 3D structural features. An important protein in this sense is the acetylcholinesterase (AChE), which is the target of many Alzheimer's dementia drugs. Consequently, the prediction of Drug-Protein Interactions (DPIs/nDPIs) between new drug candidates and specific 3D structure and targets is of major importance. To this end, we can use Quantitative Structure-Activity Relationships (QSAR) models to carry out a rational DPIs prediction. Unfortunately, many previous QSAR models developed to predict DPIs take into consideration only 2D structural information and codify the activity against only one target. To solve this problem we can develop some 3D multi-target QSAR (3D mt-QSAR) models. In this study, using the 3D MI-DRAGON technique, we have introduced a new predictor for DPIs based on two different well-known software. We have used the MARCH-INSIDE (MI) and DRAGON software to calculate 3D structural parameters for drugs and targets respectively. Both classes of 3D parameters were used as input to train Artificial Neuronal Network (ANN) algorithms using as benchmark dataset the complex network (CN) made up of all DPIs between US FDA approved drugs and their targets. The entire dataset was downloaded from the DrugBank database. The best 3D mt-QSAR predictor found was an ANN of Multi-Layer Perceptron-type (MLP) with profile MLP 37:37-24-1:1. This MLP classifies correctly 274 out of 321 DPIs (Sensitivity = 85.35%) and 1041 out of 1190 nDPIs (Specificity = 87.48%), corresponding to training Accuracy = 87.03%. We have validated the model with external predicting series with Sensitivity = 84.16% (542/644 DPIs; Specificity = 87.51% (2039/2330 nDPIs) and Accuracy = 86.78%. The new CNs of DPIs reconstructed from US FDA can be used to explore large DPI databases in order to discover both new drugs and/or targets. We have carried out some theoretical-experimental studies to illustrate the practical use of 3D MI-DRAGON. First, we have reported the prediction and pharmacological assay of 22 different rasagiline derivatives with possible AChE inhibitory activity. In this work, we have reviewed different computational studies on Drug-Protein models. First, we have reviewed 10 studies on DP computational models. Next, we have reviewed 2D QSAR, 3D QSAR, CoMFA, CoMSIA and Docking with different compounds to find Drug-Protein QSAR models. Last, we have developped a 3D multi-target QSAR (3D mt-QSAR) models for the prediction of the activity of new compounds against different targets or the discovery of new targets.

Keywords: Drug-protein interaction complex networks, protein structure networks, multi-target QSAR, markov model, AChE inhibitors.

INTRODUCTION

Yildirim *et al.* [1] have built a complex network (CN) of Drug-Protein Pairs (DPIs) as a bipartite graph made up of all DPIs for all US Food and Drug Administration (US FDA) approved drugs and proteins linked by drug-target binary associations. The resulting CN connects most drugs into a highly interlinked giant component, with strong local clustering of drugs of similar types according to Anatomical Therapeutic Chemical classification. It was motivated due to the strong incentive to develop new methods able to predict potential drug-target interactions of complex networks (CNs) made up of DPIs [2]. To this end, we can use Quantitative Structure-Activity Relationships (QSAR) models [3] to carry out a DPIs prediction. To solve this problem, we can develop some3D multi-target QSAR (3D mt-QSAR) models to predict DPIs [4]. One way to develop this class of mt-QSAR is incorporating into the QSAR equation parameters of the structure of the target (protein, DNA, RNA, etc.) in addition to the structural parameters of the drug present in classic QSAR. Some of the most widely known software we can use to reach this goal are: DRAGON, CODESSA[5], MODES-LAB[6], TOMO-COMD[7] and MARCH-INSIDE (MI)[8]. The DRAGON software is one of the most complete one,

*Address correspondence to this author at the Faculty of Pharmacy, University of Santiago de Compostela 15782, Spain; Fax: (+34) 981594912; E-mail: francisco.prado@usc.es

since it can calculate more than 1600 descriptors for drug structure including as zero- (0D) one- (1D), two- (2D), three-dimensional (3D) parameters.

Unfortunately, several QSAR models are able to predict the activity of drugs against only one target and/or are unable to codify important 3D structural features. Speck-Planche *et al.* [9, 10] have developed mt-QSAR for the design of multi-target inhibitors against chemokine receptors. This approach was focused on the construction of an mt-QSAR model for the classification and prediction of inhibitors of chemokine receptors. For instance, in a previous work we have recently developed a QSAR model based on the MARCH-INSIDE method to predict a large network of DTPs [11]. This model has been based on 2D structural parameters for drugs and 1D-structural parameters for proteins. Next, we have developed MIND-BEST [12] and NL MIND-BEST [13]. Both predictors are based on 3D structural parameters of proteins calculated with the MI software but they used only 2D structural parameters of drugs (calculated also with MI). The accuracy of the MIND-BEST model found was 86.32% and the one found with NL MIND-BEST was 90.41%. However, both models used only 2D parameters with the MI software. After that, to improve the obtained results we have used the MARCH-INSIDE (MI) software to calculate 3D structural parameters for targets and the DRAGON software to calculate 2D molecular descriptors for all drugs [14]. Using the 2D MI-DRAGON technique, we have introduced a new predictor for DPIs based on two different well-known software.

On the other hand, QSAR models can be used to explore the relationships between the structural spaces of compounds as inhibitors for specific enzymes, such as MAO inhibitors [15], HIV-1 integrase inhibitors [16], and/or protease inhibitors [17] or tyrosinase inhibitors [18-20]. In fact, almost all QSAR techniques are based on the use of molecular descriptors, which are numerical series that codify useful chemical information and enable correlations between statistical and biological properties [21, 22]. Recently, the field has moved from small molecules to proteins and other systems. For instance, González-Díaz *et al.* discussed the use of these methods, but only from the point of view of proteins [23]. Later, some groups published different papers in one special issue on QSAR, but also restricted to the field of protein and proteomics [24-30]. In other recent issue, guest-edited by González-Díaz [31] a series of papers devoted to QSAR/QSPR techniques for low-molecular-weight drugs [5, 31-39] was published. More recently, Prado-Prado *et al.* [40] have published an mt-QSAR for anti-parasitic drugs. This year has been published another issue [41] focused on QSAR/QSPR models and graph theory used to approach Drug ADMET processes and Metabolomics [42-49]. Last, one of the most recent issues published has been devoted to discuss the applications of QSAR in Pharmaceutical design [50-59]. In the present work, we have firstly reviewed the state-of-the-art on the design. We have reviewed previous works based on 2D-QSAR, 3D-QSAR, CoMFA, CoMSIA and Docking techniques, which studied different compounds to find out the structural requirements. Last, we have carried out new QSAR studies using Linear Discriminant Analysis (LDA) and compared them with Artificial Neural Network (ANN) methods using two well-known software: MARCH INSIDE and DRAGON in order to understand the essential structural

requirement for binding with the receptor for *AChE*. The topics reviewed, discussed, and/or reported in this paper are:

1. Review of Drug-target models

   1.1. MIND-BEST: Web server for drugs and target discovery

   1.2. NL MIND-BEST: a web server for ligands and proteins discovery

   1.3. 2D MI-DRAGON: a new predictor for protein-ligand interactions

   1.4. Does drug-target have a likeness?

   1.5. QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR

   1.6. Supervised prediction of drug-target interactions using bipartite local models

   1.7. Very fast prediction and rationalization of pKa values for protein-ligand complexes

   1.8. Prediction of potential drug targets based on simple sequence properties

   1.9. Importance of ligand reorganization free energy in protein-ligand prediction

   1.10. Alignment-free prediction of a drug-target complex network

2. 3D MI-DRAGON: new model for the reconstruction of US FDA drug-target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE.

   2.1. Methods

   2.2. Illustrative experiments

3. Results and Discussion

   3.1. DPI QSAR predictive models

   3.2. A Theoretical-Experimental Study using the 3D MI-DRAGON predictor

4. Conclusion

As mentioned in the previous paragraph, we can seek a QSAR predictor for DPIs using molecular descriptors of both drug and target. In this work, we have introduced for the first time 3D MI-DRAGON a new predictor for DPIs based on two different well-known software. We use the MARCH-INSIDE (MI) software to calculate 3D structural parameters for targets and the DRAGON software for 3D parameters of all DPIs present in the DrugBank database (US FDA benchmark dataset) [60-63]. Both classes of parameters were used as input of different Artificial Neuronal Network (ANN) algorithms to seek an accurate non-linear mt-QSAR predictor. 3D MI-DRAGON offers a good opportunity for fast-track calculation of all possible DPIs of one drug enabling us to re-construct large drug-target or DPI Complex Networks (CNs). In this study, we report the prediction and pharmacological assay of 22 different rasagiline derivatives with AChE inhibitory activity. The present work reports the attempts to calculate within the unified DPIs. All this can help to design new inhibitors of AChE. A very good 3D MI-DRAGON QSAR model was obtained, and the sub-

sequent combined QSAR & CN analysis may become of major importance for the prediction of the activity of new compounds against different targets or the discovery of new targets. In this sense, we have reported an illustrative study that combines both experiment and theory to show how to use this model in practical situations. We have reported the prediction and pharmacological assay of rasagiline derivatives with AChE inhibitory activity. In Fig. (**1**) we depict a flowchart with the main steps given in this work to train and validate the ANN classifier.

1. REVIEW OF DRUG-TARGET MODELS

1.1. MIND-BEST: Web Server for Drugs and Target Discovery

In this work, González-Díaz, H. *et al.*, [12] selected drug-target pairs (DTPs/nDTPs) of drugs with high affinity/non-affinity for different targets. Quantitative structure-activity relationship (QSAR) models have become a very useful tool in this context because they substantially reduce time- and resource-consuming experiments. Unfortunately, most QSAR models predict activity against only one protein target and/or they have not been implemented on a public Web server yet, freely available online to the scientific community. To solve this problem, they developed a multitarget QSAR (mt-QSAR) classifier combining the MARCH-INSIDE software for the calculation of the structural parameters of drug and target with the linear discriminant analysis (LDA) method in order to seek the best model. The accuracy of the best LDA model was 94.4% (3,859/4,086 cases) for training and 94.9% (1,909/2,012 cases) for the external validation series. In addition, they implemented the model into the Web portal Bio-AIMS as an online server entitled MARCH-INSIDE Nested Drug-Bank Exploration & Screening Tool (MIND-BEST), located at http://miaja.tic.udc.es/Bio-AIMS/MIND-BEST.php, see Fig. (**2**). This online tool is based on PHP/HTML/Python and MARCH-INSIDE routines. Finally, they illustrated two practical uses of this server with two different experiments. In experiment 1, the authors reported for the first time a MIND-BEST prediction, synthesis, characterization, and MAO-A and MAO-B pharmacological assay of eight rasagiline derivatives, promising for anti-Parkinson drug design. In experiment 2, they reported sampling, parasite culture, sample preparation, 2-DE, MALDI-TOF and -TOF/TOF MS, MASCOT search, 3D structure modeling with LOMETS, and MIND-BEST prediction for different peptides as new proteins of the ones found in the proteome of the bird parasite Trichomonas gallinae, which is promising for the discovery of antiparasitic drug targets.

1.2. NL MIND-BEST: A Web Server for Ligands and Proteins Discovery

There are many protein ligands and/or drugs described with very different affinity to a large number of target proteins or receptors. In this work, González-Díaz, H., *et al.* [13] selected ligands or drug-target pairs (DTPs/nDTPs) of drugs with high affinity/non-affinity for different targets. QSAR models have become a very useful tool in this context to substantially reduce time- and resource- consuming experiments. Unfortunately, most QSAR models predict activ-

ity against only one protein target and/or have not been implemented as a public web server freely accessible online to the scientific community. To solve this problem, the authors have developed herein a multi-target QSAR (mt-QSAR) classifier using the MARCH-INSIDE technique to calculate structural parameters of drug and target in addition to an Artificial Neuronal Network (ANN) to seek the model. The best ANN model found is a Multi-Layer Perceptron (MLP) with profile MLP 20:20-15-1:1. This MLP classifies correctly 611 out of 678 DTPs (Sensitivity=90.12%) and 3083 out of 3408 nDTPs (Specificity=90.46%), corresponding to the training series Accuracy=90.41%. The validation of the model was carried out by means of external predicting series. The model classifies correctly 310 out of 338 DTPs (Sensitivity=91.72%) and 1527 out of 1674 nDTP (Specificity=91.22%) in the validation series, corresponding to total Accuracy=91.30% for the validation series (predictability). This model favorably compares with other ANN models developed in this work and Machine Learning classifiers published before to address the same problem in different aspects. They implemented the present model at web portal Bio-AIMS as an online server called: Non-Linear MARCH-INSIDE Nested Drug-Bank Exploration & Screening Tool (NL MIND-BEST), which is located at URL: http://miaja.tic.udc.es/Bio-AIMS/NL-MIND-BEST.php, see Fig. (**3**). This online tool is based on PHP/HTML/Python and MARCH-INSIDE routines. Finally, this predictor illustrated two practical uses of this server with two different experiments. In experiment 1, they reported for the first time a Quantum QSAR study, synthesis, characterization, and experimental assay of antiplasmodial and cytotoxic activities of oxoisoaporphine alkaloids derivatives as well as NL MIND-BEST prediction of potential target proteins. In experiment 2, the authors reported sampling, parasite culture, sample preparation, 2-DE, MALDI-TOF, and -TOF/TOF MS, MASCOT search, MM/MD 3D structure modeling, and NL MIND-BEST prediction for different peptides in a new protein of the ones found in the proteome of the human parasite Giardia lamblia, which is promising for the discovery of antiparasitic drug targets.

1.3. 2D MI-DRAGON: A New Predictor for Protein-Ligand Interactions

There are many pairs of possible Drug-Protein Interactions that may take place or not (DPIs/nDPIs) between drugs with high affinity/non-affinity for different proteins. Consequently, for instance, the determination of all possible ligand-protein interactions for a single drug is expensive in terms of time and resources. In this sense, Prado-Prado, F., *et al.*, [14] can use QSAR models to carry out a rational DPIs prediction. Unfortunately, almost all QSAR models predict activity against only one target. To solve this problem we can develop multi-target QSAR (mt-QSAR) models. In this work, using the 2D MI-DRAGON technique, they introduced a new predictor for DPIs based on two different well-known software. We used the MARCH-INSIDE (MI) software to calculate 3D structural parameters for targets and the DRAGON software was used to calculate 2D molecular descriptors of all drugs which showed known DPIs present in the DrugBank database (US FDA benchmark dataset). Both classes of parameters were used as input for different Artifi-
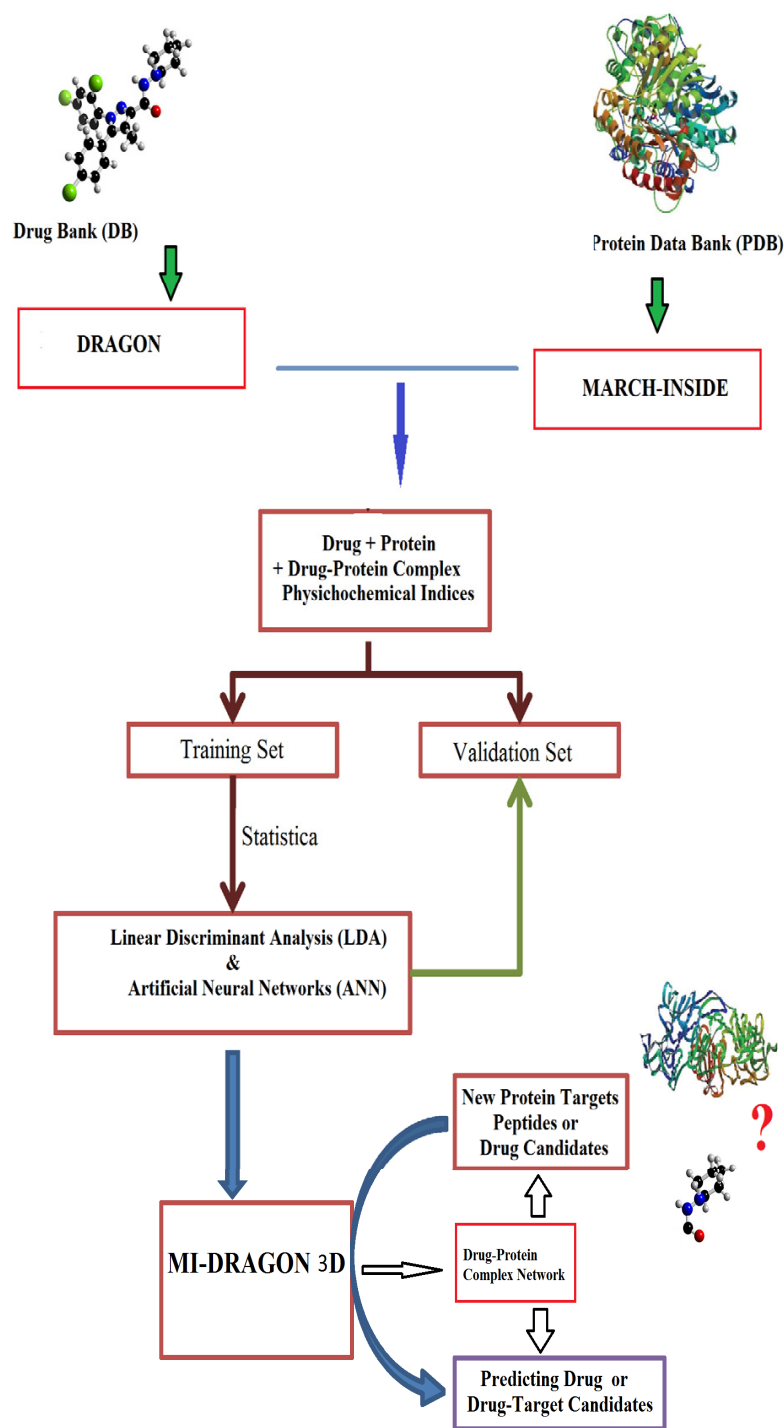
Fig. (1). Flowchart of all steps given in this work to develop the new model.

cial Neural Network (ANN) algorithms to seek an accurate non-linear mt-QSAR predictor; see Fig. (**4**) for the flowchart of all steps given in this work to develop the new model 2D MI-DRAGON. The best ANN model found is a Multi-Layer Perceptron (MLP) with profile MLP 21:21-31-1:1. This MLP classifies correctly 303 out of 339 DPIs (Sensitivity = 89.38%) and 480 out of 510 nDPIs (Specificity = 94.12%), corresponding to training Accuracy = 92.23%. The valida-

tion of the model was carried out by means of external predicting series with Sensitivity = 92.18% (625/678 DPIs; Specificity = 90.12% (730/780 nDPIs) and Accuracy = 91.06%. 2D MI-DRAGON offers a good opportunity for fast-track calculation of all possible DPIs of one drug, enabling us to re-construct large drug-target or DPI Complex Networks (CNs). For instance, the authors reconstructed the CN of the US FDA benchmark dataset with 855 nodes 519

Fig. (2). Bio-AIMS online server entitled MARCH-INSIDE Nested Drug-Bank Exploration & Screening Tool (MIND-BEST).

Fig. (3). Bio-AIMS as an online server called: Non-Linear MARCH-INSIDE Nested Drug-Bank Exploration & Screening Tool (NL MIND-BEST).
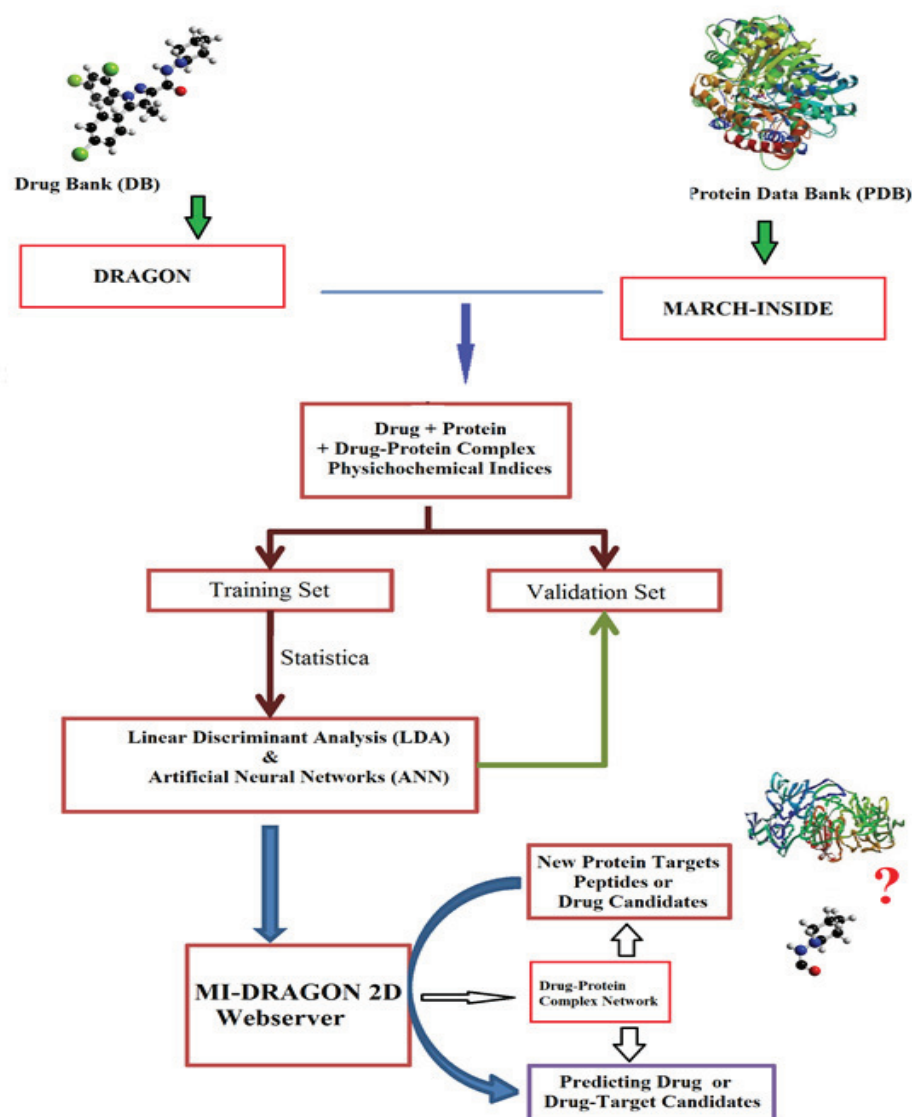
Fig. (4). Flowchart of all steps to develop the new model 2D MI-DRAGON.

drugs+336 targets. We predicted CN with similar topology (observed and predicted values of average distance are equal to 6.7 vs. 6.6). These CNs can be used to explore large DPI databases in order to discover both new drugs and/or targets. Finally, Prado-Prado, F., *et al.,*[14] illustrated in a theoretical-experimental study the practical use of 2D MI-DRAGON. They reported the prediction, synthesis, and pharmacological assay of 10 different oxoisoaporphines with MAO-A inhibitory activity. The most active compound OXO_5 presented, $IC_{(50)} = 0.00083$ muM, was notably better than the control drug Clorgyline.

1.4. Does Drug-Target have a Likeness?

The discovery of new targets that are sufficiently robust to yield marketable therapeutics is an enormous challenge. Conventional target identification approaches are disease-dependent, which require heavy experimental workload and comprehensive domain knowledge. In this work, Chen, X.,

et al.,[64] proposed that a disease-independent property of proteins, "drug-target likeness", could be explored to facilitate the genomic scale target screening in the post-genomic age. A Support Vector Machine (SVM) classifier was trained to recognize target and non-target protein sequences compiled from the Therapeutic Target Database, DrugBank, and PFam. Protein sequences are encoded by their composition, transition and distribution features of residues and Gaussian kernel function was used in SVM classification. RESULTS: SVM with a fine-tuned kernel width records 66.4 +/- 5.1% of sensitivity and 97.2 +/- 0.6% of specificity, corresponding to an overall target prediction accuracy of 94.4 +/- 0.8%. The authors concluded that despite sounding primitive, these results suggested that, similar to the "drug likeness" for small chemicals, their binding partners, drug targets, also displayed shared features which were reflected in their sequences and could be captured by statistical learning approaches. Further research on how to accurately and interpretably measure the likeness of protein being a drug target is promising. Inspired

by the progress of "drug likeness" studies, advances in protein descriptors, statistical learning algorithms and more comprehensive and accurate gold-standard data set from disease biology research may help to further define the "drug-target likeness" property of proteins.

1.5. QSAR Study of Malonyl-CoA Decarboxylase Inhibitors Using GA-MLR

Quantitative structure-activity relationships (QSAR) of a series of structural diverse malonyl-CoA decarboxylase (MCD) inhibitors have been investigated by Li, J. *et al.,*[65] using the predictive single model as well as the consensus analysis based on a new strategy proposed by them. A self-organizing map (SOM) neural network was employed to divide the whole data set into a representative training set and a test set. Then a multiple linear regression (MLR) model population was built based on the theoretical molecular descriptors selected by the Genetic Algorithm using the training set. In order to analyze the diversity of these models, multidimensional scaling (MDS) was employed to explore the model space based on the Hamming distance matrix calculated from each two models. In this space, $Q_{(2)}$ (cross-validated $R_{(2)}$) guided model selection (QGMS) strategy was performed to select submodels. Then, consensus modeling was built by two strategies, average consensus model (ACM) and weighted consensus model (WCM), where each submodel had a different weight according to the contribution of the model expressed by MLR regression coefficients. The authors obtained results which prove that QGMS is a reliable and practical method to guide the submodel selection in consensus with modeling building and their weighted consensus model (WCM) strategy is superior to the simple ACM.

1.6. Supervised Prediction of Drug-Target Interactions Using Bipartite Local Models

In silico prediction of drug-target interactions from heterogeneous biological data is critical in the search for drugs for known diseases. This problem is currently being attacked from many different points of view, a strong indication of its current importance. Precisely, being able to predict new drug-target interactions with both high precision and accuracy is the holy grail, a fundamental requirement for *in silico* methods to be useful in a biological setting. This, however, remains extremely challenging due to, amongst other things, the rarity of known drug-target interactions. Bleakley, K. *et al.,*[66] propose a novel supervised inference method to predict unknown drug-target interactions, represented as a bipartite graph. They used this method, known as bipartite local models to first predict target proteins of a given drug, then to predict drugs targeting a given protein. This gives two independent predictions for each putative drug-target interaction, which they showed it could be combined to give a definitive prediction for each interaction. The authors demonstrated the excellent performance of the proposed method in the prediction of four classes of drug-target interaction networks involving enzymes, ion channels, G protein-coupled receptors (GPCRs) and nuclear receptors in human, see Fig. (**5**). This enables us to suggest a number of new potential drug-target interactions.

1.7. Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes

Bas, D. C. *et al.,*[67] developed the PROPKA method for the prediction of the pK(a) values of ionizable residues in proteins, which is extended to include the effect of non-
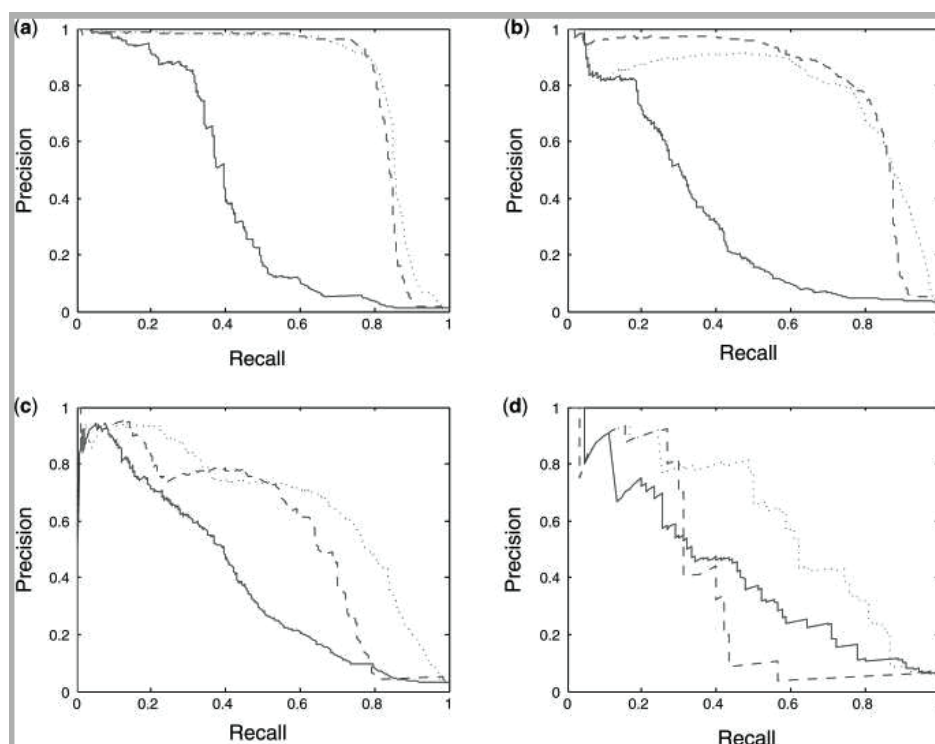

Fig. (5). PR curves for the predicted drug–target interactions using BLMs on four benchmark datasets: (**a**) enzyme, (**b**) ion channel, (**c**) GPCR and (**d**) nuclear receptor.

proteinaceous ligands on protein pK(a) values as well as to predict the change in pK(a) values of ionizable groups on the ligand itself. The authors developed this new version of PROPKA (PROPKA 2.0) see Fig. (**6**), which was, as much as possible, carried out by adapting the empirical rules underlying PROPKA 1.0 to ligand functional groups. Thus, the speed of PROPKA is retained, so that the pK(a) values of all ionizable groups are computed in a matter of seconds for most proteins. This adaptation is validated by comparing PROPKA 2.0 predictions to experimental data for 26 protein-ligand complexes including trypsin, thrombin, three pepsins, HIV-1 protease, chymotrypsin, xylanase, hydroxynitrile lyase, and dihydrofolate reductase. The protonation state changes due to plasmepsin II, cathepsin D and endothiapepsin binding to pepstatin are predicted to within 0.4 proton units at pH 6.5 and 7.0, respectively. The PROPKA 2.0 results indicate that structural changes due to ligand binding contribute significantly to the proton uptake/release, as do residues far away from the binding site, primarily due to the change in the local environment of a particular residue and hence the change in the local hydrogen bonding network. Overall, the results suggest that PROPKA 2.0 provides a good description of the protein-ligand interactions that have an important effect on the pK(a) values of titratable groups, thereby permitting fast and accurate determination of the protonation states of key residues and ligand functional groups within the binding or active site of a protein.
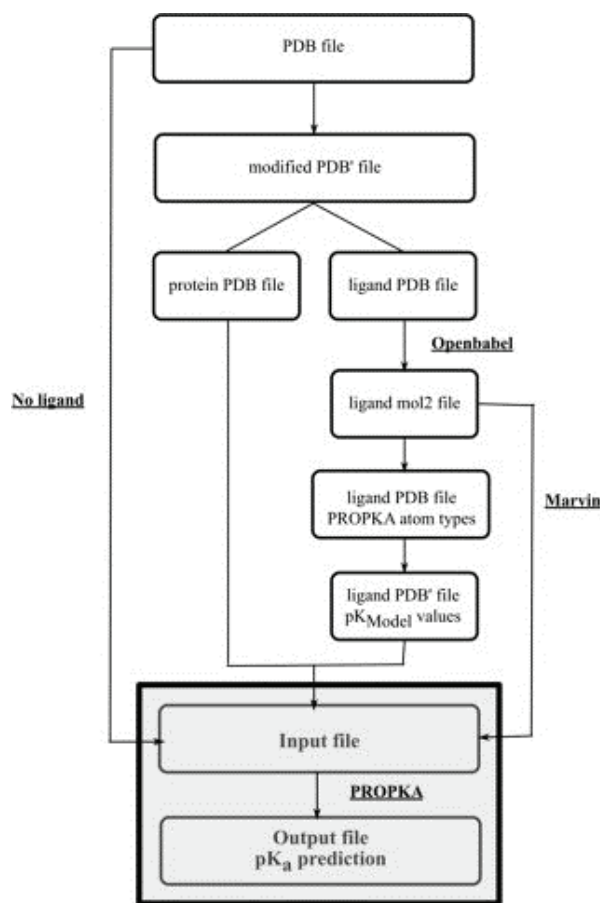

Fig. (6). Chart outlining the main steps in a PROPKA 2.0 calculation. See text for further details.

1.8. Prediction of Potential Drug Targets Based on Simple Sequence Properties

During the past decades, research and development in drug discovery have attracted much attention and efforts. However, only 324 drug targets are known for clinical drugs up to now. Identifying potential drug targets is the first step in the process of modern drug discovery for developing novel therapeutic agents. Therefore, the identification and validation of new and effective drug targets are of great value for drug discovery in both academia and pharmaceutical industry. If a protein can be predicted in advance for its potential application as a drug target, the drug discovery process targeting this protein will be greatly speeded up. Li, Q. *et al.,*[68] in their study based on the properties of known drug targets, developed a sequence-based drug target prediction method for fast identification of novel drug targets. The authors' results were based on simple physicochemical properties extracted from protein sequences of known drug targets, since several support vector machine models were constructed in this study. The best model could distinguish known drug targets from non drug targets at an accuracy of 84%. Using this model, potential protein drug targets of human origin from Swiss-Prot were predicted, some of which had already attracted much attention as potential drug targets in pharmaceutical research. They developed a drug target prediction method based solely on protein sequence information without the knowledge of family/domain annotation, or the protein 3D structure. This method can be applied in novel drug target identification and validation, as well as genome scale drug target predictions.

1.9. Importance of Ligand Reorganization Free Energy in Protein-Ligand Prediction

Accurate prediction of the binding affinities of small-molecule ligands to their biological targets is fundamental for structure-based drug design but remains a very challenging task. In their paper, Yang, C. Y. *et al.*, [69] performed computational studies to predict the binding models of 31 small-molecule Smac (the second mitochondria-derived activator of caspase) mimetics to their target, the XIAP (X-linked inhibitor of apoptosis) protein, and their binding affinities. Their results showed that computational docking was able to reliably predict the binding models, as confirmed by experimentally determined crystal structures of some Smac mimetics complexed with XIAP. However, all the computational methods they tested, including an empirical scoring function, two knowledge-based scoring functions, and MM-GBSA (molecular mechanics and generalized Born surface area), yielded poor to modest prediction for binding affinities. The linear correlation coefficient ($r_{(2)}$) value between the predicted affinities and the experimentally determined affinities was found to be between 0.21 and 0.36. The inclusion of ensemble protein-ligand conformations obtained from molecular dynamic simulations did not significantly improve the prediction. However, major improvement was achieved when the free-energy change for ligands between their free- and bound-states, or "ligand-reorganization free energy", was included in the MM-GBSA calculation, and the $r_{(2)}$ value increased from 0.36 to 0.66. The prediction was validated using 10 additional Smac mimetics designed and evaluated by an independent group. This study demonstrates

that ligand reorganization free energy plays an important role in the overall binding free energy between Smac mimetics and XIAP. This term should be evaluated for other ligand-protein systems and included in the development of new scoring functions. To the best of the authors' knowledge, this is the first computational study to demonstrate the importance of ligand reorganization free energy for the prediction of protein-ligand binding free energy.

1.10. Alignment-Free Prediction of a Drug-Target Complex Network

There are many drugs described with very different affinity to a large number of receptors. In this work, Viña, D. *et al.*, [11] selected drug-receptor pairs (DRPs) of affinity/non-affinity drugs to similar/dissimilar receptors and represented them as a large network, which may be used to identify drugs that can act on a receptor. Computational chemistry prediction of the biological activity based on QSAR substantially increases the potentialities of this kind of networks avoiding time- and resource-consuming experiments. Unfortunately, most QSAR models are unspecific or predict activity against only one receptor. To solve this problem, the authors developed an mt-QSAR classification model, see Fig. (**1**). Overall model classification accuracy was 72.25% (1390/1924 compounds) in training, 72.28% (459/635) in cross-validation. The outputs of this mt-QSAR model were used as inputs to construct a network. The observed network has 1735 nodes (DRPs), 1754 edges or pairs of DRPs with similar drug-target affinity (sPDRPs), and low coverage density d = 0.12%. The predicted network has 1735 DRPs, 1857 sPDRPs, and also low coverage density d = 0.12%. After an edge-to-edge comparison (chi-square = 9420.3; p < 0.005), the authors demonstrated that the predicted network was significantly similar to the one observed and both had a distribution closer to exponential than to normal.

2. MATERIALS AND METHODS

In this work, using the 3D MI-DRAGON technique, we have introduced for the first time a new predictor for DPIs based on two different well-known software. We have used the MARCH-INSIDE (MI)software to calculate 3D structural parameters for targets and the DRAGON software for 3D parameters of all DPIs present in the DrugBank database (US FDA benchmark dataset) [60-63]. Both classes of parameters were used as input of different Artificial Neuronal Network (ANN) algorithms to seek an accurate non-linear mt-QSAR predictor. 3D MI-DRAGON offers a good opportunity for fast-track calculation of all possible DPIs of one drug, enabling us to re-construct large drug-target or DPIs Complex Networks (CNs). In this study, we have reported the prediction and pharmacological assay of 22 different rasagiline derivatives with AChE inhibitory activity. The present work reports the attempts to calculate within unified DPIs. All this can help to design new inhibitors of AChE. A very good 3D MI-DRAGON QSAR model was obtained, and the subsequent combined QSAR & CN analysis may become of major importance for the prediction of the activity of new compounds against different targets or the discovery of new targets. In this sense, we have reported an illustrative study that combines both experiment and theory to show

how to use this model in practical situations. We have reported the prediction and pharmacological assay of rasagiline derivatives with AChE inhibitory activity.

2.1. Computational Methods

2.1.1. MOPAC AM1 Optimization Geometry Method Using CS CHEM 3D

Molecular structures of all FDA drugs were generated with CHEM 3D Ultra (version 2005). The energy of each intermediate was then minimized using the semi-empirical MOPAC method with a minimum RMS gradient of 0.100, which specifies the convergence criteria for the gradient of the potential energy surface. The geometry of the molecules was optimized and the values of the quantum chemical descriptors of each compound were calculated using AM1. AM1 theory was used with a closed shell function. The MOPAC AM1 method was selected because it was a semi-empirical quantum chemical method and the computational time was much shorter than that needed by the *ab initio* method.

2.1.2. MI-DRAGON Technique

3D Parameters for Drugs

The DRAGON software 4.0 [70] was used here to calculate the 3D parameters of drugs. It depends on whether they are computed from the chemical formula, substructure list representation, molecular graph or geometrical representation of the molecule, respectively [71, 72]. In this work, we have calculated only GETAWAY 3D descriptors. We have used these descriptors after optimizing them with 3D descriptors.

3D Parameters of Proteins

In previous works we have predicted the protein function based on different protein structural parameters derived from a Markov matrix that accounts for electrostatic interactions between aminoacid pairs in the 3D structure of the protein. One of the classes of parameters used was called the Shannon Entropy $^T\theta_k(R)$ of the Markov matrix. These values have been used here as inputs to describe information about the structure of the drug target proteins (T) in order to construct the mt-QSAR models for DTPs. The detailed explanation has been published before [73-81] and reviewed in detail more recently [33]. As follows, we provide the formula for $^T\theta_k(R)$ values and some general explanations:

$$^T\theta_k(R) = -\sum_{j\in R} {}^k p_j(R) \cdot \log\left[{}^k p_j(R)\right] \quad (1)$$

Where, $^k p_i(R)$ values are the absolute probabilities with which the effect of the electrostatic interaction propagates from the amino acid i^{th} to other amino acids j^{th} next to it and returns to i^{th} after k-steps. These probabilities refer to: amino acids considered isolated in the space (k = 0), interaction between amino acids in direct contact (k = 1) or spatial (k > 1) indirect interactions between amino acids placed at a distance equal to *k*-times the cut-off distance ($r_{ij} = k \cdot r_{cut-off}$) in the residue network. Euclidean 3D space $r_3 = (x, y, z)$ coordinates of the C_α atoms of amino acids listed in protein PDB files. For calculation, all water molecules and metal ions

were removed [23]. All calculations were carried out with our in-house MARCH-INSIDE 2.0 software [23]. For the calculation, the MARCH-INSIDE software always uses the full matrix, never a sub-matrix, but the last summation term may run either for all amino acids or only for some specific protein regions (R) denoted as: *c* for core, *i* for inner, *m* for middle, and *s* for surface regions, respectively). Consequently, we can calculate different $^T\theta_k(R)$ for the amino acids contained in the regions (*c*, *i*, *m*, *s*, or *t*) and placed at a topological distance *k* each other within this orbit (*k* is the order) [73, 74, 82-84]. In this work, we have calculated altogether 5(types of regions) x 6(orders considered) = 30 $^T\theta_k(R)$ indices for each protein.

2.1.3. Statistical Analysis

Let be $^D\theta_k(G)$ entropy-based molecular descriptors that codify information about drug structure and $^T\theta_k(R)$ entropy-based descriptors that codify information about drug target proteins; we attempt to develop a simple mt-QSAR model as a linear classifier with the general formula:

$$S(DTP)_{pred} = \sum_{k=0}^{5} a_{G,k} \cdot {}^D\theta_k(G) + \sum_{k=0}^{5} b_{R,k} \cdot {}^T\theta_k(R) + c_0 \quad (2)$$

We used the Linear Discriminating Analysis (LDA) to fit this discriminant function. The model deals with the classification of a compound set with or without affinity on different receptors. A dummy variable Affinity Class (AC) was used as input to codify the affinity. This variable indicates either high (AC = 1) or low (AC = 0) affinity of the drug by/according to the receptor. $S(DTP)_{pred}$ or DTP affinity predicted score is the output of the model and it is a continuous dimensionless score that sorts compounds from low to high affinity to the target coinciding DTPs with higher values of $S(DTP)_{pred}$ and nDTPs with lowest values. In equation (6), *b* represents the coefficients of the classification function, determined by the LDA module of the STATISTICA 6.0 software package [85]. We used the Forward Stepwise algorithm for a variable selection. The statistical significance of the LDA model was determined calculating the p-level (p) of error with the Chi-square test. We also inspected the Specificity, Sensitivity, and total Accuracy to determine the quality-of-fit to data in training. Cases for training set were selected at random out of the cases in full dataset. The remnant cases were used to validate the model. The validation of the model was corroborated with these external prediction series; these cases were never used to train the model. The ration between training/validation set was 2/1 approximately. This procedure to select training and validation sets is largely known and used to train QSAR models [86-92].

2.1.4. ANN Analysis

The non-linear mt-QSAR model was constructed using the ANN analysis. All models trained were carried out in STATISTICA 6.0 [85]. In so doing, we used a very simple type of ANN called Three Layers Perceptron (MLP-3) to fit this discriminant function. The model deals with the classification of a compound set with or without affinity on different receptors. A dummy variable Affinity Class (AC) was used as input to codify the affinity. This variable indicates either high (AC = 1) or low (AC = 0) affinity of the drug according to the receptor. $S(DTP)_{pred}$ or DTP affinity pre-

dicted score is the output of the model and it is a continuous dimensionless score that sorts compounds from low to high affinity to the target coinciding DTPs with higher values of $S(DTP)_{pred}$ and nDTPs with lowest values. In equation (2), *b* represents the coefficients of the LNN classification function, determined by the ANN module of the STATISTICA 6.0 software package [85]. We used the Forward Stepwise algorithm for a variable selection.

In addition, we can explore more complicated non-linear ANNs in order to improve the accuracy of the classifier. We processed our data with different ANNs looking for a better model. Four types of ANNs were used, namely, Probabilistic Neural Network (PNN), Radial Basic Function (RBF), Linear Neural Network (LNN), and Four Layer Perceptron (MLP-4)[40, 93]. The quality of all the ANNs (linear or non-linear) was determined calculating values of Specificity, Sensitivity, and total Accuracy to determine the quality-of-fit to data in training. The validation of the model was corroborated with external prediction series. It was also reported a ROC-curve analysis (ROC curve can be used to select an optimum decision) for both training and validation series [40, 94].

2.1.5. Data Set

The data set was made up of a set of marketed DPIs with known affinity of drugs by targets. This dataset is the same benchmark data used in previous works [1, 3, 12, 13] in this area and contains all drugs approved by the US FDA. We have downloaded this dataset from the public resource called DrugBank database [12, 13, 61-63]. The data set was made up of more than 519 drugs with their corresponding 336 targets. Subsequently, we were able to collect over 4485 cases (drug-protein interactions) instead of 519 x 336 cases. In addition, the data set was used to develop ANN models to perform the model. The names or codes for all compounds are depicted in Table 1SM of the supplementary material, due to space constraints, as well as the references consulted to compile the data in this table.

2.1.6. Complex Network Construction

We have developed a DPI network in order to achieve the drug and protein affinity with a network approach. Generally, in this network, a node may represent a drug or a target. On the other hand, the edges represent the DPIs; they express relationships between pairs of drugs with their targets [1]. Anyhow, the nodes representing targets may be of at least two types. In almost all cases reported up to date, each target is represented only once in the network. In this class of "static" DPI network, the target is depicted by the node corresponding to the X-ray structure of itself. In this work, in total we have developed two complex networks. First, we have developed the DPI networks for the observed data and secondly, the DPI network predicted by the model. The common steps to develop these networks are: first, using the Excel software in a column, we have introduced all the proteins, the drugs using quotation marks in our database. Then, in another column all the cases have been listed. The total number of vertices has been indicated at the beginning of this column, and there have been two columns of the name of drug and protein and their corresponding number of vertices. Next, bows have been placed at the end of the col-

umns in the first column, the number of vertices for the drug and in another column the one corresponding to the protein. Then, the file was saved as a .txt format file. After renaming the .txt file as a .net file we read it with the CentiBin software [95, 96]. Finally, using CentiBin we cannot only represent the network but also highlight all drugs and targets (nodes) connected by a specific edge or link (DPI). Using this software we can calculate vertex centralities to analyze the relationships between drug targets.

2.2. Illustrative Experiments

2.2.1. Synthesis of Rasagiline Derivatives

Synthesis. Synthesis of compounds **1-22** has been previously reported by us [3, 12], see Fig. (**7**).

2.2.2. Determinations of Cholinesterases Activities

Ellman's cholinesterase assay method was used to determine the *in vitro* cholinesterase activity [97]. The activity was measured by increasing in absorbance at 412 nm due to
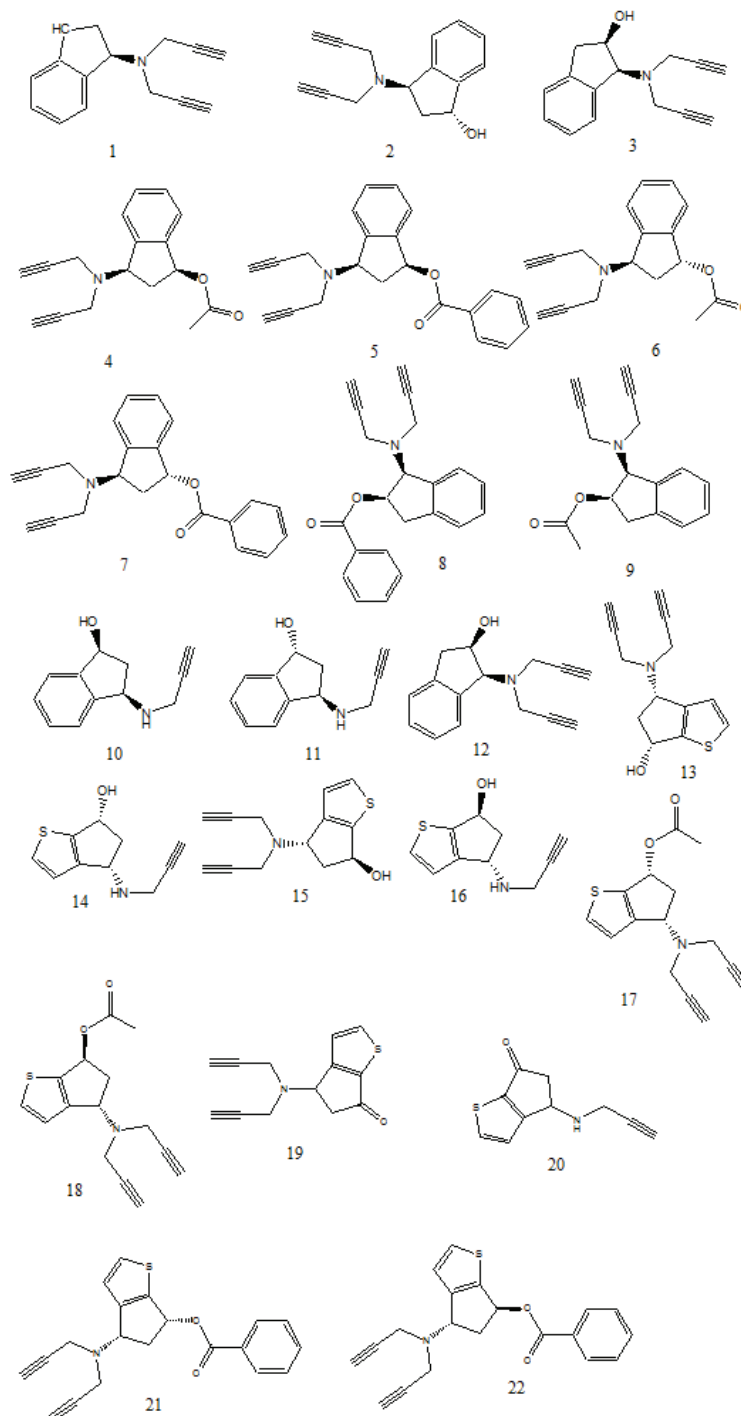

Fig. (7). Rasagiline derivatives used in this work.

the yellow color produced from the reaction of acetylthiocholine iodide with the dithiobisnitrobenzoate (DTNB) ion.

Acetylcholinesterase from human erythrocytes, recombinant acetylcholinesterase expressed in HEK 293 cells and butyrylcholinesterase from human serum were obtained from Sigma.

2.2.3. Experimental Conditions and Kinetics

Enzyme activity was measured using a FLUOstar Optima microplate reader. The assay medium contained phosphate buffer, pH 8.0, 20 mM DTNB, 0.01 U/ml of enzyme and 0.75 μM substrate (acetylthiocholine iodide or butyrylthiocholine iodide). The activity was determined by measuring the increase in absorbance at 412 nm at 1 min intervals for 10 min at 37 °C. In dose-dependent inhibition studies, the substrate was added to the assay medium containing enzyme, buffer, and DTNB with inhibitor after 10 min of incubation time. All experiments were carried out in duplicate and expressed as mean ± SEM. The relative activity is expressed as percentage ratio of enzyme activity in the absence of inhibitor, see (Table **1**).

3. RESULTS

3.1. DPIs QSAR Predictive Models

3.1.1. LDA Model

Common physicochemical properties, such as entropy, have been demonstrated to be useful on protein QSAR [98, 99]. We used these properties as input of our model in addition to drug molecular descriptors. The present model is the

first mt-QSAR one combining DRAGON and MI to predict the probability with which DPIs occur between a drug and a protein. This type of models lies within the frontiers between classic QSAR for drugs and protein QSAR [82]. Some applications for the present model include the prediction of new drugs, new protein receptors or drug targets, and drug binding sites. Detailed information on the compounds, predicted classification, and probability of affinity on different receptors of the drugs used to seek the model are shown in Table 1SM of the supplementary material. Based on the algorithms described in the Materials and Methods Section, the best linear model found was the following:

$$S(DTP)_{pred} = 11.01 \cdot d_1 - 36.37 \cdot d_2 - 12.77 \cdot d_3 - 9.34 \cdot d_4 - 52.17 \cdot d_5 + 1.62 \cdot d_6$$
$$- 1.65 \cdot d_7 - 0.11 \cdot d_8 - 0.10 \cdot d_9 + 0.25 d_{10} + 2.48 \qquad (3)$$
$$N = 4485 \qquad \chi^2 = 919.2988 \qquad p-level < 0.001$$

The nomenclature used in the descriptors of the equation is found in (Table **2**). In this equation, N is the number of cases, χ^2 is the Chi-square and p is the level of error. This model, with 10 variables, classifies correctly 256 out of 321 DPIs (Sensitivity = 79.75%) and 1014 out of 1190 nDPIs (Specificity = 85.21%). Overall training Accuracy was 84.05%. The validation of the model was carried out by means of external predicting series. The model classifies correctly 498 out of 644 DPIs (77.33%) and 2000 out of 2330 nDPIs (85.84%) in validation series. Accuracy for validation series (Predictability) was 83.99%. These results (Table **3**) indicate that we developed an accurate model according to previous reports on the use of LDA in QSAR [100, 101].

Table 1.   Inhibitory Activity of Different Rasagiline Derivatives

Compounds	hAChE (IC$_{50}$ μM)		hAChE (IC$_{50}$ μM)
1	>100 μM	12	>100 μM
2	>100 μM	13	No tested
3	>100 μM	14	No tested
4	No tested	15	No tested
5	**	16	No tested
6	No tested	17	**
7	**	18	**
8	No tested	19	**
9	**	20	**
10	>100 μM	21	**
11	>100 μM	22	**
Galantamine	1.43 ± 0.03[a]		
Eserine	151.40 ± 5.63 nM		
Tacrine	130,90 ± 6,83 nM		

Each IC$_{50}$ value is the mean ± S.E.M. from five experiments.

Table 2.   Detailed List of the Symbols and Description for All Parameters Present in the Model

Original Descriptor	Descriptor name	Code ID
H7v	H autocorrelation of lag 7 / weighted by atomic van der Waals volumes	d_1
HATS5v	leverage-weighted autocorrelation of lag 5 / weighted by atomic van der Waals volumes	d2
HATS4e	leverage-weighted autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	d3
HATS6e	leverage-weighted autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	d4
R5e+	R maximal autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	d5
$^T\theta_4(core)$	Entropy of all aminoacids placed in the core region and all the neighbors at distance $k \leq 4$	d6
$^T\theta_5(core)$	Entropy of all aminoacids placed in the core region and all the neighbors at distance $k \leq 5$	d7
$^T\theta_5(inner)$	Entropy of all aminoacids placed in the inner region and all the neighbors at distance $k \leq 5$	d8
$^T\theta_2(middle)$	Entropy of all aminoacids placed in the middle region and all the neighbors at distance $k \leq 2$	p1
$^T\theta_0(surface)$	Entropy of all aminoacids placed in the surface region and all the neighbors at distance $k \leq 0$	d9

Table 3.   Comparison of LDA and Different ANNs Classification Models

Model		Train	Stat.	Validation				
profile	Class	%	DPIs	nDPIs	Par.	%	DPIs	nDPIs
MI DRAGON 3D	DPIs	85.36	274	47	Sn	84.16	542	102
MLP	nDPIs	87.48	149	1041	Sp	87.51	291	2039
37:37-24-1:1	Total	87.03			Ac	86.79		
LDA[a]	DPIs	79.75	256	65	Sn	77.33	498	146
10:10-1:1	nDPIs	85.21	176	1014	Sp	85.84	330	2000
	Total	84.05			Ac	83.99		
PNN	DPIs	0	0	644	Sn	0	0	321
227:227-14797-2-2:1	nDPIs	100	0	2346	Sp	100	0	1174
	Total	78.46			Ac	78.53		
RBF	DPIs	47.05	303	341	Sn	52.65	169	152
1:1-1-1:1	nDPIs	56.01	1032	1314	Sp	54.86	530	644
	Total	54.08			Ac	54.38		
LNN	DPIs	53.73	346	298	Sn	45.79	147	174
227:227-1:1	nDPIs	32.05	1594	752	Sp	31.52	804	370
	Total	36.72			Ac	34.58		

DPIs: Drug-Target Pairs for compounds with high affinity; nDPIs: Drug-Target Pair for compounds with non-affinity; Stat. is statistics, Par. is parameter

3.1.2. 3D MI-DRAGON ANN Model

The previous model obtained good results with a relatively small number of parameters (10 parameters) and a linear equation. However, as result of the previous section we decided to carry out an ANN analysis to seek a better model using a non-linear method. Four types of ANNs were used, namely, Probabilistic Neural Network (PNN), Radial Basic Function (RBF), Three Layers Perceptron (MLP-3), and Four Layer Perceptron (MLP-4). See previous works about the use of these ANNs in protein QSAR [3, 13]. Fig. (**8**) depicts the network topology for some of the ANN models tested. In general, at least one ANN of every type tested was statically significant. However, one must note that the profiles of each network indicate that many of these are highly non-linear and complicated models.
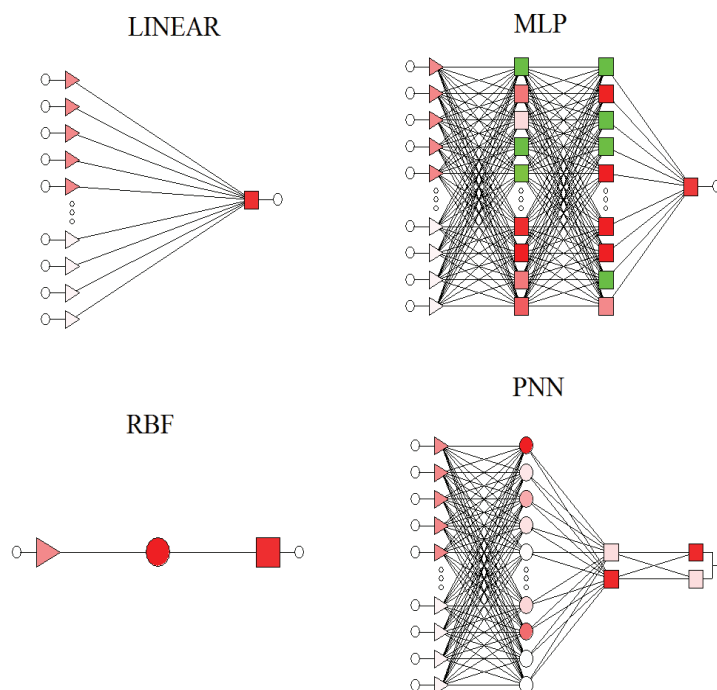
LINEAR

MLP

RBF

PNN

Fig. (8). Generic Topology of ANN models trained in this work.

Models using ANN-QSAR have been demonstrated before; see, for instance, the works carried out by Fernández and Caballero [102, 103]. We have compared different types of networks to obtain a better model. In (Table **3**) we show the classification matrix of the different networks. The profiles of networks tested were RBF 1:1-1-1:1 with only one variable; LNN 227:227-1:1, which presents many variables, and PNN 227:227-14797-2-2:1, which has a very high number of hidden neurons, see (Table **3**). Next, the simplest but most accurate ANN model found was an MLP (MLP 37:37-24-1:1) with training Accuracy = 87.03 %. This was selected as the best network found because it presents both high accuracy and an adequate number of variables accounting for features relevant for DPIs. This ANN presents 37 inputs variables ($24 d_k + 13 \Theta_m$). This leads to 37 neurons in first or input layer (I), 24 neurons in the second layer or first hidden layer (H1) and only one neuron (DPI prediction) in the output layer (O). We depict the ROC-curve for MLP 37:37-24-1:1 to show how reliable was the network model developed, in Fig. (**9**). Notably, the model presented had a ROC curve higher than 0.5. The model presented an area higher than 0.92. From now on, we call the ANN MLP 37:37-24-1:1 as the 3D MI DRAGON predictor.

Comparison with previous ML models. The ANN model shows excellent results with a relatively small number of parameters (only 37) with respect to some previously published Machine-Learning (ML) models. To assess the importance of this result, we have compared these ML models with other models used to address the same problem. For example, a notably more complicated ML model has been reported, which included a non-linear SVM model, a large number of parameters as well as many class-to-class trials rather than the single QSAR equation used in this work [64, 65]. All the other models included less than (20) input pa-

rameters or unknown parameters and some with 1000 or more (5000+) cases, and non-linear techniques such as Support vector Machine (SVM) and others [11, 66-69]. Our model is notably simpler and is the only developed in 3D parameters (drugs and proteins). However, some of these other models have low accuracy, or use the ROC curve or the Correlation coefficient as good classification, makong the task more difficult [104-110].

3.1.2. 1 3D MI-DRAGON Assembly of CNs for DPIs

The development of multi-protein CNs that incorporates protein affinity profile for drugs or the same CNs for DPIs is relevant to drug and target screening. And it is an application for this model. In order to recall the capacity of 3D MI-DRAGON to predict new CNs of DPIs, we have selected the same benchmark database used in previous works [3, 13, 14], which included US FDA approved drugs with their targets. With these goals in mind, we have developed again and manually curated the above-mentioned CN obtaining a graph with 855 vertices or nodes (drugs and proteins) and m = 1016 DPIs (edges). This CN of DPIs has D = 6.7 and average topological distances D_{ij} between all pairs of nodes. The same as before, we have developed a new CN of DPIs, but connecting only the pairs of nodes with DPIs predicted by 3D MI-DRAGON. In so doing, we have obtained a value of D = 7.2 and m = 1256 DPIs. In Fig. (**10**) we have illustrated visually both CNs (observed and predicted).

First, we have compares this predicted network (3D MI-DRAGON) with 2D MI-DRAGON predicted network [14]. We have compared to observe the similar or dissimilar topology (connectivity patterns structure) between them. Measuring in terms of TIs such as: number of nodes (n), number of edges (m), Wiener index (W), diameter (D), the Randic connectivity index (Xr), topological distance (Dist),
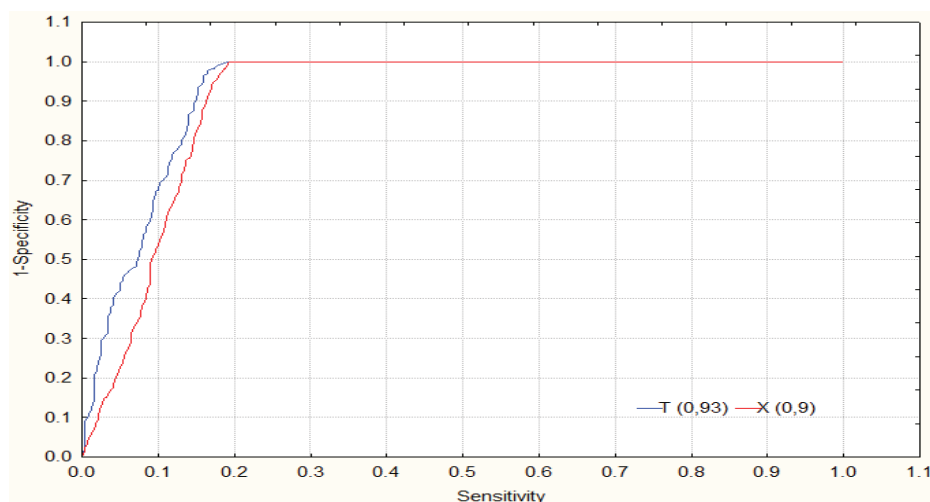
Fig. (9). ROC Curve for 3D MI-DRAGON predictor (red = train series, blue = validation series).
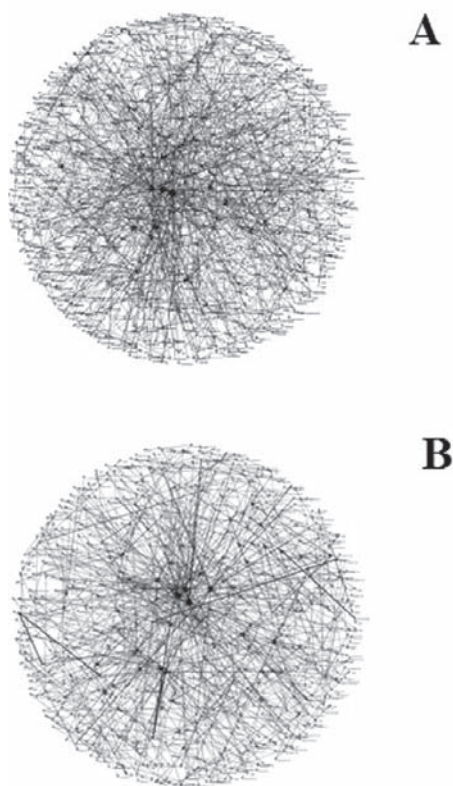

Fig. (10). Observed vs. Predicted drug-target complex networks.

network average values for radiality (R), node degree (δ), eccentricity (E). (Table **4**) shows that all the TIs are similar excepting n, m and w. That means both CNs have a high similarity between them. These results are very interesting, because our 3D MI-DRAGON model presents similar results to the 2D MI-DRAGON model, whose results have been published successfully before.

   To see how reliable and valid our model is, we have not only compared it to TIs to observe similarity between both predicted networks, but we have also studied the centrality

analysis of given networks. This type of drug screening and drug target discovery is the calculation of those nodes (drugs or proteins) which are more relevant or important (central) in the graph. To this end, we can use numerical parameters that quantify the importance of a node in a graph, which are called node centralities C_t of type t [111]. The identification of these nodes using node centralities may help us to identify the most relevant drugs or proteins in analogy to similar procedures developed for PINs; networks of Protein-Protein Interactions (PPIs) [112]. (Table **5**) shows the predicted results of both node degree centrality (C_δ) and closeness centrality (C_{clo}) for proteins and drugs present in the database and compares them with the predicted results of 2D-MI-DRAGON model. The parameter C_δ measures the local importance of a node by counting the number of nodes directly attached to him [112]. Conversely, C_{clo} measures the global importance of a node in a CN by taking in consideration the inverse of the sum of D_{ij} ($C_{clo} = 1/\Sigma D_{ij}$) [113]. Consequently, the higher C_δ the higher is the local importance of the node but the higher C_{clo}, the lower is the global importance of the node. For instance, the protein 1HA2 is one important protein both locally and globally in this CNs with lower $C_{clo} > 4$ and a $C_\delta = 26$. It means that this protein is both locally and globally important because it is the target of many drugs (high C_δ). This result is similar to the one obtained by the 2D MI-DRAGON model. Another interesting result was referred to Simvastatin. Simvastatin is a hypolipidemic drug used to control elevated cholesterol, or hypercholesterolemia. It is a member of the statin class of pharmaceuticals. The primary use of Simvastatin is for the treatment of dyslipidemia and the prevention of cardiovascular disease [114, 115]. Depending on our aims, the most important nodes in pharmacological terms not necessarily have to be the most central in the graph (those with higher C_δ and lower C_{clo}), see (Table **5**). We show in this example, our model predicts efficiently. We found that the 3D MI-DRAGON model shows very similar results to the previous model, which has been published with excellent results. In Table 2SM of supplementary material we show all node degrees and closeness results.

Table 4.   Comparison 3D MI-DRAGON versus 2D MI-DRAGON

2D MI-DRAGON	Value	TIs	Value	3D MI-DRAGON
	706	n	59	
	907	m	631	
	1826812	W	2057954	
	18	D	19	
	255.09	Xr	266.39	
	2.49	δ	2.44	
	6.7	Dist	7.2	
	0.078	E	0.083	
	12.43	R	11.39	

[a]The TIs used are: number of nodes (n), number of edges (m), Wiener index (W), diameter (D), the Randic connectivity index (Xr), topological distance (Dist), network average values for radiality (R), node degree (δ), eccentricity (E).

Table 5.   Results of Node Degree (C_δ) and Closeness Centrality (C_{clo}) for 20 Proteins and Drugs

Drug/PDB	C_δ 2D-MI-DRAGON	C_δ 3D MI-DRAGON	Drug/PDB	C_{clo} 2D-MI-DRAGON	C_{clo} 3D MI-DRAGON
1HA2	44	26	1HA2	4.80	3.54
1BNA	36	40	Simvastatin	4.22	4.11
NADH	35	33	Gliclazide	4.17	3.48
1R5K	27	29	Saquinavir	4.16	3.46
Simvastatin	18	17	1BNA	4.15	3.46
1EMI	16	21	Cefalotin	4.13	2.98
1CZM	14	17	Atorvastatin	4.09	3.44
1NHZ	14	14	1A8M	4.09	3.75
1MO8	14	14	1XF0	4.07	3.09
1UZF	13	13	Estrone	4.06	2.61
1SQN	13	13	Ketoprofen	4.02	2.86
1T9N	13	12	Testosterone	4.02	2.42
1BYW	11	15	1TZI	4.02	3.77
1VRU	11	10	1KED	4.00	3.16
1E3G	11	10	Captopril	3.99	3.49
Atorvastatin	11	10	Liothyronine	3.97	3.21
1ZNC	10	11	Diflunisal	3.96	3.20
1ODW	10	10	Halothane	3.95	3.11
Pyridoxal Phosphate	9	9	Digitoxin	3.94	3.45
1HWL	9	9	Pyridoxine	3.94	3.08

3.2. A Theoretical-Experimental Study Using the 3D MI-DRAGON Predictor

Finally, we have illustrated in a theoretical-experimental study the practical use of 3D MI-DRAGON. We have reported the prediction, synthesis, and pharmacological assay of 20 different rasagiline derivatives with AChE inhibitory activity.

3.2.1. 3D MI-DRAGON Prediction of Rasagiline Derivatives vs. AChE

In this *in silico* experiment we have used 3D MI-DRAGON to predict the interaction of the rasagiline derivatives with respect to AChE. To this end, we have downloaded the 3D structure of AChE protein with PDB ID 1EEA and calculated their structural parameters with MI. We have also generated the SMILE codes for these compounds and we have used MOPAC AM1 Optimization geometry method for these compounds to calculate their 3D structural parameters with DRAGON. Next, we have predicted their propensity to undergo DPIs with AChE using as inputs for the 3D MI-DRAGON predictor the structural parameters of both drugs and proteins. In (Table **6**) we have confronted the results obtained using this model and the outcomes of the pharmacological assay. No compounds are selective inhibitors of AChE, which is why we used as control galantamine for AChE. We have considered that the observed class for active compounds OC = 1 if compound IC_{50} < 10 μM; this cutoff is in the similar range as others used in previous works [15, 116]. As we can see in this table, all the rasagiline derivative compounds present some activity, but none of these compounds have inhibitory activity in the pharmacological assays. All our compounds in the pharmacological assay (OC = 0) were inactive. 3D MI-DRAGON predicted as inactive all compounds, except for 3 of them. The model classified correctly 19 of 22 compounds tested (86.36%). In this test, our model was compared with pharmacological testing of 22 compounds synthesized by us. And we could observe the effectiveness of our model with experimental data. Moreover, we note that the model predicted all compounds tested as inactive, which is important because the model allows discriminating between active and inactive compounds. However, some compounds were not tested by the pharmacological assay, those compounds were predicted as inactive using 3D MI-DRAGON model. We discarded the pharmaceutical assays of these compounds because we consider that our model is reliable. This kind of model can be used to save efforts and money when performing the pharmacological tests. This is a good example of how reliable is the 3D MI-DRAGON model.

3.2.2. 3D MI-DRAGON Complex Network of Rasagiline Derivatives vs. US FDA Proteins

An additional use of 3D MI-DRAGON was to carry out the *in silico* or virtual screening of the new compounds with respect to all the other targets previously approved by US FDA [14, 117]. It may help to find new targets for these drugs or discard possible toxicological effects depending on the other targets predicted and/or discarded for these compounds. This type of experiment is of major importance due to the cost in terms of animal sacrifice, time, materials and human resources of the experimental assay of all compounds

against all these targets, see recent reviews by Duardo-Sánchez *et al.* [34, 118-120]. In fact, over a decade, the US FDA has been engaged in the applied research, development, and evaluation of computational toxicology methods used to support the safety assessment of a diverse set of regulated products. The basis for evaluating computational toxicology methods is multi-factorial, including the potential for increased efficiency, reduction of the number of animals used, lower costs, and the need to explore emerging technologies that support the goals of the US FDA's Critical Path Initiative (e.g. to make decisions that support the information available early in the drug review process)[121].
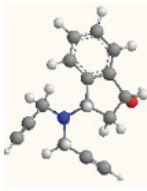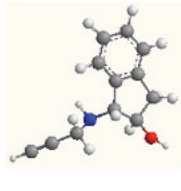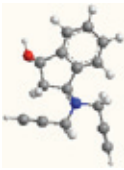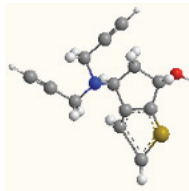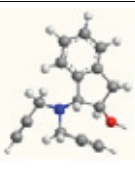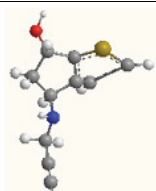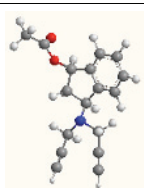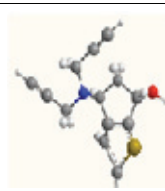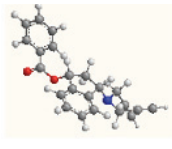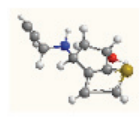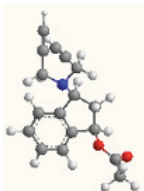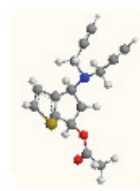
In this experiment, we have downloaded the 3D structure of all proteins that are targets of the US FDA approved drugs. Next, we have calculated the structural parameters of all these proteins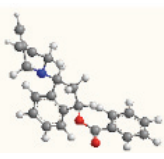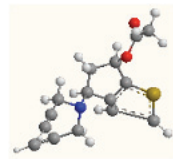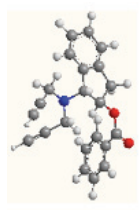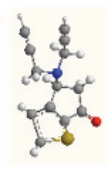 with MI. We have also generated the SMILE codes for these compounds and we have used MOPAC AM1 Optimization geometry method for these compounds to calculate their 3D structural parameters with DRAGON. Next, we have predicted their propensity to undergo DPIs with all US FDA proteins using as inputs for the 3D MI-DRAGON predictor the structural parameters of both drugs and proteins. We have predicted all proteins in the FDA dataset *vs.* the 22 rasagiline derivatives. We have found that most of the 22 derivatives were predicted as non-active (low DPIs scores) against most proteins in the FDA database. Consequently, 3D MI-DRAGON has predicted a high selectivity of rasagiline derivatives as AChE inhibitors. We were able to reach this goal because the model predicted these compounds as non-active with respect to most proteins that were targets of the FDA drugs.

Using these results, we have developed a DP-CN for rasagiline derivatives and the FDA dataset see Fig. (**11**). Consequently, we have obtained a CN with 87 nodes (FDA drugs, proteins, or rasagiline derivatives) and 166 DP (edges, DTPs). As in this network we can see that protein 1EEA (AChE) is predicted to interact with compound 3, this protein is an AChE target [122]. These results are good because they are consistent with the experimental results presented in this paper in which the compound 3 had shown low AChE activity. The use of such complex networks can help us find and predict new drug-protein interactions, and therefore find new drugs with improved biological activity and fewer side effects, especially in neurological diseases.
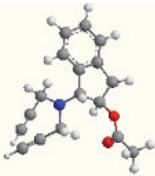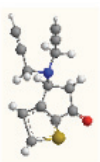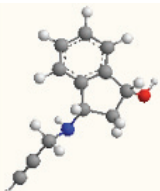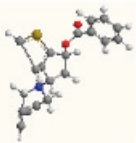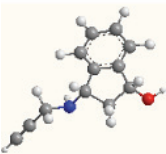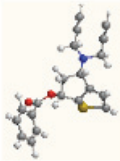
4. CONCLUSIONS

The 3D MI-DRAGON predictor is based on structural parameters of drugs, calculated with DRAGON and parameters of proteins calculated with MI. It is possible to seek excellent predictors for DPIs using as input structural parameters of drugs and proteins calculated with different programs and combined with ANN models. Combining MARCH-INSIDE and DRAGON approach and ANN is possible to seek an mt-QSAR classifier to predict with Accuracy > 85% the probability of drugs to bind more than 500 different drug target proteins approved by the US FDA. The 3D MI-DRAGON predictor is also useful to assemble CNs of DPIs. These computationally assembled CNs offer an alternative to discover new drugs or targets, and explore the selectivity of drugs. In this work, we have exemplified these conclusions

Table 6. **Prediction of Rasagiline derivatives with 3D MI-DRAGON Predictor**

Drug	OC	PC	Score	Structure	DRUG	OC	PC	Score	Structure
1	0	0	0.95		12	0	0	0.63	
2	0	0	0.95		13	0	0	1.00	
3	0	0	0.86		14	0	0	0.87	
4	0	0	0.95		15	0	0	1.00	
5	0	1	0.52		16	0	0	0.87	
6	0	0	0.95		17	0	0	1.00	
7	0	1	0.52		18	0	0	1.00	
8	0	0	0.88		19	0	0	1.00	

(Table 6) contd….

Drug	OC	PC	Score	Structure	DRUG	OC	PC	Score	Structure
9	0	0	0.89		20	0	0	1.00	
10	0	0	0.75		21	0	0	0.97	
11	0	1	0.27		22	0	0	0.97	

OC = Observed class; PC = Predicted class
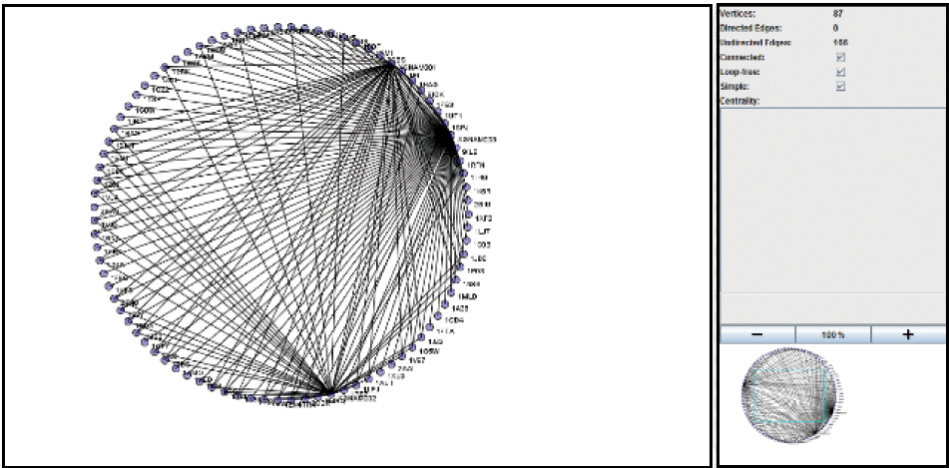
Fig. (11). DP-CN for rasagiline derivatives and the FDA dataset.

through the experimental-theoretical study of the AChE activity of new rasagiline derivatives.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Yildirim, M.A.; Goh, K.I.; Cusick, M.E.; Barabasi, A.L.; Vidal, M. Drug-target network. *Nature Biotechnology,* **2007**, *25*, 1119-1126.

[2]   Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics,* **2008**, *24*, i232-240.

[3]   Prado-Prado, F.; Garcia-Mera, X.; Abeijon, P.; Alonso, N.; Caamano, O.; Yanez, M.; Garate, T.; Mezo, M.; Gonzalez-Warleta, M.; Muino, L.; Ubeira, F.M.; Gonzalez-Diaz, H. Using entropy of drug and protein graphs to predict FDA drug-target network: theoretic-experimental study of MAO inhibitors and hemoglobin

peptides from Fasciola hepatica. *Eur J Med Chem,* **2011**, *46*, 1074-1094.

[4] Prado-Prado, F.J.; Borges, F.; Perez-Montoto, L.G.; Gonzalez-Diaz, H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *Eur J Med Chem,* **2009**, *44*, 4051-4056.

[5] Helguera, A.M.; Combes, R.D.; Gonzalez, M.P.; Cordeiro, M.N. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr Top Med Chem,* **2008**, *8*, 1628-1655.

[6] Estrada, E.; Molina, E.; Nodarse, D.; Uriarte, E. Structural contributions of substrates to their binding to P-Glycoprotein. A TOPS-MODE approach. *Curr Pharm Des,* **2010**, *16*, 2676-2709.

[7] Marrero-Ponce, Y.; Casanola-Martin, G.M.; Khan, M.T.; Torrens, F.; Rescigno, A.; Abad, C. Ligand-based computer-aided discovery of tyrosinase inhibitors. Applications of the TOMOCOMD-CARDD method to the elucidation of new compounds. *Curr Pharm Des,* **2010**, *16*, 2601-2624.

[8] Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Perez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr Drug Metab,* **2010**, *11*, 379-406.

[9] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. Chemoinformatics in Multi-Target Drug Discovery for Anti-Cancer Therapy: In Silico Design Of Potent And Versatile Anti-Brain Tumor Agents. *Anticancer Agents Med Chem,* **2011**.

[10] Speck-Planche, A.; Kleandrova, V.V. In silico design of multi-target inhibitors for C-C chemokine receptors using substructural descriptors. *Mol Divers,* **2011**.

[11] Vina, D.; Uriarte, E.; Orallo, F.; Gonzalez-Diaz, H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol Pharm,* **2009**, *6*, 825-835.

[12] Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C.R.; Pazos, A.; Dea-Ayuela, M.A.; Gomez-Munoz, M.T.; Garijo, M.M.; Sansano, J.; Ubeira, F.M. MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from Trichomonas gallinae. *J Proteome Res,* **2011**, *10*, 1698-1718.

[13] Gonzalez-Diaz, H.; Prado-Prado, F.; Sobarzo-Sanchez, E.; Haddad, M.; Maurel Chevalley, S.; Valentin, A.; Quetin-Leclercq, J.; Dea-Ayuela, M.A.; Teresa Gomez-Munos, M.; Munteanu, C.R.; Jose Torres-Labandeira, J.; Garcia-Mera, X.; Tapia, R.A.; Ubeira, F.M. NL MIND-BEST: a web server for ligands and proteins discovery--theoretic-experimental study of proteins of Giardia lamblia and new compounds active against Plasmodium falciparum. *J Theor Biol,* **2011**, *276*, 229-249.

[14] Prado-Prado, F.; Garcia-Mera, X.; Escobar, M.; Sobarzo-Sanchez, E.; Yanez, M.; Riera-Fernandez, P.; Gonzalez-Diaz, H. 2D MI-DRAGON: a new predictor for protein-ligands interactions and theoretic-experimental studies of US FDA drug-target network, oxoisoaporphine inhibitors for MAO-A and human parasite proteins. *Eur J Med Chem,* **2011**, *46*, 5838-5851.

[15] Santana, L.; Uriarte, E.; González-Díaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *Journal of Medicinal Chemistry,* **2006**, *49*, 1149-1156.

[16] Marrero-Ponce, Y. Linear indices of the "molecular pseudograph's atom adjacency matrix": definition, significance-interpretation, and application to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors. *J Chem Inf Comput Sci,* **2004**, *44*, 2010-2026.

[17] Vilar, S.; Santana, L.; Uriarte, E. Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. *J Med Chem* **2006**, *49*, 1118-1124.

[18] Marrero-Ponce, Y.; Khan, M.T.; Casanola Martin, G.M.; Ather, A.; Sultankhodzhaev, M.N.; Torrens, F.; Rotondo, R. Prediction of tyrosinase inhibition activity using atom-based bilinear indices. *ChemMedChem,* **2007**, *2*, 449-478.

[19] Casanola-Martin, G.M.; Marrero-Ponce, Y.; Khan, M.T.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorganic & Medicinal Chemistry,* **2007**, *15*, 1483-1503.

[20] Casanola-Martin, G.M.; Marrero-Ponce, Y.; Khan, M.T.; Ather, A.; Khan, K.M.; Torrens, F.; Rotondo, R. Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental *in vitro* assays. *Eur J Med Chem,* **2007**, *42*, 1370-1381.

[21] Nunez, M.B.; Maguna, F.P.; Okulik, N.B.; Castro, E.A. QSAR modeling of the MAO inhibitory activity of xanthones derivatives. *Bioorg Med Chem Lett,* **2004**, *14*, 5611-5617.

[22] Terada, M.; Inaba, M.; Yano, Y.; Hasuma, T.; Nishizawa, Y.; Morii, H.; Otani, S. Growth-inhibitory effect of a high glucose concentration on osteoblast-like cells. *Bone,* **1998**, *22*, 17-23.

[23] González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics,* **2008**, *8*, 750-778.

[24] Zhao, C.J.; Dai, Q.Y. [Recent advances in study of antinociceptive conotoxins]. *Yao Xue Xue Bao,* **2009**, *44*, 561-565.

[25] Jacob, R.B.; McDougal, O.M. The M-superfamily of conotoxins: a review. *Cellular and Molecular Life Sciences,* **2010**, *67*, 17-27.

[26] Giuliani, A.; Di Paola, L.; Setola, R. Proteins as Networks: A Mesoscopic Approach Using Haemoglobin Molecule as Case Study. *Curr Proteomics,* **2009**, *6*, 235-245.

[27] Vilar, S.; Gonzalez-Diaz, H.; Santana, L.; Uriarte, E. A network-QSAR model for prediction of genetic-component biomarkers in human colorectal cancer. *Journal of Theoretical Biology,* **2009**, *261*, 449-458.

[28] Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Prado-Prado, F.J.; Uriarte, E.; Bolas-Fernandez, F.; Podda, G.; Pazos, A.; Munteanu, C.R.; Ubeira, F.M.; Gonzalez-Diaz, H. 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochimica et Biophysica Acta,* **2009**, *1794*, 1784-1794.

[29] Torrens, F.; Castellano, G. Topological Charge-Transfer Indices: From Small Molecules to Proteins. *Curr Proteomics,* **2009**, 204-213.

[30] Vázquez, J.M.; Aguiar, V.; Seoane, J.A.; Freire, A.; Serantes, J.A.; Dorado, J.; Pazos, A.; Munteanu, C.R. Star Graphs of Protein Sequences and Proteome Mass Spectra in Cancer Prediction. *Curr Proteomics,* **2009**, *6*, 275-288.

[31] Gonzalez-Diaz, H. Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR). *Curr Top Med Chem,* **2008**, *8*, 1554.

[32] Ivanciuc, O. Weka machine learning for predicting the phospholipidosis inducing potential. *Curr Top Med Chem,* **2008**, *8*, 1691-1709.

[33] Gonzalez-Diaz, H.; Prado-Prado, F.; Ubeira, F.M. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem,* **2008**, *8*, 1676-1690.

[34] Duardo-Sanchez, A.; Patlewicz, G.; Lopez-Diaz, A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. *Curr Top Med Chem,* **2008**, *8*, 1666-1675.

[35] Wang, J.F.; Wei, D.Q.; Chou, K.C. Drug candidates from traditional chinese medicines. *Curr Top Med Chem,* **2008**, *8*, 1656-1665.

[36] Gonzalez, M.P.; Teran, C.; Saiz-Urra, L.; Teijeira, M. Variable selection methods in QSAR: an overview. *Curr Top Med Chem,* **2008**, *8*, 1606-1627.

[37] Caballero, J.; Fernandez, M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). *Curr Top Med Chem,* **2008**, *8*, 1580-1605.

[38] Wang, J.F.; Wei, D.Q.; Chou, K.C. Pharmacogenomics and personalized use of drugs. *Curr Top Med Chem,* **2008**, *8*, 1573-1579.

[39] Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem,* **2008**, *8*, 1555-1572.

[40] Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorganic & Medicinal Chemistry,* **2010**, *18*, 2225-2231.

[41] Gonzalez-Diaz, H. Network topological indices, drug metabolism, and distribution. *Curr Drug Metab, 11*, 283-284.

[42]   Khan, M.T. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr Drug Metab, 11*, 285-295.

[43]   Mrabet, Y.; Semmar, N. Mathematical methods to analysis of topology, functional variability and evolution of metabolic systems based on different decomposition concepts. *Curr Drug Metab, 11*, 315-341.

[44]   Martinez-Romero, M.; Vazquez-Naya, J.M.; Rabunal, J.R.; Pita-Fernandez, S.; Macenlle, R.; Castro-Alvarino, J.; Lopez-Roses, L.; Ulla, J.L.; Martinez-Calvo, A.V.; Vazquez, S.; Pereira, J.; Porto-Pazos, A.B.; Dorado, J.; Pazos, A.; Munteanu, C.R. Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. *Curr Drug Metab, 11*, 347-368.

[45]   Zhong, W.Z.; Zhan, J.; Kang, P.; Yamazaki, S. Gender specific drug metabolism of PF-02341066 in rats--role of sulfoconjugation. *Curr Drug Metab, 11*, 296-306.

[46]   Wang, J.F.; Chou, K.C. Molecular modeling of cytochrome P450 and drug metabolism. *Curr Drug Metab, 11*, 342-346.

[47]   Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Perez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr Drug Metab, 11*, 379-406.

[48]   Garcia, I.; Diop, Y.F.; Gomez, G. QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr Drug Metab, 11*, 307-314.

[49]   Chou, K.C. Graphic rule for drug metabolism systems. *Curr Drug Metab, 11*, 369-378.

[50]   Concu, R.; Podda, G.; Ubeira, F.M.; Gonzalez-Diaz, H. Review of QSAR Models for Enzyme Classes of Drug Targets: Theoretical Background and Applications in Parasites, Hosts, and other Organisms. *Current Pharmaceutical Design,* **2010**, *16*, 2710-2723.

[51]   Estrada, E.; Molina, E.; Nodarse, D.; Uriarte, E. Structural Contributions of Substrates to their Binding to P-Glycoprotein. A TOPS-MODE Approach. *Current Pharmaceutical Design,* **2010**, *16*, 2676-2709.

[52]   Garcia, I.; Fall, Y.; Gomez, G. QSAR, Docking, and CoMFA Studies of GSK3 Inhibitors. *Current Pharmaceutical Design,* **2010**, *16*, 2666-2675.

[53]   González-Díaz, H. QSAR and Complex Networks in Pharmaceutical Design, Microbiology, Parasitology, Toxicology, Cancer, and Neurosciences. *Current Pharmaceutical Design,* **2010**, *16*, 2598-2600.

[54]   Gonzalez-Diaz, H.; Romaris, F.; Duardo-Sanchez, A.; Perez-Mototo, L.G.; Prado-Prado, F.; Patlewicz, G.; Ubeira, F.M. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, aplications, and legal issues. *Current Pharmaceutical Design,* **2010**, *16*, 2737-2764.

[55]   Marrero-Ponce, Y.; Casanola-Martin, G.M.; Khan, M.T.; Torrens, F.; Rescigno, A.; Abad, C. Ligand-Based Computer-Aided Discovery of Tyrosinase Inhibitors. Applications of the TOMOCOMD-CARDD Method to the Elucidation of New Compounds. *Current Pharmaceutical Design,* **2010**, *16*, 2601-2624.

[56]   Munteanu, C.R.; Fernandez-Blanco, E.; Seoane, J.A.; Izquierdo-Novo, P.; Rodriguez-Fernandez, J.A.; Prieto-Gonzalez, J.M.; Rabunal, J.R.; Pazos, A. Drug Discovery and Design for Complex Diseases through QSAR Computational Methods. *Current Pharmaceutical Design,* **2010**, *16*, 2640-2655.

[57]   Roy, K.; Ghosh, G. Exploring QSARs with Extended Topochemical Atom (ETA) Indices for Modeling Chemical and Drug Toxicity. *Current Pharmaceutical Design,* **2010**, *16*, 2625-2639.

[58]   Speck-Planche, A.; Scotti, M.T.; de Paulo-Emerenciano, V. Current pharmaceutical design of antituberculosis drugs: future perspectives. *Current Pharmaceutical Design,* **2010**, *16*, 2656-2665.

[59]   Vazquez-Naya, J.M.; Martinez-Romero, M.; Porto-Pazos, A.B.; Novoa, F.; Valladares-Ayerbes, M.; Pereira, J.; Munteanu, C.R.; Dorado, J. Ontologies of drug discovery and design for neurology, cardiology and oncology. *Current Pharmaceutical Design,* **2010**, *16*, 2724-2736.

[60]   Kirchmair, J.; Markt, P.; Distinto, S.; Schuster, D.; Spitzer, G.M.; Liedl, K.R.; Langer, T.; Wolber, G. The Protein Data Bank (PDB), its related services and software tools as key components for in silico guided drug discovery. *J Med Chem,* **2008**, *51*, 7021-7040.

[61]   Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A.C.; Wishart, D.S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res,* **2011**, *39*, D1035-1041.

[62]   Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res,* **2008**, *36*, D901-906.

[63]   Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res,* **2006**, *34*, D668-672.

[64]   Chen, X.; Fang, Y.; Yao, L.; Chen, Y.; Xu, H. Does drug-target have a likeness? *Methods Inf Med,* **2007**, *46*, 360-366.

[65]   Li, J.; Lei, B.; Liu, H.; Li, S.; Yao, X.; Liu, M.; Gramatica, P. QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling. *J Comput Chem,* **2008**, *29*, 2636-2647.

[66]   Bleakley, K.; Yamanishi, Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics,* **2009**, *25*, 2397-2403.

[67]   Bas, D.C.; Rogers, D.M.; Jensen, J.H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins,* **2008**, *73*, 765-783.

[68]   Li, Q.; Lai, L. Prediction of potential drug targets based on simple sequence properties. *BMC Bioinformatics,* **2007**, *8*, 353.

[69]   Yang, C.Y.; Sun, H.; Chen, J.; Nikolovska-Coleska, Z.; Wang, S. Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. *J Am Chem Soc,* **2009**, *131*, 13709-13721.

[70]   Talete srl. DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.3 ed.; 2005.

[71]   Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors.*, 2000.

[72]   Papa, E.; Villa, F.; Gramatica, P. Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in Pimephales promelas (fathead minnow). *J Chem Inf Model,* **2005**, *45*, 1256-1266.

[73]   González-Díaz, H.; Pérez-Castillo, Y.; Podda, G.; Uriarte, E. Computational Chemistry Comparison of Stable/Nonstable Protein Mutants Classification Models Based on 3D and Topological Indices. *Journal of Computational Chemistry,* **2007**, *28*, 1990-1995.

[74]   Gonzalez-Diaz, H.; Saiz-Urra, L.; Molina, R.; Gonzalez-Diaz, Y.; Sanchez-Gonzalez, A. Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments. *J Comput Chem,* **2007**, *28*, 1042-1048.

[75]   Aguero-Chapin, G.; Antunes, A.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. Comparative Study of Topological Indices of Macro/Supramolecular RNA Complex Networks. *J Chem Inf Model,* **2008**, *48*, 2265–2277.

[76]   Cruz-Monteagudo, M.; Munteanu, C.R.; Borges, F.; Cordeiro, M.N.D.S.; Uriarte, E.; Chou, K.-C.; González-Díaz, H. Stochastic molecular descriptors for polymers. 4. Study of complex mixtures with topological indices of mass spectra spiral and star networks: The blood proteome case. *Polymer,* **2008**, *49*, 5575-5587.

[77]   Dea-Ayuela, M.A.; Perez-Castillo, Y.; Meneses-Marcel, A.; Ubeira, F.M.; Bolas-Fernandez, F.; Chou, K.C.; Gonzalez-Diaz, H. HP-Lattice QSAR for dynein proteins: experimental proteomics (2D-electrophoresis, mass spectrometry) and theoretic study of a Leishmania infantum sequence. *Bioorganic & Medicinal Chemistry,* **2008**, *16*, 7770-7776.

[78]   Aguero-Chapin, G.; Gonzalez-Diaz, H.; de la Riva, G.; Rodriguez, E.; Sanchez-Rodriguez, A.; Podda, G.; Vazquez-Padron, R.I. MMM-QSAR recognition of ribonucleases without alignment: comparison with an HMM model and isolation from Schizosaccharomyces pombe, prediction, and experimental assay of a new sequence. *J Chem Inf Model,* **2008**, *48*, 434-448.

[79]   Ferino, G.; Gonzalez-Diaz, H.; Delogu, G.; Podda, G.; Uriarte, E. Using spectral moments of spiral networks based on PSA/mass spectra outcomes to derive quantitative proteome-disease relationships (QPDRs) and predicting prostate cancer. *Biochemical and Biophysical Research Communications,* **2008**, *372*, 320-325.

[80] Gonzalez-Diaz, H.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Prado-Prado, F.J.; Aguero-Chapin, G.; Bolas-Fernandez, F.; Vazquez-Padron, R.I.; Ubeira, F.M. QSAR for RNases and theoretic-experimental study of molecular diversity on peptide mass fingerprints of a new Leishmania infantum protein. *Molecular Diversity,* **2009**.

[81] Aguero-Chapin, G.; Varona-Santos, J.; de la Riva, G.A.; Antunes, A.; Gonzalez-Villa, T.; Uriarte, E.; Gonzalez-Diaz, H. Alignment-Free Prediction of Polygalacturonases with Pseudofolding Topological Indices: Experimental Isolation from Coffea arabica and Prediction of a New Sequence. *J Proteome Res, 2009*, *8*, 2122-2128.

[82] Gonzalez-Diaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J Proteome Res,* **2007**, *6*, 904-908.

[83] Gonzalez-Diaz, H.; Molina, R.; Uriarte, E. Recognition of stable protein mutants with 3D stochastic average electrostatic potentials. *FEBS Letters,* **2005**, *579*, 4297-4301.

[84] Concu, R.; Podda, G.; Uriarte, E.; Gonzalez-Diaz, H. Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *Journal of Computational Chemistry,* **2009**, *30*, 1510-1520.

[85] StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, *www.statsoft.com.Statsoft*, Inc., 6.0; 2002.

[86] Casanola-Martin, G.M.; Marrero-Ponce, Y.; Khan, M.T.; Khan, S.B.; Torrens, F.; Perez-Jimenez, F.; Rescigno, A.; Abad, C. Bond-based 2D quadratic fingerprints in QSAR studies: virtual and *in vitro* tyrosinase inhibitory activity elucidation. *Chem Biol Drug Des,* **2010**, *76*, 538-545.

[87] Castillo-Garit, J.A.; Vega, M.C.; Rolon, M.; Marrero-Ponce, Y.; Kouznetsov, V.V.; Torres, D.F.; Gomez-Barrio, A.; Bello, A.A.; Montero, A.; Torrens, F.; Perez-Gimenez, F. Computational discovery of novel trypanosomicidal drug-like chemicals by using bond-based non-stochastic and stochastic quadratic maps and linear discriminant analysis. *Eur J Pharm Sci, 2010*, *39*, 30-36.

[88] Gozalbes, R.; Barbosa, F.; Nicolai, E.; Horvath, D.; Froloff, N. Development and validation of a pharmacophore-based QSAR model for the prediction of CNS activity. *ChemMedChem, 2009*, *4*, 204-209.

[89] Marrero-Ponce, Y.; Meneses-Marcel, A.; Rivera-Borroto, O.M.; Garcia-Domenech, R.; De Julian-Ortiz, J.V.; Montero, A.; Escario, J.A.; Barrio, A.G.; Pereira, D.M.; Nogal, J.J.; Grau, R.; Torrens, F.; Vogel, C.; Aran, V.J. Bond-based linear indices in QSAR: computational discovery of novel anti-trichomonal compounds. *J Comput Aided Mol Des, 2008*, *22*, 523-540.

[90] Patankar, S.J.; Jurs, P.C. Classification of inhibitors of protein tyrosine phosphatase 1B using molecular structure based descriptors. *J Chem Inf Comput Sci, 2003*, *43*, 885-899.

[91] Murcia-Soler, M.; Perez-Gimenez, F.; Garcia-March, F.J.; Salabert-Salvador, M.T.; Diaz-Villanueva, W.; Medina-Casamayor, P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J Mol Graph Model, 2003*, *21*, 375-390.

[92] Cercos-del-Pozo, R.A.; Perez-Gimenez, F.; Salabert-Salvador, M.T.; Garcia-March, F.J. Discrimination and molecular design of new theoretical hypolipaemic agents using the molecular connectivity functions. *J Chem Inf Comput Sci, 2000*, *40*, 178-184.

[93] Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Pazos, A.; Prado-Prado, F.J.; Gonzalez-Diaz, H. Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J Proteome Res, 2010*, *9*, 1182-1190.

[94] Prado-Prado, F.J.; Gonzalez-Diaz, H.; Santana, L.; Uriarte, E. Unified QSAR approach to antimicrobials. Part 2: predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorganic & Medicinal Chemistry, 2007*, *15*, 897-902.

[95] Junker, B.H.; Koschutzki, D.; Schreiber, F. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics, 2006*, *7*, 219.

[96] Koschützki, D. *CentiBiN Version 1.4.2*, 2006.

[97] GL, E.; KD, C.; Jr, A.V.; RM, F.-S. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochemistry Pharmacology,* **1961**, *7*, 88-95.

[98] Ivanciuc, O.; Oezguen, N.; Mathura, V.S.; Schein, C.H.; Xu, Y.; Braun, W. Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins. *Current Medicinal Chemistry, 2004*, *11*, 583-593.

[99] Schein, C.H.; Ivanciuc, O.; Braun, W. Common physical-chemical properties correlate with similar structure of the IgE epitopes of peanut allergens. *Journal of Agricultural and Food Chemistry,* **2005**, *53*, 8752-8759.

[100] Alvarez-Ginarte, Y.M.; Marrero-Ponce, Y.; Ruiz-Garcia, J.A.; Montero-Cabrera, L.A.; Vega, J.M.; Noheda Marin, P.; Crespo-Otero, R.; Zaragoza, F.T.; Garcia-Domenech, R. Applying pattern recognition methods plus quantum and physico-chemical molecular descriptors to analyze the anabolic activity of structurally diverse steroids. *Journal of Computational Chemistry,* **2007**.

[101] Morales, A.H.; Rodríguez-Borges, J.E.; García-Mera, X.; Fernández, F.; Dias-Sueiro-Cordeiro, M.N. Probing the Anticancer Activity of Nucleoside Analogues: A QSAR Model Approach Using an Internally Consistent Training Set. *Journal of Medicinal Chemistry, 2007*, *50*, 1537-1545.

[102] Fernandez, M.; Caballero, J.; Tundidor-Camba, A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. *Bioorg Med Chem, 2006*, *14*, 4137-4150.

[103] Caballero, J.; Fernandez, M. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. *J Mol Model, 2006*, *12*, 168-181.

[104] Mascarenhas, N.M.; Ghoshal, N. Combined ligand and structure based approaches for narrowing on the essential physicochemical characteristics for CDK4 inhibition. *J Chem Inf Model, 2008*, *48*, 1325-1336.

[105] Vijayan, R.S.; Bera, I.; Prabu, M.; Saha, S.; Ghoshal, N. Combinatorial library enumeration and lead hopping using comparative interaction fingerprint analysis and classical 2D QSAR methods for seeking novel GABA(A) alpha(3) modulators. *J Chem Inf Model, 2009*, *49*, 2498-2511.

[106] Pham, T.A.; Jain, A.N. Parameter estimation for scoring protein-ligand interactions using negative training data. *J Med Chem, 2006*, *49*, 5856-5868.

[107] Raha, K.; Peters, M.B.; Wang, B.; Yu, N.; Wollacott, A.M.; Westerhoff, L.M.; Merz, K.M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discov Today, 2007*, *12*, 725-731.

[108] Zhang, S.; Golbraikh, A.; Tropsha, A. Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces. *J Med Chem, 2006*, *49*, 2713-2724.

[109] Michel, J.; Verdonk, M.L.; Essex, J.W. Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? *J Med Chem, 2006*, *49*, 7427-7439.

[110] Naumann, T.; Matter, H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J Med Chem, 2002*, *45*, 2366-2378.

[111] Gonzalez-Diaz, H.; Gonzalez-Diaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics, 2008*, *8*, 750-778.

[112] Jeong, H.; Mason, S.P.; Barabasi, A.L.; Oltvai, Z.N. Lethality and centrality in protein networks. *Nature, 2001*, *411*, 41-42.

[113] Estrada, E. Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics, 2006*, *6*, 35-40.

[114] Todd, P.A.; Goa, K.L. Simvastatin. A review of its pharmacological properties and therapeutic potential in hypercholesterolaemia. *Drugs, 1990*, *40*, 583-607.

[115] Hernandez-Perera, O.; Perez-Sala, D.; Navarro-Antolin, J.; Sanchez-Pascuala, R.; Hernandez, G.; Diaz, C.; Lamas, S. Effects of the 3-hydroxy-3-methylglutaryl-CoA reductase inhibitors, atorvastatin and simvastatin, on the expression of endothelin-1 and endothelial nitric oxide synthase in vascular endothelial cells. *J Clin Invest, 1998*, *101*, 2711-2719.

[116] Santana, L.; Gonzalez-Diaz, H.; Quezada, E.; Uriarte, E.; Yanez, M.; Vina, D.; Orallo, F. Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and B inhibitors. *J Med Chem, 2008*, *51*, 6740-6751.

[117]   Nguyen, N.T.; Cook, D.M.; Bero, L.A. The decision-making process of US Food and Drug Administration advisory committees on switches from prescription to over-the-counter status: a comparative case study. *Clin Ther,* **2006,** *28,* 1231-1243.

[118]   González-Díaz, H.; Prado-Prado, F.; Pérez-Montoto, L.G.; Duardo-Sánchez, A.; López-Díaz, A. QSAR Models for Proteins of Parasitic Organisms, Plants and Human Guests: Theory, Applications, Legal Protection, Taxes, and Regulatory Issues. *Curr Proteomics,* **2009,** *6,* 214-227.

[119]   González-Díaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Pérez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. Review of MARCH-INSIDE & Complex Networks prediction of Drugs: ADMET, Anti-parasite Activity, Metabolizing Enzymes and Cardiotoxicity Proteome Biomarkers *Current Drug Metabolism,* **2010,** *11,* 379-406.

[120]   Gonzalez-Diaz, H.; Romaris, F.; Duardo-Sanchez, A.; Perez-Montoto, L.G.; Prado-Prado, F.; Patlewicz, G.; Ubeira, F.M. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Curr Pharm Des,* **2010,** *16,* 2737-2764.

[121]   Yang, C.; Valerio, L.G., Jr.; Arvidson, K.B. Computational toxicology approaches at the US Food and Drug Administration. *Altern Lab Anim,* **2009,** *37,* 523-531.

[122]   Raves, M.L.; Harel, M.; Pang, Y.P.; Silman, I.; Kozikowski, A.P.; Sussman, J.L. Structure of acetylcholinesterase complexed with the nootropic alkaloid, (-)-huperzine A. *Nat Struct Biol,* **1997,** *4,* 57-63.