Biochemistry & Pharmacology: Open Access

State of Art: Review of Theoretical Study of GSK-3 β and a New Neural Networks QSAR Studies for the Design of New Inhibitors Using 2D-Descriptors

García I[1] and Prado-Prado F[2*]

[1]Department of Organic Chemistry, University of Vigo, 36200, Spain
[2]Biomedical Sciences Department, Health Sciences Division, University of Quintana Roo, 77039, Mexico

Abstract

Alzheimer's disease (AD) is characterized by several pathologies, as this disease involves neuropathological lesions in the brain. Indeed, a wealth of evidence suggests that β-amyloid is central to the pathophysiology of AD and is likely to play an early role in this intractable neurodegenerative disorder. AD is the most prevalent form of dementia, and current indications show that twenty-nine million people live with AD worldwide, a figure expected to rise exponentially over the coming decades. Clearly, blocking disease progression or, in the best-case scenario, preventing AD altogether would be of benefit in both social and economic terms. However, current AD therapies are merely palliative and only temporarily slow cognitive decline, and treatments that address the underlying pathologic mechanisms of AD are completely lacking. While familial AD (FAD) is caused by autosomal dominant mutations in either amyloid precursor protein (APP) or the presenilin (PS1, PS2) genes. First, we have reviewed 2D QSAR, 3D QSAR, CoMFA, CoMSIA and docking for GSK-3α and GSK-3β with different compounds to find out their structural requirements. Next, we develop a QSAR for GSK-3β, because is one of the most important enzymes that intervenes in neuropathological disease such as Alzheimer. QSAR could play an important role in studying these GSK-3 inhibitors. For this reason we developed QSAR models for GSK-3β, LDA, ANNs and CT from more than 40000 cases with more than 2400 different molecules inhibitors of GSK-3β obtained from ChEMBL database server; in total we used more than 45000 different molecules to develop the QSAR models. We used 237 molecular descriptors calculated with DRAGON software. The model correctly classified 1310 out of 1643 active compounds (79.7%) and 24823 out of 26156 non-active compounds (94.9%) in the training series. The overall training performance was 94.0%. Validation of the model was carried out using an external predicting series. In this series the model classified correctly 757 out of 940 (80.5%) active compounds and 14 166 out of 14 937 non-active compounds (94.8%). The overall predictability performance was 94.0%. In this work, we propose five types of non Linear ANN and we show that it is another alternative model to the already existing ones in the literature, such as LDA. The best model obtained was RBF 166:166-402-1:1 which had an overall training performance of 94.2%. All this can help to design new inhibitors of GSK-3β. The present work reports the attempts to calculate within a unified framework probabilities of GSK-3β inhibitors against different molecules found in the literature.

Keywords: GSK-3β; QSAR; Artificial neural network; Linear neural network; Linear discriminant analysis

Introduction

Glycogen synthase kinase-3 (GSK-3) has two isoforms, GSK-3α and GSK-3β, [1] and they are serine/threonine kinases involved in numerous cellular processes and diverse diseases as Alzheimer disease, cancer, and diabetes. GSK-3α and GSK-3β have been shown to be present in mammals and the latter is specifically expressed in the central nervous system [2,3]. In particular, GSK-3β is well known to play critical roles in oxidative stress-induced neurodegenerative diseases such as Alzheimer´s disease (AD) [2,4]. Despite intensive investigation into the physiological roles of GSK-3 isoforms, the basis for their differential activities remains unresolved. A more comprehensive understanding of the mechanistic basis for GSK-3 isoform-specific functions could lead to the development of isoform-specific inhibitors [5]. GSK-3β knock-out mice die *in utero* [6], whereas GSK-3α knock-out mice are viable and display improved glucose tolerance in response to glucose load and elevated hepatic glycogen storage and insulin sensitivity [7,8].

Alzheimer´s disease [9] is a serious and degenerative disorder that causes a gradual loss of neurons, and in spite of the efforts realized by the big pharmaceutical companies of the world, the origin of this pathology is still not very clear. β-amyloid (Aβ) is an important protein implicated in the pathogenesis of AD, but the mechanism by which it causes neurotoxicity is still unknown [10,11]. In particular, there are few literature reports to study the direct link between the pathological hyperphosphorylation of tau protein, a microtubule-associated protein, and the formation of neurofibrillary tangles (NFT) [12]. The last decades had marked a very significant era of AD research. During this period, the nature of amyloid plaques and NTFs, the two histopathological hallmarks of AD, had been elucidated. Recent research efforts have led to several hypotheses to explain AD. Amyloid β toxicity is believed to play a primary role in the development of AD [13]. GSK-3β activity may increase with aging [14], which is consistent with the fact that aging is the most important risk factor for AD. Both *in vitro* and *in vivo* studies have demonstrated that inhibition of GSK-3β, can reverse hyperphosphorylation of tau and prevent behavioral impairments in mice [15-20]. These studies make GSK-3β inhibition very attractive as a therapeutic target for AD [21].

In the last years, a number of publications have been published suggesting GSK-3 as a target for the treatment of AD. There are two isoforms of GSK-3, GSK-3α and GSK-3β, both sharing a high homology

Citation: García I, Prado-Prado F (2015) State of Art: Review of Theoretical Study of GSK-3β and a New Neural Networks QSAR Studies for the Design of New Inhibitors Using 2D-Descriptors. Biochem Pharmacol (Los Angel) 4: 170. doi:10.4172/2167-0501.1000170

at their catalytic site, but the α form possesses an extended N-terminus with respect to the β form [22,23]. The phosphorylation of proteins by GSK-3 is an important link in neural function [24-26]. There are two characteristic neuropathological hallmarks of AD, Neurofibrillary Tangles (NFT's) and an increased production of amyloid beta (Aβ) peptides, where NFT's are composed of highly phosphorylated forms of the microtubule-associated protein tau [27] and studies have shown that GSK-3 is one of the main *in vivo* players of phosphorylation of tau protein [28]. It has been reported that Lithium, a GSK-3 inhibitor, blocks production of Aβ peptides by interfering with APP cleavage at γ-secretase step, where the target for Lithium is GSK-3α [22,29]. Phiel et al. [29] showed that selective reduction in concentration of the α isoform led to a decrease in the concentration of Aβ40 and Aβ42, primary constituents of amyloid plaques in AD. Thus, inhibition of GSK-3α could potentially provide dual therapy against AD, preventing the buildup of amyloid plaques and of neurofibrillary tangles [29-31].

GSK-3β is a serine/threonine kinase and is thought to be a key factor for aberrant tau phosphorylation [32]. Activated GSK-3β coexists with progression of NFT's and neurodegeneration in the AD brain [33-35]. A conditional GSK-3β overexpressing transgenic mouse exhibits persistent tau hyperphosphorylation, pretangle-like somatodendritic localization of tau, neuronal death in hippocampus and cognitive deficits [36,37]. These studies suggest that GSK-3β is associated with AD progression, and GSK-3β inhibition is expected to be a promising therapeutic approach for AD.

In this sense, quantitative structure-activity relationships (QSAR) could play an important role in studying these β and γ-secretase inhibitors. QSAR models are necessary in order to guide the β and γ-secretase inhibitors.

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 [38], HIV-1 integrase inhibitors [39], and/or protease inhibitors [40] or tyrosinase inhibitors [41-43]. 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 [44,45]. Recently, the field has moved from small molecules to proteins and other systems. For instance, González-Díaz et al. have discussed the use of these methods but only from the point of view of proteins [46]. Later, some groups have published different papers in one special issue on QSAR but they have been also restricted to the field of protein and proteomics [47-53]. In other recent issue, guest-edited by González-Díaz [54] a series of papers have been published, devoted to QSAR/QSPR techniques for low-molecular-weight drugs [54-63]. Most recently, Prado-Prado et al. [64] have published a mt-QSAR for anti-parasitic drugs. This year we have published another issue [65] focused on QSAR/QSPR models and a graph theory used to approach Drug ADMET processes and Metabolomics [66-73]. Last, one of the most recent issues published has discussed the applications of QSAR in Pharmaceutical Design [74-83].

The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors [12] in the last years. QSARs can be used as predictive tools for the development of molecules [84,85]. The QSAR approach involves the development of models that relate the structure of drugs with their biological activity against different targets [86,87]. Furthermore, there are multiple chemometric approaches that can, in principle, be selected for this step. Multiple linear regression (MLR), LDA, partial least squares (PLS) and different kinds of artificial neural networks

can be used to relate molecular structure (represented by molecular descriptors) with biological properties. The ANNs are particularly useful in QSAR studies in which the linear models fit poorly due to high data complexity; an example was the work of Prado-Prado et al. in which four types of non-ANN were developed to calculate within an unified framework probabilities of antiparasitic action of drugs against different parasite species [64,88,89]. There are several different kinds of ANN and these include multilayer perceptron (MLP), radial basis functions (RBF) and PNNs; the latter ANN is a variant of RBF systems. In particular, PNN is a type of neural network that uses a kernel-based approximation to form an estimate of the probability density functions of classes in a classification problem [90]. In the present work, 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, in this review, we developed quantitative structure-activity relationships (QSAR) models for GSK-3β, linear discriminant analysis (LDA) [91] and linear artificial neural networks (ANNs) from more than 40000 cases with more than 24000 different inhibitors of GSK-3β obtained from ChEMBL database http://www.ebi.ac.uk/chembldb/index.php/target/browser/classification [92,93]; in total we used more than 45000 different cases to develop the QSAR models. In addition, we did a study of different fragments that exist in the molecules of the database in order to see which fragments had more influence in the activity, and which fragments interact more with the protein. As there are very studies with GSK-3β that can be found in the literature the design of new inhibitors of this enzyme is very important for study of the neurodegenerative diseases [94,95]. The topics reviewed, discussed, and/or reported in this paper are:

1. Studies of GSK-3α inhibitors

1.1. 2D-QSAR for 3-anilino-4-phenylmaleimides

1.2. 3D-QSAR and docking of 3-anilino-4-phenylmaleimides

1.3. QSAR studies of Some GSK-3α Inhibitory pyrimidines

2. Studies of GSK-3β inhibitors

2.1. Design, synthesis and structure-activity relationships of 1,3,4-oxadiazole derivatives

2.2. Linear/Nonlinear Regression Methods for Prediction of Glycogen Synthase Kinase-3β Inhibitory Activities

2.3. Molecular modeling, docking and 3D-QSAR studies for maleimides

2.4. Molecular Docking and biological testing of new GSK-3β inhibitors

2.5. 3D-QSAR Modeling of Paullones

2.6. Modeling of Binding Mode of Benzo[*e*]isoindole-1,3-diones

3. QSAR studies of GSK-3β

3.1. Theoretical study of GSK-3β: Neural Networks QSAR studies

Studies of GSK-3α Inhibitors

2D-QSAR for 3-anilino-4-phenylmaleimides

Sivaprakasam et al. [31] reported in their study a 2D-QSAR exploration of the physicochemical (hydrophobic, electronic, and steric) and structural requirements among 3-anilino-4-phenylmaleimides toward GSK-3α binding. Using Fujita-Ban and Hansch QSAR analyses, electronic and steric interactions at the 4-phenyl ring and hydrophobic
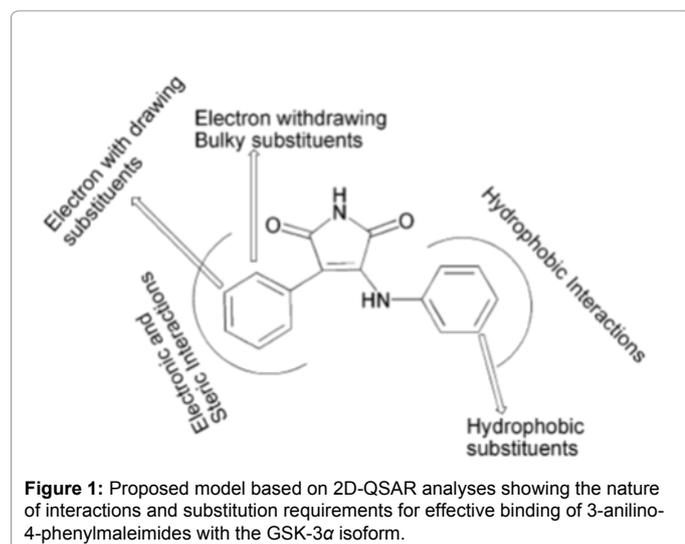
interactions at the 3-anilino ring were shown to be crucial. Hansch-type QSAR was still widely used in the lead optimization stage of synthetic and other projects.

Fujita-Ban analysis of 3-anilino-4-phenylmaleimides revealed that certain structural features such as Cl, OCH_3, and NO2 mono substitution at any position around the 4-phenyl ring were favorable for GSK-3α inhibition. Substituents at the 3-anilino ring such as 3-Cl, 4-Cl, 5-Cl, 3-COOH, 4-OH, and 4-SCH$_3$ were positively and 3-OH was negatively correlated with GSK-3α inhibitory activity.

Through Hansch QSAR analyses, they found that the GSK-3α inhibitory activity was enhanced by: 1. Electron-withdrawing, bulky *ortho* substituents at 4-phenyl ring; 2. 4-chloro substitution around anilino ring; 3. 3-anilino rather than 3-N-methylanilino derivatives; 4. Hydrophobic *meta* substituents on the anilino ring. Overall, QSAR models 13a and 14a suggested electronic and steric effects at the 4-phenyl ring and hydrophobic effects at the 3-anilino or 3-N-methylanilino ring were crucial. Their 2D-model (Figure 1) illustrated these effects which are essential for binding the maleimides to the GSK-3α enzyme. Their analysis provided key information regarding ligand–target interactions which they believed would help medicinal chemists to design more potent GSK-3α inhibitors.

3D-QSAR and docking of 3-anilino-4-phenylmaleimides

3D-QSAR analyses were reported in this article [96], using CoMFA and CoMSIA and molecular docking studies on 3-anilino-4-phenylmaleimides as GSK-3α inhibitors, in order to better understand the mechanism of action and structure-activity relationship of these compounds. The comparison of the active site residues of GSK-3α showed that all the key amino acids involved in polar interactions with the maleimides for the β isoform were the same in the α isoform, except for Asp133 in the β isoform, which was replaced by Glu196 in the α isoform. The authors prepared a homology model for GSK-3α and showed that the change from Asp to Glu should not affect maleimide binding significantly. Their best CoMFA model contained steric and electrostatic fields and had $n = 56$, $q^2 = 0.844$, $r^2 = 0.942$, $SEE = 0.104$, $F = 162.49$ and $r^2_{pred} = 0.779$ for five components. CoMFA electrostatic contours revealed that increased negative charge at the *meta* position of the 4-phenyl ring was favorable for the activity. They found that electron withdrawing groups at the *meta* and *para* positions around the anilino ring were important for enhancing activity.

Figure 1: Proposed model based on 2D-QSAR analyses showing the nature of interactions and substitution requirements for effective binding of 3-anilino-4-phenylmaleimides with the GSK-3α isoform.

Electron-withdrawing bulky *ortho* substituents on the 4-phenyl ring were conducive to GSK-3α inhibition. CoMSIA model showed the importance of hydrogen bond donor groups on these ligands for enhanced activity. The best CoMSIA model (S + E + D) had $n = 56$, $q^2 = 0.833$, $r^2 = 0.932$, $SEE = 0.113$, $F = 111.67$ and $r^2_{pred} = 0.803$ for six components. Comparatively, 3-N-methylanilino derivatives were less active than 3-anilino derivatives.

Docking studies revealed the binding poses of three subclasses of these ligands, namely anilino, N-methylanilino and indoline derivatives, within the active site of the β isoform, and helped to explain the difference in their inhibitory activity.

QSAR studies of some GSK-3α inhibitory pyrimidines

Jamloky et al. studied in this paper [22] a series of pyrimidines which was performed to gain structural insight into the binding mode of the molecules to the GSK-3α. The molecular modeling studies were performed using CS Chem. Office 2001 molecular modeling software version 6.0. MOPAC module was used to minimize the energy and calculate the descriptors. The thermodynamic and steric features of the pyrimidines were highly correlated with GSK-3α inhibitory activity. The positive coefficient of PMI-Y in the model suggested that the presence of bulky substituents positioned towards the Y-axis of the molecule would enhance the GSK-3α inhibitory activity. The observation supports the hypothesis that the presence of the bulky substituents like bromine with inherent hydrophobic character may be involved in the nonspecific interaction with the ATP binding site. The results of the study suggested that the introduction of bulky groups at C-5 position of the hydrophobic interaction with the ATP binding site of the enzyme may be attributed to the strain exerted by the two adjacent phenyl rings on the planar pyrazolo (3,4-*b*) pyridine ring, thereby partly disrupting the hydrogen bonding interaction between nitrogen in the pyrazolo group and the complementary group in the enzyme.

Studies of GSK-3β Inhibitors

Design, synthesis and structure-activity relationships of 1,3,4-oxadiazole derivatives

Saitoh et al. [97] reported design, synthesis and structure–activity relationships of a novel series of oxadiazole derivatives as GSK-3β inhibitors. Among these inhibitors, compound 20x showed highly selective and potent GSK-3β inhibitory activity *in vitro* and its binding mode was determined by obtaining the X-ray co-crystal structure of 20x (Figure 2) and GSK-3β (Figure 3). The hydrogen bonding interaction of the benzimidazole core with the hinge region and the oxadiazole with Asp200 were observed. Additionally, the interaction of 4-methoxyphenyl group with Arg141 was also observed.

Linear/nonlinear regression methods for prediction of glycogen synthase Kinase-3β inhibitory activities

Freitas et al. [98] applied linear/nonlinear regression methods as multiple linear regression (MLR), artificial neural network (ANN), and support vector machines (SVM) with a series of glycogen synthase kinase-3β (GSK-3β) inhibitors using calculated Dragon descriptors. Few variables were selected from a pool of calculated Dragon descriptors through three different feature selection methods, namely genetic algorithm (GA), successive projections algorithm (SPA), and fuzzy rough set ant colony optimization (fuzzy rough set ACO). The fuzzy rough set ACO/SVM-based model gave the best estimation/prediction results, demonstrating the nonlinear nature of this analysis
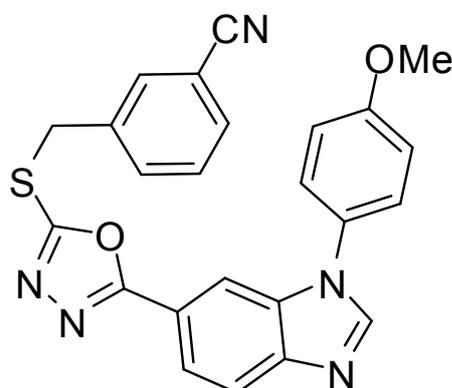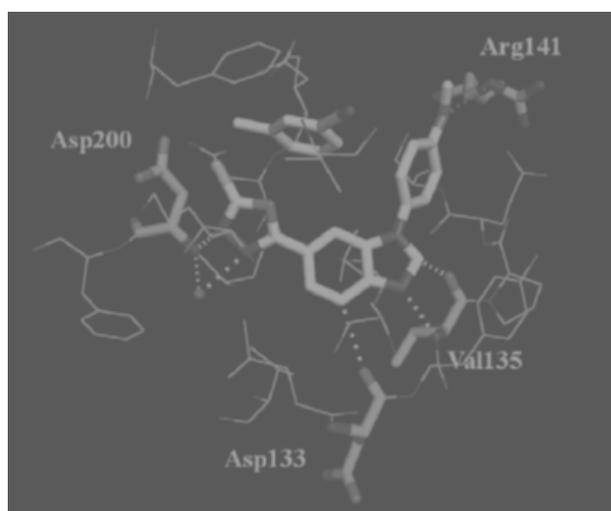
Figure 2: Structure of 20x.

Figure 3: X-ray co-crystal structure of 1 in complex with GSK-3β.

and suggesting fuzzy rough set ACO, introduced in chemistry for the first time, as an improved variable selection method in QSAR for the class of GSK-3β inhibitors. MLR yielded QSAR models only reasonably predictable, with r^2 ranging from 0.77 to 0.81 and r^2_{test} of 0.67 to 0.76, ANN and specially SVM were capable of estimating and predicting biological activities very accurately.

Molecular modeling, docking and 3D-QSAR studies for maleimides

Hwan-Kim et al. [99] carried out molecular modeling and docking studies with three-dimensional quantitative structure relationships (3D-QSAR) to determine the correct binding mode of glycogen synthase kinase 3β (GSK-3β) inhibitors. For the 3D-QSAR (CoMFA and CoMSIA), they used 51 substituted benzofuran-3-yl-(indol-3-yl) maleimides. Two binding modes of the inhibitors to the binding site of GSK-3β were analyzed. The binding mode 1 yielded better 3D-QSAR correlations using both CoMFA and CoMSIA methodologies. The three-component CoMFA model from the steric and electrostatic fields for the experimentally determined pIC$_{50}$ values had the following statistics: $R^2(cv) = 0.386$ and $SE(cv) = 0.854$ for the cross-validation, and

$R^2 = 0.811$ and $SE = 0.474$ for the fitted correlation. $F_{(3.47)} = 67.034$, and probability of $R^2 = 0$ (3.47) = 0.000. The binding mode suggested by the results of this study was consistent with the preliminary results of X-ray crystal structures of inhibitor-bound GSK-3β. The 3D-QSAR models were used for the estimation of the inhibitory potency of two additional compounds.

Molecular docking and biological testing of new GSK-3β inhibitors

Lavrovskii et al. [100] used a series of new heteroaryl-substituted oxadiazole-5-carboxamide inhibitors of GSK-3β. Molecular docking was used for the rational selection of synthesized compounds for the subsequent biological testing. It was established that the inhibitory activity of the synthesized compounds strongly depends on the character of substituents in the phenyl ring and the nature of terminal heterocyclic fragments. The most active compounds inhibit GSK-3β at IC$_{50}$ in the micro molar range and could be considered as potential drug candidates.

3D-QSAR modeling of paullones

Osolodkin et al. [101] carried out a 3D-QSAR study which suggested ways of modification of the molecule to increase its physiological activity. A comparative molecular field analysis (CoMFA) [7] and a comparative molecular similarity indices analysis (CoMSIA) [8] are among the most widely used 3D-QSAR methods. The energy of Van der Waals and electrostatic interactions of a probe atom (with the charge +1) with molecules of the training set (CoMFA) or the electrostatic, Van der Waals, hydrophobic, and donor/acceptor similarity indices (CoMSIA) were used as descriptors. The equation for activity prediction was derived using the partial least squares (PLS) method. The advantages of the methods were the ability of graphic representation of PLS model coefficients and the fact that they allowed the user to suggest substitutions affecting activity and/or selectivity of the molecules. The authors built a new 3D-QSAR model for GSK-3β inhibition by paullones by means of the CoMFA method. This model can be used as a guide for designing new paullone GSK-3β inhibitors.

Modeling of binding mode of Benzo[*e*]isoindole-1,3-diones

Yang et al. [102] synthesized benzo[*e*]isoindole-1,3-dione derivatives and the effects on GSK-3β activity and zebrafish embryo growth were evaluated. A series of derivatives showed obvious inhibitory activity against GSK-3β. The most potent inhibitor, 7,8-dimethoxy-5-methylbenzo[*e*]isoindole-1,3-dione, showed nanomolar IC$_{50}$ and obvious phenotype on zebrafish embryo growth associated with the inhibition of GSK-3β at low micro molar concentration. The interaction mode between this compound and GSK-3β was characterized by computational modeling. To rationalize the structure-activity relationships of these compounds, the binding modes of the most potent inhibitors 8a and 8b (Figure 4) were modeled using docking simulations. Compounds 8a and 8b were docked into the ATP binding site of GSK-3β, and the binding modes of the lowest energy were analyzed. Compounds 8a and 8b fit the ATP pocket of GSK-3β well. The maleimide motif of type II formed a pair of hydrogen bonds with the hinge region (Glu133 and Val135) of GSK-3β, similar to the binding mode of other known maleimides GSK-3β inhibitors. The two methoxy oxygen atoms formed another two hydrogen bonds with the positively charged Lys85. The methyl group of the methoxy at C-8 position docked to the small back cleft of GSK-3β. This binding mode explicitly explained the important role of the two methoxy groups at C-7 and C-8 positions. Another result was the 4-ethyl group of 8b
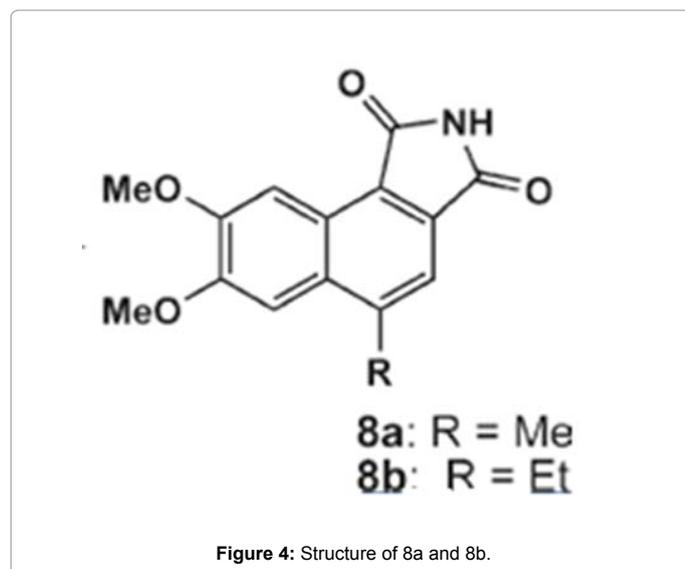
Figure 4: Structure of 8a and 8b.

docks to the minor hydrophobic pocket made up of Ile62 and Val70 in front of the ATP binding site of GSK-3β (Figure 5), which contributed to its higher binding affinity compared to 8a. The docking results also provided a template to understand the structure-activity relationships of other compounds.

QSAR Studies of GSK-3β

Theoretical study of GSK-3β: Neural Networks QSAR studies for the design of new inhibitors using 2D-descriptors

Alzheimer´s disease [9] is a serious and degenerative disorder that causes a gradual loss of neurons, and in spite of the efforts realized by the big pharmaceutical companies of the world, the origin of this pathology is still not very clear. β-amyloid (Aβ) is an important protein implicated in the pathogenesis of AD, but the mechanism by which it causes neurotoxicity is still unknown [10,11]. In particular, there are few literatures report to study the direct link between the pathological hyperphosphorylation of tau protein, a microtubule-associated protein, and the formation of neurofibrillary tangles (NFT) [12]. The last decades had marked a very significant era of AD research. During this period, the nature of amyloid plaques and NTFs, the two histopathological hallmarks of AD, had been elucidated. Recent research efforts have led to several hypotheses to explain AD. Amyloid β toxicity is believed to play a primary role in the development of AD [13]. GSK-3β activity may increase with aging [14], which is consistent with the fact that aging is the most important risk factor for AD. Both *in vitro* and *in vivo* studies have demonstrated that inhibition of GSK-3β, can reverse hyperphosphorylation of tau and prevent behavioral impairments in mice [15-20]. These studies make GSK-3β inhibition very attractive as a therapeutic target for AD [21].

We developed quantitative structure-activity relationships (QSAR) models for GSK-3β, linear discriminant analysis (LDA) [91] and linear artificial neural networks (ANNs) from more than 40000 cases with more than 24000 different molecules inhibitors of GSK-3β obtained from ChEMBL database http://www.ebi.ac.uk/chembldb/index.php/target/browser/classification [92,93]; in total we used more than 45000 different molecules to develop the QSAR models. In addition, we did a study of different fragments that exist in the molecules of the database

in order to see which fragments had more influence in the activity, and which fragments interact more with the protein. As there many studies with GSK-3β that can be found in the literature the design of new inhibitors of this enzyme is very important for the study of neurodegenerative diseases [94,95].

Methods

Linear classifier

A database from ChEMBL database [92] containing assayed GSK-3β inhibitors was used (Table SM from the Supplementary Material). The DRAGON software 4.0 [14] was utilized here and provides 1664 descriptors classified as zero- (0D) one- (1D), two- (2D) and three-dimensional (3D) descriptors depending on the fact they are computed from the chemical formula, substructure list representation, molecular graph or geometrical representation of the molecule, respectively [103]. In this work, we calculated the following descriptors: 2D autocorrelations, Burden eigenvalues, topological charge indices, eigenvalue-based indices, functional group counts, atoms-centred fragments, charge descriptors and molecular properties. The QSAR model was constructed with the multivariate regression technique, the LDA, employing the Forward stepwise method for the selection of variables. All statistical analyses and data exploration were carried out in STATISTICA 6.0 [104]. In the actual work, the independent data test is used by splitting the data randomly in a training series used for a model construction and a cross-validation (CV) one. The general formula of the QSAR classification function is the following:

$$GSKI-3\beta_{score} = \sum W_m \cdot {}^m 2D_i + W_0 \qquad (1)$$

where GSKI-3β$_{score}$ is the continuous and dimensionless score value for the GSKI-3β/non-GSKI-3β classification that gives relatively higher values to molecules with more probability to act as GSKI-3β, ^m2D_i are the $2Ds$ of type m, W_m is the coefficient (weights) of these indices in the QSAR model and W_0 is the independent term.

The reported statistical parameters of the QSAR model are the following: N, χ^2, F, and p-level as well as Sensitivity, Specificity, and Accuracy for both training and CV [104]. N is the number of molecules used to train the model, λ is Wilks statistic parameter, χ^2 is Chi-square and p-level is the probability of error.
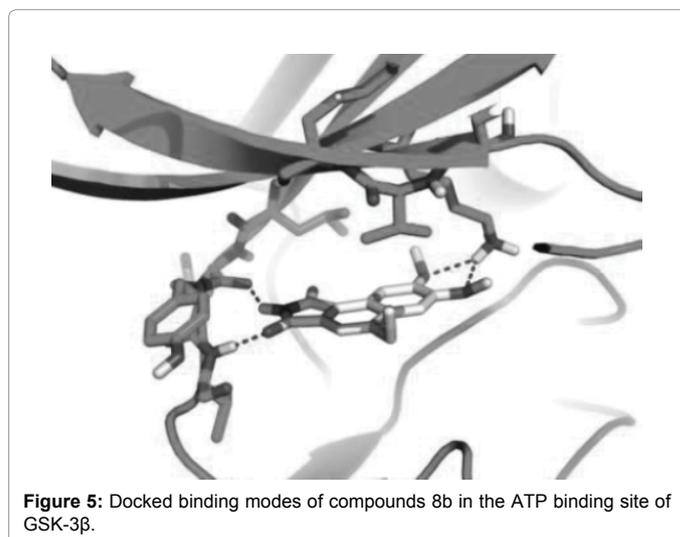

Figure 5: Docked binding modes of compounds 8b in the ATP binding site of GSK-3β.

Nonlinear classifiers

We processed our data with different ANNs using the STATISTICA 6.0 software [104] looking for a better model to predict activity against GSK-3β. Five types of ANNs were used, namely, Probabilistic Neural Network (PNN), Radial Basis Function (RBF) [105], Three Layers Perceptron (MLP-3), and Four Layer Perceptron (MLP-4) and Linear (LNN). The profile of a ANN is: Ni:I-H1-H2-O:No. It means that we have inputs variables (Ni), neurons in the input layer (I), neurons in the first hidden layer (H1), in the second hidden layer (H2), neuron in the output layer (O) and output variable (No).

We can used a very simple type of ANN called Linear Neural Network (RBF) 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 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 RBF classification function, determined by the ANN module of the STATISTICA 6.0 software package [104]. We used Forward Stepwise algorithm for a variable selection.

Let be $^k\chi(G)$ drugs molecular descriptors and $^k\xi(R)$ receptor or drug target descriptors for different drugs (d) with different receptor; we can attempt to develop a simple linear classifier of mt-QSAR type with the general formula:

$$S\left(DTP\right)_{pred} = \sum_{k=0}^{5} b_i\left(G_k\right) \cdot {}^k\chi\left(G\right) + b_0 \qquad (2)$$

We assessed the quality of models with different statistical parameters like Specificity (Equation 2), Sensitivity (Equation 3), Accuracy (Equation 4) and ROC curve (Receiver Operating Characteristic curve) which is a graphical plot of the sensitivity, or true positives, vs. (1−specificity), or false positives,

$$specificity = \frac{NTN}{NTN + NFP} \qquad (3)$$

$$sensitivity = \frac{NTP}{NTP + NFN} \qquad (4)$$

$$accuracy = \frac{NTP + NTN}{NTP + FN + FP + TN} \qquad (5)$$

where NTN means number of true negatives, NFP is number of false positives, NTP is number of true positives, NFN is number of false negatives, FN is false negatives, FP is false positives and TN is true negatives.

The data set used in this article was obtained from ChEMBL database [92,93]. It has more than 56000 cases and more than 24000 different compounds inhibitors of GSK-3β. In total we used more than 45000 different molecules to develop the QSAR models obtained in ChEMBL. This is a database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data). ChEMBL

normalises the bioactivities into a uniform set of end-points and units where possible, and also tags the links between a molecular target and a published assay with a set of varying confidence levels. The data is abstracted and curated from the primary scientific literature, and covers a significant fraction of the structure activity relationship (SAR) and discovery of modern drugs. The codes and activity for all compounds as well as the references used to collect them are depicted in Table SM of the supplementary material file.

Results and Discussion

LDA

In this paper we obtained a LDA study with Equation 6, and we can observe that eighteen variables entry inside equation:

$$GSKI - 3\beta_{score} = -18.1 \cdot D1 + 18.9 \cdot D2 + 0.7 \cdot D3 + 22.0 \cdot D4 - 37.7 \cdot D5 + 11.7 \cdot D6 + 6.6 \cdot D7 - 8.5 \cdot D8 - 20.9 \cdot D9 + 5.4 \cdot D10 + 7.2 \cdot D11 - 2.4 \cdot D12 + 1.2 \cdot D13 + 14.3 \cdot D14 - 14.6 \cdot D15 - 10.8 \cdot D16 + 12.4 \cdot D17 + 20.1 \cdot D18 - 9.4 \cdot D19 + 13.6 \cdot D20 + 8.8 \cdot D21 + 14.1 \cdot D22 - 168.3 \cdot D23 - 122.8 \cdot D24 - 2.8 \cdot D25 + 0.1 \cdot D26 - 0.7 \cdot D27 - 86.4 \qquad (6)$$

$$N = 45{,}299 \qquad \lambda = 0.64 \qquad \chi^2 = 12201.95 \qquad p-level < 0.001$$

In Table 1, we show the code names of descriptors used in the equation 6. The nomenclature used in the descriptors of the equation is the same as establishing the Dragon software, where N is the number of compounds used for training, λ is the Wilks statistic parameter, χ^2 is the Chi-square and p is the level of error. The model correctly classified 1310 out of 1643 active compounds (79.7%) and 24823 out of 26156 non-active compounds (94.9%) in the training series. The overall training performance was 94.0%. Validation of the model was carried out using an external predicting series. In this series the model classified correctly 757 out of 940 (80.5%) compounds and 14166 out of 14937 non-active compounds (94.8%). The overall predictability performance was 94.0% (Table 2).

ANN models

The ANN models are non-linear models useful to predict the biological activity of a large datasets of molecules. This technique is an alternative to linear methods such as LDA [106,107]. Figure 6 depicts the networks maps for some of the ANN models. In general, at least one ANN of every types tested was statically significant. However, one must note that the profiles of each network indicate that these are highly nonlinear and complicated models [108-110].

In Figure 7, we depict the ROC-curve [111,112] for RBF tested. Notably, almost model presented and an area under curve higher than 0.5 (the value for a random classifier). The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before [113]; see, for instance, the work of Fernandez and Caballero [114]. The same is true about the ANNs tested, where is illustrated ROC-curves of ANN RBF with an area higher than 0.99. To show how important is this result, we compared the present model with other model used to address the same problem. We processed our data with ANNs looking for a better model. In general, the ANN RBF tested was statically significant [107].

The network found was RBF and it showed training performance higher than 94.2%. The summary of results is showed in Table 2. After direct inspection of the results reported in Table 2 for ANN methods, we can conclude that a complex ANN method is a good method to predict the activity. We compare different types of networks to obtain a better model; Table 2 shows the classification matrix of the different networks. RBF 166:166-402-1:1 was taken as the main network because it presented a wider range of variables, 166 inputs in the first layer and 166 neurons in second layer, and two sets of cases (Training and Validation). Another tested networks found were LNN 233:233-1:1 and

Definition	Name Descriptor
D1	ATS1m
D2	ATS2m
D3	ATS8m
D4	ATS3v
D5	ATS3e
D6	MATS3m
D7	MATS4m
D8	MATS3e
D9	MATS2p
D10	GATS1v
D11	GATS4v
D12	GATS1e
D13	GATS7e
D14	GATS3p
D15	BELm4
D16	BELm5
D17	BELv2
D18	BEHe1
D19	BEHe8
D20	BELe5
D21	BELe8
D22	BELp4
D23	JGI4
D24	JGI7
D25	Ui
D26	AMR
D27	MLOGP

Table 1: Code names of the different molecular descriptor used in the equation 6.

Model	Train			Stat.	Validation		
profile	Active	Non-Active	%	Par.	%	Active	Non-Active
	1310	333	79.9	Sn	80.5	757	183
LDA	1333	24823	94.9	Sp	94.8	771	14166
			94.0	Ac	94.0		
RBF	1552	100	94.0	Sn	94.3	889	53
166:166-402-1:1	1572	25613	94.2	Sp	94.1	909	14611
			94.0	Ac	94.2		

Table 2: Comparison of LDA and ANN classification model.

LNN 232:232-1:1 presented low accuracy and PNN 233:233-20619-2-2:1 had a very low percentage of DTPs leading to possible errors in the model although its accuracy was very good, (Table 1). We depict the ROC-curve for RBF 166:166-402-1:1 to show how reliable was the network model developed, (Figure 7).

Conclusion

The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors. Nowadays, theoretical studies such as QSAR models have become a very useful tool in this context to substantially reduce time and resources consuming experiments. In this work we developed a new LDA model using the Dragon descriptors, with a large data base using about 20000 different drugs obtained from the ChEMBL server. We conclude that a large database gives a much more precise model; the use of tools such as ChEMBL database enables us to develop
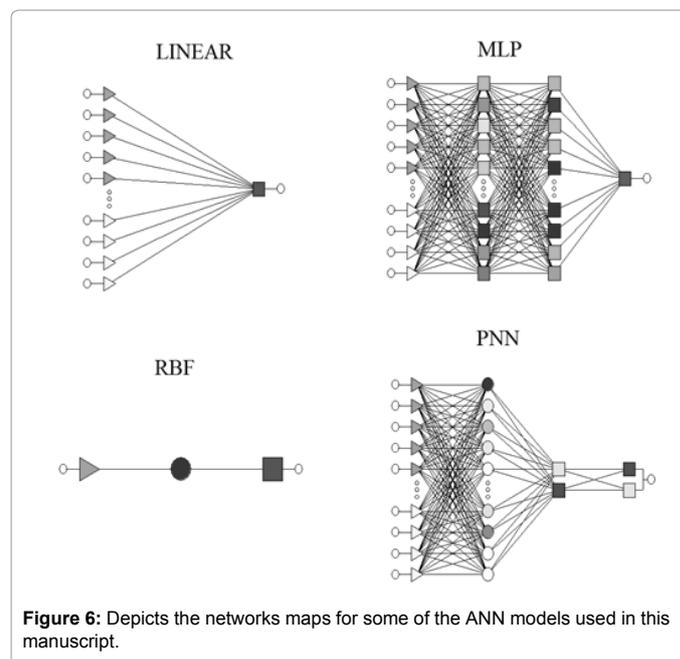

Figure 6: Depicts the networks maps for some of the ANN models used in this manuscript.
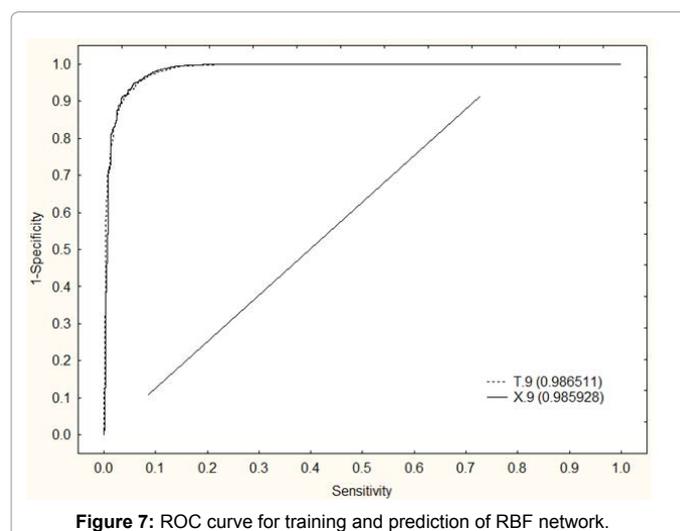

Figure 7: ROC curve for training and prediction of RBF network.

models with large data bases, and this helps us to make the results more reliable. To improve the model we developed non-linear models and compared them to LDA. We proposed non-linear models, and for the first time, we proposed ANN models based on Dragon Descriptors series of GSK-3β, and we concluded that they are alternative methods to study the activity of different families of molecules compared with other methods found in the literature.

References

1. Woodgett JR (1990) Molecular cloning and expression of glycogen synthase kinase-3/factor A. EMBO J 9: 2431-2438.

2. Cai F, Wang F, Lin FK, Liu C, Ma LQ, et al. (2008) Redox modulation of long-term potentiation in the hippocampus via regulation of the glycogen synthase kinase-3beta pathway. Free Radic Biol Med 45: 964-970.

3. Woodgett JR (1991) cDNA cloning and properties of glycogen synthase kinase-3. Methods Enzymol 200: 564-577.

4. Maiese K, Chong ZZ (2004) Insights into oxidative stress and potential novel therapeutic targets for Alzheimer disease. Restor Neurol Neurosci 22: 87-104.

5. Buescher JL, Phiel CJ (2010) A noncatalytic domain of glycogen synthase kinase-3 (GSK-3) is essential for activity. J Biol Chem 285: 7957-7963.

6. Hoeflich KP, Luo J, Rubie EA, Tsao MS, Jin O, et al. (2000) Requirement for glycogen synthase kinase-3beta in cell survival and NF-kappaB activation. Nature 406: 86-90.

7. Hooper C, Killick R, Lovestone S (2008) The GSK3 hypothesis of Alzheimer's disease. J Neurochem 104: 1433-1439.

8. MacAulay K, Doble BW, Patel S, Hansotia T, Sinclair EM, et al. (2007) Glycogen synthase kinase 3alpha-specific regulation of murine hepatic glycogen metabolism. Cell Metab 6: 329-337.

9. Olson RE (2000) Secretase inhibitors as therapeutics for Alzheimer´s disease. Annu Rep Med Chem 35: 31-40.

10. Deguchi K (2006) Insights into oxidative stress and potential novel therapeutic targets for Alzheimer disease. J Cereb Blood Flow Metab 26: 1263-1273.

11. Stabenfeldt SE, García AJ, LaPlaca MC (2006) Thermoreversible laminin-functionalized hydrogel for neural tissue engineering. J Biomed Mater Res A 77: 718-725.

12. Huang J, Chen YJ, Bian WH, Yu J, Zhao YW, et al. (2010) Unilateral amyloid-beta25-35 injection into the rat amygdala increases the expressions of aberrant tau phosphorylation kinases. Chin Med J (Engl) 123: 1311-1314.

13. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297: 353-356.

14. Wen Y, Planel E, Herman M, Figueroa HY, Wang L, et al. (2008) Interplay between cyclin-dependent kinase 5 and glycogen synthase kinase 3 beta mediated by neuregulin signaling leads to differential effects on tau phosphorylation and amyloid precursor protein processing. J Neurosci 28: 2624-2632.

15. Pérez M, Hernández F, Lim F, Díaz-Nido J, Avila J (2003) Chronic lithium treatment decreases mutant tau protein aggregation in a transgenic mouse model. J Alzheimers Dis 5: 301-308.

16. Engel T, Goñi-Oliver P, Lucas JJ, Avila J, Hernández F (2006) Chronic lithium administration to FTDP-17 tau and GSK-3beta overexpressing mice prevents tau hyperphosphorylation and neurofibrillary tangle formation, but pre-formed neurofibrillary tangles do not revert. J Neurochem 99: 1445-1455.

17. Engel T, Hernández F, Avila J, Lucas JJ (2006) Full reversal of Alzheimer's disease-like phenotype in a mouse model with conditional overexpression of glycogen synthase kinase-3. J Neurosci 26: 5083-5090.

18. Le Corre S, Klafki HW, Plesnila N, Hübinger G, Obermeier A, et al. (2006) An inhibitor of tau hyperphosphorylation prevents severe motor impairments in tau transgenic mice. Proc Natl Acad Sci U S A 103: 9673-9678.

19. Nakashima H, Ishihara T, Suguimoto P, Yokota O, Oshima E, et al. (2005) Chronic lithium treatment decreases tau lesions by promoting ubiquitination in a mouse model of tauopathies. Acta Neuropathol 110: 547-556.

20. Noble W, Planel E, Zehr C, Olm V, Meyerson J, et al. (2005) Inhibition of glycogen synthase kinase-3 by lithium correlates with reduced tauopathy and degeneration in vivo. Proc Natl Acad Sci U S A 102: 6990-6995.

21. Gong CX, Iqbal K (2008) Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease. Curr Med Chem 15: 2321-2328.

22. Jamloki A, Karthikeyan C, Sharma SK, Hari Narayana Murthy NS, et al. (2006) QSAR Studies on Some GSK-3a Inhibitory 6-aryl-pyrazolo-(3,4-b) pyrimidines. Asian Journal of Biochemistry 1: 236-243.

23. Ali A, Hoeflich KP, Woodgett JR (2001) Glycogen synthase kinase-3: properties, functions, and regulation. Chem Rev 101: 2527-2540.

24. Martinez A, Castro A, Dorronsoro I, Alonso M (2002) Glycogen synthase kinase 3 (GSK-3) inhibitors as new promising drugs for diabetes, neurodegeneration, cancer, and inflammation. Med Res Rev 22: 373-384.

25. Martinez A, Alonso M, Castro A, Pérez C, Moreno FJ (2002) First non-ATP competitive glycogen synthase kinase 3 beta (GSK-3beta) inhibitors: thiadiazolidinones (TDZD) as potential drugs for the treatment of Alzheimer's disease. J Med Chem 45: 1292-1299.

26. Eldar-Finkelman H (2002) Glycogen synthase kinase 3: an emerging therapeutic target. Trends Mol Med 8: 126-132.

27. Lee VM, Goedert M, Trojanowski JQ (2001) Neurodegenerative tauopathies. Annu Rev Neurosci 24: 1121-1159.

28. Flaherty DB, Soria JP, Tomasiewicz HG, Wood JG (2000) Phosphorylation of human tau protein by microtubule-associated kinases: GSK3beta and cdk5 are key participants. J Neurosci Res 62: 463-472.

29. Phiel CJ, Wilson CA, Lee VM, Klein PS (2003) GSK-3alpha regulates production of Alzheimer's disease amyloid-beta peptides. Nature 423: 435-439.

30. Bhat RV, Budd Haeberlein SL, Avila J (2004) Glycogen synthase kinase 3: a drug target for CNS therapies. J Neurochem 89: 1313-1317.

31. Sivaprakasam P, Xie A, Doerksen RJ (2006) Probing the physicochemical and structural requirements for glycogen synthase kinase-3alpha inhibition: 2D-QSAR for 3-anilino-4-phenylmaleimides. Bioorg Med Chem 14: 8210-8218.

32. Ishiguro K, Takamatsu M, Tomizawa K, Omori A, Takahashi M, et al. (1992) Tau protein kinase I converts normal tau protein into A68-like component of paired helical filaments. J Biol Chem 267: 10897-10901.

33. Pei JJ, Tanaka T, Tung YC, Braak E, Iqbal K, et al. (1997) Distribution, levels, and activity of glycogen synthase kinase-3 in the Alzheimer disease brain. J Neuropathol Exp Neurol 56: 70-78.

34. Pei JJ, Braak E, Braak H, Grundke-Iqbal I, Iqbal K, et al. (1999) Distribution of active glycogen synthase kinase 3beta (GSK-3beta) in brains staged for Alzheimer disease neurofibrillary changes. J Neuropathol Exp Neurol 58: 1010-1019.

35. Baum L, Hansen L, Masliah E, Saitoh T (1996) Glycogen synthase kinase 3 alteration in Alzheimer disease is related to neurofibrillary tangle formation. Mol Chem Neuropathol 29: 253-261.

36. Lucas JJ, Hernández F, Gómez-Ramos P, Morán MA, Hen R, et al. (2001) Decreased nuclear beta-catenin, tau hyperphosphorylation and neurodegeneration in GSK-3beta conditional transgenic mice. EMBO J 20: 27-39.

37. Hernández F, Borrell J, Guaza C, Avila J, Lucas JJ (2002) Spatial learning deficit in transgenic mice that conditionally over-express GSK-3beta in the brain but do not form tau filaments. J Neurochem 83: 1529-1533.

38. Santana L, Uriarte E, González-Díaz H, Zagotto G, Soto-Otero R, et al. (2006) A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. J Med Chem 49: 1149-1156.

39. Marrero-Ponce Y (2004) 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 44: 2010-2026.

40. Vilar S, Santana L, Uriarte E (2006) Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. J Med Chem 49: 1118-1124.

41. Marrero-Ponce Y, Khan MT, Casañola Martín GM, Ather A, Sultankhodzhaev MN, et al. (2007) Prediction of tyrosinase inhibition activity using atom-based bilinear indices. Chem Med Chem 2: 449-478.

42. Casanola Martin, Marrero Ponce Y, Khan MT, Ather A, Sultan S, et al. (2007) TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. Bio Org Med Chem 15: 1483-1503.

43. Casañola-Martín GM, Marrero-Ponce Y, Khan MT, Ather A, Khan KM, et al. (2007) Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental in vitro assays. Eur J Med Chem 42: 1370-1381.

44. Núñez MB, Maguna FP, Okulik NB, Castro EA (2004) QSAR modeling of the MAO inhibitory activity of xanthones derivatives. Bioorg Med Chem Lett 14: 5611-5617.

45. Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. Wiley VCH.

46. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E (2008) Proteomics, networks and connectivity indices. Proteomics 8: 750-778.

47. Zhao CJ, Dai QY (2009) [Recent advances in study of antinociceptive conotoxins]. Yao Xue Xue Bao 44: 561-565.

48. Jacob RB, McDougal OM (2010) The M-superfamily of conotoxins: a review. Cell Mol Life Sci 67: 17-27.

49. Giuliani A, Di Paola L, Setola R (2009) Proteins as Networks: A Mesoscopic Approach Using Haemoglobin Molecule as Case Study. Curr Proteomics 6: 235-245.

50. Vilar S, González-Díaz H, Santana L, Uriarte E (2009) A network-QSAR model for prediction of genetic-component biomarkers in human colorectal cancer. J Theor Biol 261: 449-458.

51. Concu R, Dea-Ayuela MA, Perez-Montoto LG, Prado-Prado FJ, Uriarte E, et al. (2009) 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. Biochim Biophys Acta 1794: 1784-1794.

52. Torrens F, Castellano G (2009) Topological Charge-Transfer Indices: From Small Molecules to Proteins. Curr Proteomics 6: 204-213.

53. Vázquez JM, Vanessa A, Seoane JA, Freire A, Serantes JA, et al. (2009) Star Graphs of Protein Sequences and Proteome Mass Spectra in Cancer Prediction. Curr Proteomics 6: 275-288.

54. González-Díaz H (2008) Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR). Curr Top Med Chem 8: 1554.

55. Ivanciuc O (2008) Weka machine learning for predicting the phospholipidosis inducing potential. Curr Top Med Chem 8: 1691-1709.

56. González-Díaz H, Prado-Prado F, Ubeira FM (2008) Predicting antimicrobial drugs and targets with the March-Inside approach. Curr Top Med Chem 8: 1676-1690.

57. Duardo-Sánchez A, Patlewicz G, López-Díaz A (2008) Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. Curr Top Med Chem 8: 1666-1675.

58. Wang JF, Wei DQ, Chou KC (2008) Drug candidates from traditional chinese medicines. Curr Top Med Chem 8: 1656-1665.

59. Helguera AM, Combes RD, González MP, Cordeiro MN (2008) Applications of 2D descriptors in drug design: a Dragon tale. Curr Top Med Chem 8: 1628-1655.

60. González MP, Terán C, Saíz-Urra L, Teijeira M (2008) Variable selection methods in QSAR: an overview. Curr Top Med Chem 8: 1606-1627.

61. Caballero J, Fernandez M (2008) 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 8: 1580-1605.

62. Wang JF, Wei DQ, Chou KC (2008) Pharmacogenomics and personalized use of drugs. Curr Top Med Chem 8: 1573-1579.

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

64. Prado-Prado FJ, García-Mera X, González-Díaz H (2010) Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. Bioorg Med Chem 18: 2225-2231.

65. González-Díaz H (2010) Network topological indices, drug metabolism, and distribution. Curr Drug Metab 11: 283-284.

66. Khan MT (2010) Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. Curr Drug Metab 11: 285-295.

67. Mrabet Y, Semmar N (2010) Mathematical methods to analysis of topology, functional variability and evolution of metabolic systems based on different decomposition concepts. Curr Drug Metab 11: 315-341.

68. Martínez-Romero M, Vázquez-Naya JM, Rabuñal JR, Pita-Fernández S, Macenlle R, et al. (2010) Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. Curr Drug Metab 11: 347-368.

69. Zhong WZ, Zhan J, Kang P, Yamazaki S (2010) Gender specific drug metabolism of PF-02341066 in rats--role of sulfoconjugation. Curr Drug Metab 11: 296-306.

70. Wang JF, Chou KC (2010) Molecular modeling of cytochrome P450 and drug metabolism. Curr Drug Metab 11: 342-346.

71. González-Díaz H, Duardo-Sanchez A, Ubeira FM, Prado-Prado F, Pérez-Montoto LG, et al. (2010) 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.

72. García I, Diop YF, Gómez G (2010) QSAR & complex network study of the HMGR inhibitors structural diversity. Curr Drug Metab 11: 307-314.

73. Chou KC (2010) Graphic rule for drug metabolism systems. Curr Drug Metab 11: 369-378.

74. Concu R, Podda G, Ubeira FM, González-Díaz H (2010) Review of QSAR models for enzyme classes of drug targets: Theoretical background and applications in parasites, hosts, and other organisms. Curr Pharm Des 16: 2710-2723.

75. Estrada E, Molina E, Nodarse D, Uriarte E (2010) Structural contributions of substrates to their binding to P-Glycoprotein. A TOPS-MODE approach. Curr Pharm Des 16: 2676-2709.

76. García I, Fall Y, Gómez G (2010) QSAR, docking, and CoMFA studies of GSK3 inhibitors. Curr Pharm Des 16: 2666-2675.

77. González-Díaz H (2010) QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences. Curr Pharm Des 16: 2598-2600.

78. Gonzalez-Diaz H, Romaris F, Duardo-Sanchez A, Pérez-Montoto LG, Prado-Prado F, et al. (2010) Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, aplications, and legal issues. Current Pharmaceutical Design 16: 2737-2764.

79. Marrero-Ponce Y (2010) Ligand-Based Computer-Aided Discovery of Tyrosinase Inhibitors. Applications of the Tomocomd-Cardd Method to the Elucidation of New Compounds. Current Pharmaceutical Design16: 2601-2624.

80. Munteanu CR, Fernández-Blanco E, Seoane JA, Izquierdo-Novo P, Rodríguez-Fernández JA, et al. (2010) Drug discovery and design for complex diseases through QSAR computational methods. Curr Pharm Des 16: 2640-2655.

81. Roy K, Ghosh G (2010) Exploring QSARs with Extended Topochemical Atom (ETA) indices for modeling chemical and drug toxicity. Curr Pharm Des 16: 2625-2639.

82. Speck-Planche A, Scotti MT, de Paulo-Emerenciano V (2010) Current pharmaceutical design of antituberculosis drugs: future perspectives. Current Pharmaceutical Design 16: 2656-2665.

83. Vázquez-Naya JM, Martínez-Romero M, Porto-Pazos AB, Novoa F, Valladares-Ayerbes M, et al. (2010) Ontologies of drug discovery and design for neurology, cardiology and oncology. Curr Pharm Des 16: 2724-2736.

84. Chou KC (2004) Structural bioinformatics and its impact to biomedical science. Curr Med Chem 11: 2105-2134.

85. Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ (2006) Progress in computational approach to drug development against SARS. Curr Med Chem 13: 3263-3270.

86. Prado-Prado FJ, González-Díaz H, de la Vega OM, Ubeira FM, Chou KC, et al. (2008) Unified QSAR Approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. Bioorg Med Chem 16: 5871-5880.

87. Prado-Prado FJ, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou KC, et al. (2009) Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. Bioorg Med Chem 17: 569-575.

88. Vanyúr R, Héberger K, Jakus J (2003) Prediction of anti-HIV-1 activity of a series of tetrapyrrole molecules. J Chem Inf Comput Sci 43: 1829-1836.

89. Vilar S, Santana L, Uriarte E (2006) Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. J Med Chem 49: 1118-1124.

90. Mosier PD, Jurs PC (2002) QSAR/QSPR studies using probabilistic neural networks and generalized regression neural networks. J Chem Inf Comput Sci 42: 1460-1470.

91. Prado-Prado FJ, Borges F, Perez-Montoto LG, González-Díaz H (2009) Multi-

target spectral moment: QSAR for antifungal drugs vs. different fungi species. Eur J Med Chem 44: 4051-4056.

92. Overington J (2009) ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. J Comput Aided Mol Des 23: 195-198.

93. Welchman R (2010) Advances and Progress in Drug Design - SMi's ninth annual meeting. IDrugs 13: 239-242.

94. Prado-Prado FJ, Ubeira FM, Borges F, González-Díaz H (2010) Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. J Comput Chem 31: 164-173.

95. Prado-Prado FJ, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou KC, et al. (2009) Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. Bioorg Med Chem 17: 569-575.

96. Prasanna S, Daga PR, Xie A, Doerksen RJ (2009) Glycogen synthase kinase-3 inhibition by 3-anilino-4-phenylmaleimides: insights from 3D-QSAR and docking. J Comput Aided Mol Des 23: 113-127.

97. Saitoh M (2009) Design, synthesis and structure–activity relationships of 1,3,4-oxadiazole derivatives as novel inhibitors of glycogen synthase kinase-3beta. Bioorganic & Medicinal Chemistry 17: 2017-2029.

98. Freitas MP, M Goodarzi, Jensen R (2009) Feature Selection and Linear/Nonlinear Regression Methods for the Accurate Prediction of Glycogen Synthase Kinase-3ß Inhibitory Activities. J Chem Inf Model 49: 824-832.

99. Kim KH, Gaisina I, Gallier F, Holzle D, Blond SY, et al. (2009) Use of molecular modeling, docking, and 3D-QSAR studies for the determination of the binding mode of benzofuran-3-yl-(indol-3-yl)maleimides as GSK-3beta inhibitors. J Mol Model 15: 1463-1479.

100. Ryzhova EA, Koryakova AG, Bulanova EA, Mikitas OV, Karapetyan RN (2009) Syntheis, Molecular Docking, and Biological Testing of New Selective Inhibitors of Glycogen Synthase Kinase 3beta. Pharmaceutical Chemistry Journal 43: 148-153.

101. Osolodkin DI, Shulga DA, Tsareva DA, Oliferenko AA, Palyulin VA, et al. (2010) The choice of atomic charges calculation scheme in 3D-QSAR modelling of GSK-3β inhibition by paullones. Dokl Biochem Biophys 434: 274-278.

102. Zou H, Zhou L, Li Y, Cui Y, Zhong H, et al. (2010) Benzo[e]isoindole-1,3-diones as potential inhibitors of glycogen synthase kinase-3 (GSK-3). Synthesis, kinase inhibitory activity, zebrafish phenotype, and modeling of binding mode. J Med Chem 53: 994-1003.

103. Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. Wiley VCH.

104. StatSoft.Inc, Statistica (2002) (data analysis software system) version 6.0, www.statsoft.com.Statsoft, Inc.

105. Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou O, Alexandridis A (2006) A novel RBF neural network training methodology to predict toxicity to Vibrio fischeri. Mol Divers 10: 213-221.

106. Concu R, Podda G, Ubeira FM, González-Díaz H (2010) Review of QSAR models for enzyme classes of drug targets: Theoretical background and applications in parasites, hosts, and other organisms. Curr Pharm Des 16: 2710-2723.

107. Roy K, Mandal AS (2008) Development of linear and nonlinear predictive QSAR models and their external validation using molecular similarity principle for anti-HIV indolyl aryl sulfones. J Enzyme Inhib Med Chem 23: 980-995.

108. Roy K, Mandal AS (2009) Predictive QSAR modeling of CCR5 antagonist piperidine derivatives using chemometric tools. J Enzyme Inhib Med Chem 24: 205-223.

109. Roy K, Pratim Roy P (2009) Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. Eur J Med Chem 44: 2913-2922.

110. Patra JC, Singh O (2009) Artificial neural networks-based approach to design ARIs using QSAR for diabetes mellitus. J Comput Chem 30: 2494-2508.

111. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29-36.

112. González-Díaz H, Bonet I, Terán C, De Clercq E, Bello R, et al. (2007) ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. Eur J Med Chem 42: 580-585.

113. Agüero-Chapin G, Varona-Santos J, de la Riva GA, Antunes A, González-Vlla T, et al. (2009) Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from Coffea arabica and prediction of a new sequence. J Proteome Res 8: 2122-2128.

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