Article

Using Topological Indices to Predict Anti-Alzheimer and Anti-Parasitic GSK-3 Inhibitors by Multi-Target QSAR *in Silico* Screening

Isela García *, Yagamare Fall and Generosa Gómez

Department of Organic Chemistry, Faculty of Chemistry, University of Vigo, Spain

* Author to whom correspondence should be addressed; E-Mail: iselapintos@yahoo.es;
Tel.: +34-986-813-679; Fax: +34-986-812-262.

Abstract: *Plasmodium falciparum*, *Leishmania*, *Trypanosomes*, are the causers of diseases such as malaria, leishmaniasis and African trypanosomiasis that nowadays are the most serious parasitic health problems worldwide. The great number of deaths and the few drugs available against these parasites, make necessary the search for new drugs. Some of these antiparasitic drugs also are GSK-3 inhibitors. GSKI-3 are candidates to develop drugs for the treatment of Alzheimer's disease. In this work topological descriptors for a large series of 3,370 active/non-active compounds were initially calculated with the ModesLab software. Linear Discriminant Analysis was used to fit the classification function and it predicts heterogeneous series of compounds like paullones, indirubins, meridians, *etc*. This study thus provided a general evaluation of these types of molecules.

Keywords: glycogen synthase kinase-3 (GSK-3) inhibitors; Alzheimer's disease; *Plasmodium falciparum*; *Trypanosoma brucei*; *Leishmania*; QSAR

1. Introduction

Neurofibrillary tangles (NFTs) are one of the characteristic neuropathological lesions of Alzheimer's disease (AD) [1] and other neurodegenerative processes such as frontotemporal dementia, Pick's disease, progressive supranuclear palsy, and corticobasal degeneration [2].

Alzheimer's disease affects 5–10% of the population over 65 years of age. The dementia associated with AD results from the selective death of neurons, which is associated with several anatomo-pathological hallmarks such as senile neuritic plaques and neurofibrillary tangles [3,4]. Three molecular actors clearly play a role in the development of AD: the amyloid β peptide (Aβ), presenilins-1 and -2 and the microtubule-associated protein tau.

Glycogen Synthase Kinase-3 (GSK-3) is a senile-threonine kinase ubiquitously expressed and involved in the regulation of many cell functions [5]. GSK-3 was originally identified as one of the five protein kinases that phosphorylate glycogen synthase [6], being implicated in type-2 diabetes [7]. GSK-3 is also known to phosphorylate the microtubule-associated protein tau in mammalian cell [8]. This hyperphosphorylation is an early event in neurodegenerative conditions, such as Alzheimer's disease [9], also involved is a second kinase, called CDK-5 [10].

Two GSK-3 genes (α and β) have been cloned from vertebrates. The intrinsic biochemical properties of GSK-3 are also conserved, with greater differences between GSK-3α and GSK-3β than between species [11].

Interest in GSK-3 has grown far beyond glycogen metabolism during the past decade and GSK-3 is known to occupy a central stage in many cellular and physiological events, including Wnt and Hedgehog signaling, transcription, insulin action, cell-division cycle, response to DNA damage, cell death, cell survival, patterning and axial orientation during development, differentiation, neural functions, circadian rhythm and others. In 1988 Ishiguro *et al.* [12] isolated one enzyme when they were studying an extract of the brain, which showed the generation of paired helical filaments of Tau protein kinase, typical injury of Alzheimer's disease.

In this moment, there is an increasing interest in the evaluation of kinases from unicellular parasites and fungies as targets for potential new anti-parasitic drugs. The evolutionary difference between unicellular kinases and their human homologues might be sufficient to allow the design of parasite-specific inhibitors. The *Plasmodium falciparum* genome contains 65 genes that encode kinases, including three forms of GSK-3. *P. falciparum* is one of the *Plasmodium* species that cause malaria in humans. It is transmitted by the female *Anopheles* mosquito. Malaria has the highest rates of complications and mortality; for example in 2006 it accounted for 91% of all 247 million human malaria infections (98% in Africa) and 90% of the deaths.

Leishmania is a genus of trypanosome protozoa, and is the parasite responsible for the disease leishmaniasis. *Leishmania* commonly infects hyraxes, canids, rodents and humans; and currently affects 12 million people in the world. There are many drawbacks to current chemotherapy for leishmaniasis, including problems of low efficacy, severe toxic side effects, and emerging drug resistance [13-15].

Leishmania parasites posses a complex life cycle in which the parasite passes between the sandfly vector and the mammalian host, during which time the parasite oscillates between rapidly dividing and cell cycle-arrested forms. The cell cycle of *Leishmania* is closely regulated, as in other eukaryotes, and integrated with its differentiation between the various life cycle stages.

Trypanosomes are a group of kinetoplastic protozoa distinguished by having only a single flagellum. All members are exclusively parasitic, found primary in insects. A few genera have life-cycles involving a secondary host, which may be a vertebrate or a plant. These include several species that cause major diseases in humans. African Trypanosomiasis disease is caused by members of the

Trypanosma brucei complex, is a serious health threat. It is estimated that 300,000 to 500,000 humans in sub-Saharan African are infected. Despite the critical need, the available therapies are becoming less satisfactory due to the rising level of resistance to the available drugs, the long period of treatment required to achieve a cure, and the unacceptable and sometimes severe adverse effects associated with current therapies [16]. However, optimization of the selectivity of drug candidates for parasite kinases becomes an issue due to the highly conserved amino acids and protein conformation of the catalytic domains [17-20]. Our main objective in this work was to identify numerous examples of anti-parasitic, anti-fungi, *etc.* and GSK-3 inhibitors in order to obtain a model for the optimization of selected target inhibitors for drug development [21,22].

There are more than one thousand theoretical descriptors available in the literature to represent molecular structures, and one usually faces the problem of selecting those which are the most representative for any particular property under consideration. Topological indices [23-33], the most commonly used molecular descriptors, have been widely used in the correlation of physicochemical properties of organic compounds. In chemical graph theory, molecular structures are normally represented as hydrogen depleted graphs, whose vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulas can be then assimilated to undirected and finite multigraphs with labeled vertices, commonly known as molecular graphs. Topological indices, also known as graph theoretical indices, are descriptors that characterize molecular graphs and contain a large amount of information about the molecule, including the numbers of hydrogen and non-hydrogen atoms bonded to each non-hydrogen atom, the details of the electronic structure of each atom, and the molecular structural features [34]. In this article, we use ModesLab software [35] for calculating Topological indices (see Table 1), which is very important in the area of development of molecular descriptors and its applications to quantitative structure-property (QSPR), quantitative structure-activity (QSAR) relationship and drug design. ModesLab also provides a very useful way to define the properties of atoms, bonds and fragments by an extension of SMILES language and use these properties in molecular descriptors calculations.

Table 1. Topological Indices used in the present study.

Index	Description
$\chi(P)$, $\chi(C)$, $\chi(PC)$, $\chi(Ch)$	Randic branching index
$\chi_v(P)$, $\chi_v(C)$, $\chi_v(PC)$, $\chi_v(Ch)$	Valence connectivity
$e(P)$, $e(C)$, $e(pC)$, $e(Ch)$	Epsilon index
$^1\kappa$, $^2\kappa$, $^3\kappa$	Kappa index
$^1\kappa(alpha)$, $^2\kappa(alpha)$, $^3\kappa(alpha)$	Kappa (alpha) index
$\emptyset$	Flexibility index
M_1	Zagreb M_1 index
M_2	Zagreb M_2 index
H	Harary number
J	Balaban index

The development of QSARs using simple molecular indices appears to be a promising alternative or complementary technique to drug-protein docking, high-throughput screening and combinatorial chemistry techniques. 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 [36-38].

A large number of examples have been published in which the use of molecular descriptors has become in a rational alternative to massive synthesis and screening of compounds in medicinal chemistry [39,40]. The principal deficiency in the use of some molecular indices concerns their lack of physical meaning.

2. Results and Discussion

2.1. General QSAR for GSK-3 Inhibitors

The development of a discriminant function [41] that allows the classification of organic compounds as active or non-active is the key step in the present approach for the discovery of GSK-3 inhibitors. It was therefore necessary to select a training data set of GSK-3 inhibitors containing wide structural variability. Linear Discriminant Analysis (LDA) was used to construct the classifiers. One of the most important steps in this work was the organization of the spreadsheet containing the raw data used as input for the LDA because this is not a classic classifier. Herein, the schematisation of the paper is peculiar. Our expectation is to use a two-group discriminant function to classify compounds into two possible groups: compounds that belong to a particular group and compounds that do not belong to this group. To this end, we have to indicate somehow what group we pretend to predict in each case. In this regard, we made the following steps, these steps are essentially the same given by Concu *et al.* [42,43] for their QSAR study of six classes of enzymes. In this study, each compound may be assayed on q^{th} different sets of conditions in the pharmacological tests, which are defined as Compound Assay Conditions query (CAC_q). These conditions are indicated in the supplementary material. The selection here of discriminant techniques instead of regression techniques was determined by the lack of homogeneity in the conditions under which these values were measured. As reported in different sources, numerous IC_{50} values lie within a range rather than a single value. In other cases, the activity is not scored in terms of IC_{50} values but is quoted as inhibitory percentages at a given concentration. Once the training series had been designed, forward-stepwise Linear Discriminant Analysis (LDA) was carried out in order to derive the QSAR (see **Equation 1**):

$$i - \text{GSK} - 3_{TI} = 12.2755 \cdot {}^6\chi(Ch) + 5.9786 \cdot {}^5e(Ch) + 0.243 \cdot {}^1\kappa - 0.8553 \cdot \phi$$
$$+ 1.9549 \cdot {}^3\chi(C)_{dif} - 1.2747 \cdot {}^5\chi_v(P)_{dif} - 43.1819 \cdot {}^4e(C)_{dif} \quad (1)$$
$$- 1.2034 \cdot {}^5e(P)_{dif} + 3.705 \cdot {}^6e(P)_{dif} - 0.5713 \cdot {}^6e(pC)_{dif} - 0.2127$$
$$\lambda = 1.00 \qquad F = 163.45 \qquad p < 1.00$$

The statistical significance of this model was determined by examining Wilk's λ statistic, Fisher ratio (F), and the p-level (p). The model is based on two types of parameters, the first type are parameters for single molecules. The type one includes first fourth parameters in the model. The following parameters presented in Table 1: randic branching index ($\chi(Ch)$, $\chi(C)$), epsilon index ($e(Ch)$, $e(C)$, $e(P)$, $e(pC)$), kappa index (κ), flexibility index ($\varnothing$) and valence connectivity ($\chi_v(P)$). In addition, we can see in the model parameters that quantify the difference between the structure of the drug and the structure of the drugs active for a given set of conditions CAC_q (see the last six parameters in the

Equation 1a-1f). We quantify this information in terms of the difference between the descriptors (Ds) of the drug and the average of Ds of active drugs for a given condition (see Methods section). The exact formulas for these terms present in the model are:

$$^3\chi(C)_{dif} = {}^3\chi(C) - {}^3\overline{\chi}(C) \quad (1a)$$

$$^5\chi_v(P)_{dif} = {}^5\chi_v(P) - {}^5\overline{\chi}_v(P) \quad (1b)$$

$$^4e(C)_{dif} = {}^4e(C) - {}^4\overline{e}(C) \quad (1c)$$

$$^5e(P)_{dif} = {}^5e(P) - {}^5\overline{e}(P) \quad (1d)$$
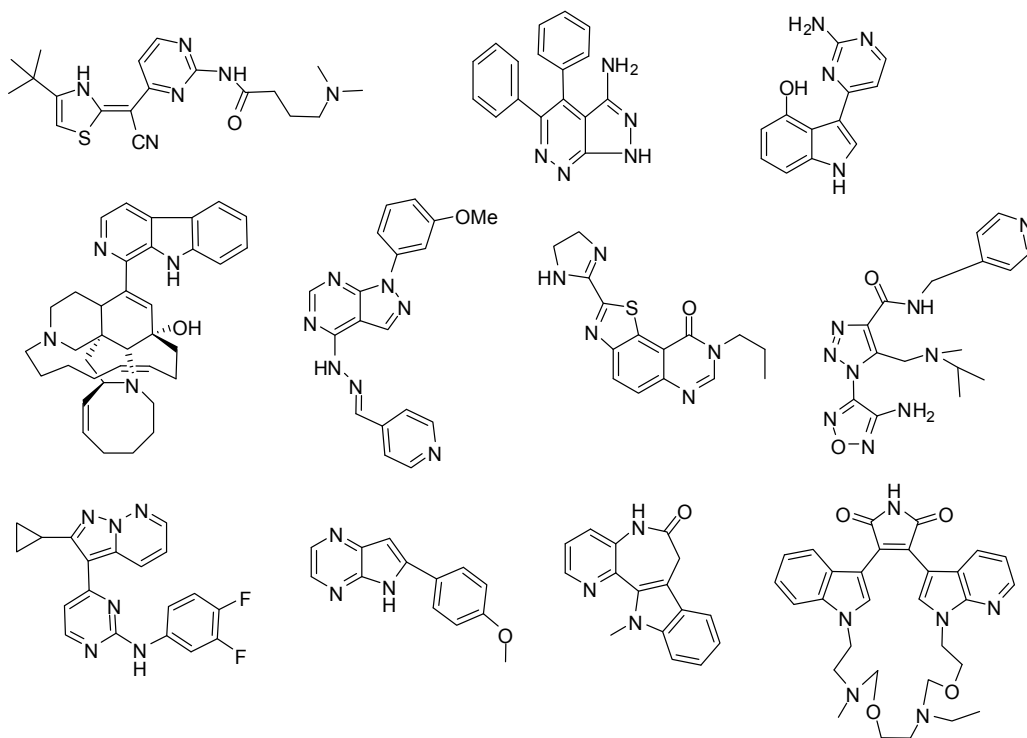
$$^6e(P)_{dif} = {}^6e(P) - {}^6\overline{e}(P) \quad (1e)$$

$$^6e(pC)_{dif} = {}^6e(pC) - {}^6\overline{e}(pC) \quad (1f)$$

The introduction of this last parameter in the model (1) coincides with the i-GSK-3 activity observed in organic acid compounds [44] and drug connectivity may become an interesting alternative for fast computational pre-screening of large series of compounds in order to rationalize synthetic efforts [45-51] complementary to more elaborated techniques 3D-QSAR, CoMFA, and CoMSIA studies that depend on a detailed knowledge of 3D structure. In any case, these present models are of more general application than the other known methods that apply only to compounds tested in only one CAC and/or belonging to only one homogeneous structural class of compounds. Confirmation of this statement comes from the fact that the present classification function has given rise to an efficient separation of all compounds: with Accuracy = 99.1% in training series and Accuracy = 86.8% in validation series for the topological function, see Table 2 for details. The names, observed classification, predicted classification and subsequent probabilities for all 3,370 compounds in training and average validation are given as supplementary material. This level of total Accuracy, Sensitivity and Specificity is considered as excellent by other researches that have used LDA for QSAR studies and taking into account the great variety of compounds (see Figure 1), due to the fact that their structures are very different; see for instance the works of Garcia-Domenech, Prado-Prado and Marrero-Ponce *et al.* [52-67].

Table 2. Training and validation results.
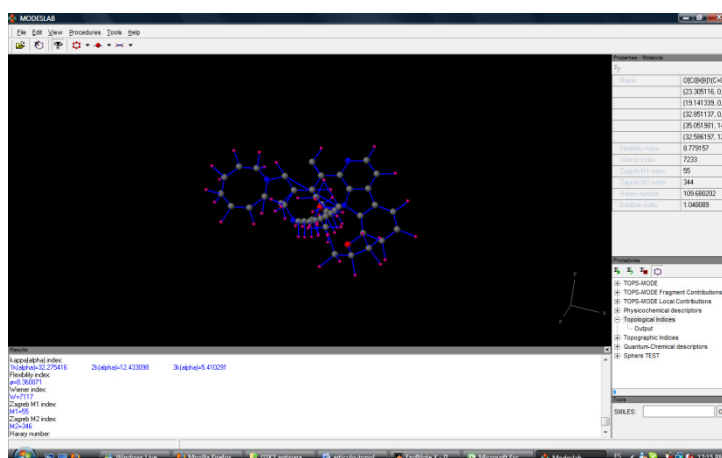
Group	Parameter	%	GSKI-3	Non-active
		Training		
GSKI-3	*Sensitivity*	95.3	854	42
Non-active	*Specificity*	82.8	77	371
Total	*Accuracy*	91.1		
		Validation		
GSKI-3	*Sensitivity*	95.3	282	14
Non-active	*Specificity*	84.6	179	985
Total	*Accuracy*	86.8		

Figure 1. Some compounds studied in this article.

3. Materials and Methods

3.1. Computational Methods

The dataset formed by 3,370 cases was divided into 43 groups depending of their GSK-3 inhibitory activity and antiparasitic, antifungi, *etc.* activity. The model was obtained with topological descriptors. ModesLab (Molecular Descriptor Laboratory) version 1.5 software [35] was used to calculate all descriptors (see Figure 2).

Figure 2. ModesLab software.

The total analyzed variables were 189. The variables $^3\chi(Ch)$, $^4\chi(Ch)$, $^3\chi_v(Ch)$, $^4\chi_v(Ch)$, $^3e(Ch)$, $^4e(Ch)$, $^3\chi(Ch)_{avg}$, $^4\chi(Ch)_{avg}$, $^3\chi_v(Ch)_{avg}$, $^4\chi_v(Ch)_{avg}$, $^3e(Ch)_{avg}$, $^4e(Ch)_{avg}$, $^3\chi(Ch)_{dif}$, $^4\chi(Ch)_{dif}$, $^3\chi_v(Ch)_{dif}$, $^4\chi_v(Ch)_{dif}$, $^3e(Ch)_{dif}$, $^4e(Ch)_{dif}$ was eliminated of the database because all of them were equal to 0 and

constants. The quality of the model was determined by examining the Wilk's statistic, the square of Mahalanobis distance (D^2), the Fisher ratio (F) and the number of variables in the equation. Discrimination functions were obtained by using the forward-stepwise linear discriminant analysis as implemented in Statistica 6.0.

3.2. Multi-target Linear Discriminant Analysis (LDA)

In this regard, we performed the following steps, which are essentially the same given by Concu *et al.* [42,43] for their QSAR study of six classes of enzymes:

(1). We created a raw data representing each compound input as a vector made up of 1 output variable, 189 structural variables (inputs) divided in values (see the first term of the Equation (1)), averages (see the second term of the Equation (1)) and differences between values and averages (see the third term of the Equation (1)); and the CAC_q variable. CAC_q is an auxiliary not used to construct the model.

(2). The first element (output) is a dummy variable (Boolean) called Observed Group (OG); $OG = 0$ if the compound belongs to the class to which we refer in CAC_q and 1 otherwise ($OG = 1$). We could repeat each compound more than once in the raw data. In fact, we could repeat each compound 43 times corresponding to 43 CAC_q Assay Conditions (see **Table 3**). The first time we used the $CAC_q = CAC$ number. It means that we used the real CAC class of the compound in CAC_q. In this case, the LDA model had to give the highest probability to the group $OG = 0$ because it had to predict the real class of the compound. The remnant 43 times we use an CAC class number different to the real in CAC_q and then the LDA model had to predict the highest probability for the group $OG = 1$. This indicated that the compound did not belong to this group.

Table 3. Compound Assay Conditions query (CACq).

Group	*Parameter*	*Enzyme*	*Isoform*	*Enzyme/ Organism*	*Species*	*Activity (1,0)*	*Condition*	*Observ.*
1	%	GSK-3	beta	enzyme	no	0	0	=
2	cKi	GSK-3	alfa	enzyme	no	0	100	>
3	EC_{50} (µM)	no	no	no	Cell Efficacy	0	2	>
4	EC_{50} (µM)	no	no	no	glycogen synthesis stimulation	0	inactive	=
5	EC_{50} (µM)	no	no	no	β-catenin synthesis	0	2	>
6	EC_{50} (µM)	no	no	parasite	*T. brucei*	0	2	>
7	EC_{50} (µM)	no	no	virus	HIV-1	0	NA	=
8	ED_{50} (µM)	no	no	parasite	*L. donovani*	0	5	<
9	IC_{50} (ng/mL)	no	no	parasite	*P. falciparum* (chloroquine resistant W2 clone)	0	NA	=
10	IC_{50} (ng/mL)	no	no	parasite	*P. falciparum* (chloroquine sensitive D6 clone)	0	NA	=

Table 3. *Cont.*

11	IC$_{50}$ (nM)	GSK-3	alfa	enzyme	no	0	2000	>
12	IC$_{50}$ (nM)	GSK-3	beta	enzyme	no	0	2000	>
13	IC$_{50}$ (nM)	GSK-3	nd	enzyme	no	0	2000	>
14	IC$_{50}$ (μg/mL)	no	no	bacterium	*M. intracellulare*	0	NA	=
15	IC$_{50}$ (μg/mL)	no	no	bacterium	MRS	0	NA	=
16	IC$_{50}$ (μg/mL)	no	no	bacterium	*S. aureus*	0	NA	=
17	IC$_{50}$ (μg/mL)	no	no	cell line	Human Vero cells	0	NC	<>
18	IC$_{50}$ (μg/mL)	no	no	fungus	*C. neoformans*	0	NA	=
19	IC$_{50}$ (μg/mL)	no	no	parasite	*L. donovani*	0	NA	=
20	IC$_{50}$ (μM)	GSK-3	beta	enzyme	no	0	2	>
21	IC$_{50}$ (μM)	GSK-3	nd	enzyme	no	0	2	>
22	IC$_{50}$ (μM)	GSK-3	no	parasite	*P. falciparum*	0	20	>
23	IC$_{50}$ (μM)	GSK-3	α/β	enzyme	no	0	2	>
24	IC$_{50}$ (μM)	no	no	bacterium	*M. intracellulare*	0	—	=
25	IC$_{50}$ (μM)	no	no	bacterium	MRSA	0	—	=
26	IC$_{50}$ (μM)	no	no	cell line	Hep2	0	NA	=
27	IC$_{50}$ (μM)	no	no	cell line	HT29	0	NA	=
28	IC$_{50}$ (μM)	no	no	cell line	Human Vero cells	0	NA	=
29	IC$_{50}$ (μM)	no	no	cell line	Human Vero cells	0	NC	<>
30	IC$_{50}$ (μM)	no	no	cell line	LMM3	0	NA	=
31	IC$_{50}$ (μM)	no	no	cell line	PTP	0	2	>
32	IC$_{50}$ (μM)	no	no	fungus	*C. albicans*	0	—	=
33	IC$_{50}$ (μM)	no	no	fungus	*C. neoformans*	0	—	=
34	IC$_{50}$ (μM)	no	no	parasite	*L. mexicana*	0	2	>
35	IC$_{50}$ (μM)	no	no	parasite	*P. falciparum*	0	2	>
36	IC$_{50}$ (μM)	no	no	parasite	*P. falciparum* D6	0	NA	=
37	IC$_{50}$ (μM)	no	no	parasite	*P. falciparum* W2	0	NA	=
38	IC$_{50}$E-9 (M)	GSK-3	nd	enzyme	no	0	2	>
39	IC$_{90}$ (μg/mL)	no	no	parasite	*L. donovani*	0	NA	=
40	MIC (μg/mL)	no	no	bacterium	*M. tuberculosis* (H37Rv)	0	NA	=
41	pIC$_{50}$	GSK-3	beta	enzyme	no	0	0	=
42	pIC$_{50}$	GSK-3	nd	enzyme	no	0	0	=
43	IC$_{50}$ (μM)	no	no	no	Cell Efficacy	0	2	>

The problem in this type of organization of raw data is that the descriptor values are compound constants. Consequently, if these latter and LDA are based only on these values, they will necessarily fail when we change OG values. An inconvenient in this regard occurs if we pretend to use the model

for a real enzyme, since we have only one unspecific prediction and we need 43 specific probabilities, one confirming the real class and 42 giving low probabilities for the other CAC_q. We can solve this problem introducing variables characteristic of each CAC class referred on the CAC_q but without giving information in the input about the real CAC class of the protein. To this end, we used the average value of each descriptor for all enzymes that belonged to the same CAC class. We also calculated the deviation of all the descriptors from the respective group indicated in CAC_q. Altogether, we have then (63 topological descriptor values) + (63 topological descriptor values average values for CAC class) + (63 topological descriptor values deviations values for CAC class average) = 189 input variables. It is of major importance to understand that we never used as input CAC_q, so the model only includes as input the physicochemical-topological descriptor values for the protein entry and the average and deviations of these values from the CAC_q, which is not necessarily the real CAC class. The general formula for this class of LDA model is shown below (see **Equation 2**), where S(E) is not the probability but a real valued score that predicts the propensity of a compound to act as an inhibitor of a given class:

$$
\begin{aligned}
S(E) &= \sum_{k,G,Dt} b_k \cdot {}^k D_t(G) + \sum_{k,G,D} c_k \cdot {}^k \overline{D}_t(G) + \sum_{k,G,D} d_k \cdot \left({}^k D_t(G) - {}^k \overline{D}_t(G) \right) + a_0 \\
&= \sum_{k,G,Dt} b_k \cdot {}^k D_t(G) + \sum_{k,G,D} c_k \cdot {}^k \overline{D}_t(G) + \sum_{k,G,D} d_k \cdot \Delta^k D_t(G) + a_0
\end{aligned}
\tag{2}
$$

In a compact notation we write ${}^k D_t(G)$, where D_t is the type of descriptor; G is the types of subgraphs studied in the molecular connectivity G = path (*p*), clusters (*C*), path-clusters (*pC*) and chains (rings) (*Ch*).

LDA forward stepwise analysis was carried out for variable selection to build up the models [68]. All the variables included in the model were standardized in order to bring them onto the same scale. Subsequently, a standardized linear discriminant equation that allows comparison of their coefficients was obtained [69]. The square of Canonical regression coefficient (Rc) and Wilk's statistics (U) were examined in order to assess the discriminatory power of the model (U = 0 perfect discrimination, being 0 < U < 1); the separation of the two groups was statistically verified by the Fisher ratio (F) test with an error level p < 0.05.

3.3. Data Set

The data set was conformed to a set of marketed and/or reported drugs/receptor pairs where affinity/non-affinity of drugs with the receptors was established taking into consideration the IC50, *k*i, *pk*i, values. In consequence, we managed to collect 1,192 examples of active compounds in different CAC_q. In addition, we used a negative control series of 2,178 cases of non-active compounds in different CAC_q. In the two data sets used, there were the following training series: 474 active compounds plus 896 non-active compounds (1,370 in total), predicting series: 296 + 1,164 = 1,460 in total. Due to space constraints the names or codes for all compounds are listed in supplementary material **SM1** in the Supporting Information, as well as the references consulted to compile the data in this table. This series is composed at random by the most representative families of GSK-3 inhibitors taken from the literature (supplementary material **SM2**). The remaining compounds were a

heterogeneous series of inactive compounds, including members of the aforementioned families and compounds included in the Merck index [70].

4. Conclusions

In this work we have shown that the ModesLab methodology using topological indices can be considered a good alternative for developing GSK-3 inhibitors in a fast and efficient way with respect to other methods of the literature. This approach is able to correctly classify the GSK-3 inhibitory activity of compounds with different structural patterns.

Acknowledgements

References

1. Olson, R.E. Secretase inhibitors as therapeutics for Alzheimer's disease. *Annu. Rep. Med. Chem.* **2000**, *35*, 31-40.
2. Brion, J.P.; Anderton, B.H.; Authelet, M.; Dayanandan, R.; Leroy, K.; Lovestone, S.; Ocatve, J.N.; Pradier, L.; Touchet, N.; Tremp, G. Neurofibrillary Tangles and Tau Phosphorilation. *Biochem. Soc. Symp.* **2001**, *67*, 81-88.
3. Braak, H.; Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **1991**, *82*, 239-259
4. Iqbal, K.; Alonso, A.; Chen, S.; Chohan, M.O.; El-Akkad, E.; Gong, C.; Khatoon, S.; Li, B.; Liu, F.; Rahman, A.; TAnimukai, H.; Grundke-Iqbal, I. Tau pathology in Alzheimer disease and other tauopathies. *BBA-Mol. Basis Dis.* **2005**, *1739*, 198-210.
5. Cohen, P.; Frame, S. The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 769-776.
6. Woodgett, J.R.; Cohen, P. Multisite phosphorilation of glycogen synthase. Molecular basis for the sustrate specificity of glycogen synthase kinase-3 and casein kinase-II (glycogen synthase kinase-5). *Biochim. Biophys. Acta* **1984**, *788*, 339-347.
7. Nikoulina, S.E.; Ciaraldi, T.P.; Mudailar, S.; Mohideen, P.; Carter, L.; Henry, R.R. Potential role of glycogen synthase kinase-3 in skeletal muscle insulin resistance of type 2 diabetes. *Diabetes* **2000**, *49*, 263-271.
8. Lovestone, S.; Reynolds, C.H.; Latimer, D.; Davis, D.R.; Anderton, B.H.; Gallo, J.M.; Hanger, D.; Mulot, S.; Marquardt, B. Alzheimer's disease-like phosphorylation of the microtubule-associated protein tau by glycogen synthase kinase-3 in transfected mammalian cells. *Curr. Biol.* **1994**, *4*, 1077-1086.
9. Imahori, K.; Uchida, T. Physiology and pathology of tau protein kinases in relation to Alzheimer's disease. *J. Biochem.* **1997**, *121*, 179-188.

10. Takashima, A.; Murayama, M.; Yasutake, K.; Takahashi, H.; Yokoyama, M.; Ishiguro, K. Involvement of cyclin dependent kinase5 activator p25 on tau phosphorylation in mouse brain. *Neurosci. Lett.* **2001**, *306*, 37-40.

11. Ryves, W.J.; Harwood, A.J. Lithium Inhibits Glycogen Synthase Kinase-3 by Competition for Magnesium. *Biochem. Biophys. Res. Commun.* **2001**, *280*, 720-725.

12. Ishiguro, K.; Ihara, Y.; Uchida, T.; Imahori, K.A. Novel Tubulin-Dependent Protein Kinase Forming a Paired Helical Filament Epitope on Tau. *J. Bio. Chem.* **1988**, *104*, 319-321.

13. Arana, B.; Rizzo, N.; Diaz, A. Chemotherapy of cutaneous leishmaniasis: A review. *Med. Microbiol. Immunol.* **2001**, *190*, 93-95.

14. Bryceson, A. Current issues in the treatment of visceral leishmaniasis. *Med. Microbiol. Immunol.* **2001**, *190*, 85-87.

15. Sundar, S. Treatment of visceral leishmaniasis. *Med. Microbiol. Immunol.* **2001**, *190*, 89-92.

16. Fairlamb, A.H. Chemotherapy on human African trypanosomiasis: Current and future prospects. *Trends Parasitol.* **2003**, *19*, 488-494.

17. Copeland, R.A.; Pompliano, D.L.; Meek, T.D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **2006**, *5*, 730-732.

18. Liao, J.J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 409-424.

19. Pink, R.; Hudson, A.; Mouries, M.A.; Bending, M. Opportunities and chanllenges in antiparasitic drug discovery. *Nat. Rev. Drug Discov.* **2005**, *4*, 727-740.

20. Plyte, S.E.; Hughes, K.; Nilkolakaki, E.; Pulverer, B.J.; Woodgett, J.R. Glycogen Synthase Kinase-3: Functions in oncogenesis and development. *Biochim. Biophys. Acta* **1992**, *1114*, 147-162.

21. Dajani, R.; Fraser, E.; Roe, S.M.; Young, N.; Good, V.; Dale, T.C.; Pearl, L.H. Crystal structure of glycogen synthase kinase-3 beta: Structural basic for phosphate-primed subtrate specificity and autoinhibition. *Cell (Cambridge, Mass.)* **2001**, *105*, 721-732.

22. Ojo, K.K.; Gillespie, R.G.; Riechers, A.; Napuli, A.J.; Verlinde, C.L.; Buckner, F.S.; Gelb, M.H.; Domostoj, M.M.; Wells, S.J.; Scheer, A.; Wells, T.N.C.; Voorhis, C.V. Glycogen Synthase Kinase 3 is a potential drug target for african trypanosomiasis therapy. *Antimicrob. Agents Chemother.* **2008**, *52*, 3710-3717.

23. González-Díaz, H.; Munteanu, C.R. *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*; Transworld Research Network: Kerala, India, 2010.

24. Gonzalez-Díaz, 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.

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

26. González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices. *Curr. Top Med. Chem.* **2007**, *7,* 1015-1029.

27. 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.

28. 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.

29. 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.

30. 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.

31. Khan, M. T. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr. Drug Metab.* **2010**, *11*, 285-295.

32. Garcia, I.; Diop, Y.F.; Gomez, G. QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr. Drug Metab.* **2010**, *11*, 307-314.

33. 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.; *et al.* Artificial intelligence techniques for colorectal cancer drug metabolism: Ontology and complex network. *Curr. Drug Metab.* **2010**, *11*, 347-368.

34. Mao, B.Y.; Chou, K.C.; Maggiora, G.M. Topological analysis of hydrogen bonding in protein structure. *Eur. J. Biochem.* **1990**, *188*, 361-365.

35. Estrada, E.; Gutiérrez, Y. *ModesLab*, versión 1.5, 2002-2004.

36. 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.

37. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley-VCH: Weinheim, Germany, 2000.

38. Freund, J.A.; Poschel, T. Stochastic processes in physics, chemistry, and biology. In *Lect. Notes Phys.*; Springer-Verlag: Berlin, Germany, 2000.

39. Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573-1588.

40. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2001**, *43*, 1975-1985.

41. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Chemometric Methods in Molecular Design*; Van Waterbeemd, H., Ed.; Wiley-VCH: New York, NY, USA, 1995; Volume 2, pp. 265-282.

42. 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. *Biochim. Biophys. Acta.* **2009**, *1794*, 1784-1794.

43. Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Bolas-Fernandez, F.; Prado-Prado, F.J.; Podda, G.; Uriarte, E.; Ubeira, F.M.; Gonzalez-Diaz, H. Prediction of Enzyme Classes from 3D Structure: A General Model and Examples of Experimental-Theoretic Scoring of Peptide Mass Fingerprints of Leishmania Proteins. *J. Proteome Res.* **2009**, *8*, 4372-4382.

44. Konda, V.R.; Desai, A.; Darland, G.; Bland, J.S.; Tripp, M.L. Rho iso-alpha acids from hops inhibit the GSK-3/NF-kappaB pathway and reduce inflammatory markers associated with bone and cartilage degradation. *J. Inflamm.* **2009**, *6*, 26-34.

45. Rochais, C.; Duc, N.V.; Lescot, E.; Sopkova-de Oliveira Santos, J.; Bureau, R.; Meijer, L.; Dallemagne, P.; Rault, S. Synthesis of new dipyrrolo- and furopyrrolopyrazinones related to tripentones and their biological evaluation as potential kinases (CDKs1-5, GSK-3) inhibitors. *Eur. J. Med. Chem.* **2009**, *44*, 708-716.

46. Simon, D.; Benitez, M.J.; Gimenez-Cassina, A.; Garrido, J.J.; Bhat, R.V.; Diaz-Nido, J.; Wandosell, F. Pharmacological inhibition of GSK-3 is not strictly correlated with a decrease in tyrosine phosphorylation of residues 216/279. *J. Neurosci. Res.* **2008**, *86*, 668-674.

47. Patel, D.S.; Bharatam, P.V. Selectivity criterion for pyrazolo[3,4-b]pyrid[az]ine derivatives as GSK-3 inhibitors: CoMFA and molecular docking studies. *Eur. J. Med. Chem.* **2008**, *43*, 949-957.

48. Jacquemard, U.; Dias, N.; Lansiaux, A.; Bailly, C.; Loge, C.; Robert, J.M.; Lozach, O.; Meijer, L.; Merour, J.Y.; Routier, S. Synthesis of 3,5-bis(2-indolyl)pyridine and 3-[(2-indolyl)-5-phenyl]pyridine derivatives as CDK inhibitors and cytotoxic agents. *Bioorg. Med. Chem.* **2008**, *16*, 4932-4953.

49. Xiao, J.; Guo, Z.; Guo, Y.; Chu, F.; Sun, P. Inhibitory mode of N-phenyl-4-pyrazolo[1,5-b] pyridazin-3-ylpyrimidin-2-amine series derivatives against GSK-3: Molecular docking and 3D-QSAR analyses. *Protein Eng. Des. Sel.* **2006**, *19*, 47-54.

50. Tavares, F.X.; Boucheron, J.A.; Dickerson, S.H.; Griffin, R.J.; Preugschat, F.; Thomson, S.A.; Wang, T.Y.; Zhou, H.Q. N-Phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amines as potent and selective inhibitors of glycogen synthase kinase 3 with good cellular efficacy. *J. Med. Chem.* **2004**, *47*, 4716-4730.

51. Olesen, P.H.; Sorensen, A.R.; Urso, B.; Kurtzhals, P.; Bowler, A.N.; Ehrbar, U.; Hansen, B.F. Synthesis and *in vitro* characterization of 1-(4-aminofurazan-3-yl)-5-dialkylaminomethyl-1H-[1,2,3]triazole-4-carboxyl ic acid derivatives. A new class of selective GSK-3 inhibitors. *J. Med. Chem.* **2003**, *46*, 3333-3341.

52. Calabuig, C.; Anton-Fos, G.M.; Galvez, J.; Garcia-Domenech, R. New hypoglycaemic agents selected by molecular topology. *Int. J. Pharm.* **2004**, *278*, 111-118.

53. Garcia-Garcia, A.; Galvez, J.; de Julian-Ortiz, J.V.; Garcia-Domenech, R.; Munoz, C.; Guna, R.; Borras, R. New agents active against Mycobacterium avium complex selected by molecular topology: A virtual screening method. *J. Antimicrob. Chemother.* **2004**, *53*, 65-73.

54. Prado-Prado, F.J.; Ubeira, F.M.; Borges, F.; Gonzalez-Diaz, H. Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J. Comput. Chem.* **2009**, *31*, 164-173.

55. Prado-Prado, F.J.; Martinez de la Vega, O.; Uriarte, E.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. 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.* **2009**, *17*, 569-575.

56. Prado-Prado, F.J.; de la Vega, O.M.; Uriarte, E.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. 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.* **2009**, *17*, 569–575.

57. 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.

58. Prado-Prado, F.J.; Gonzalez-Diaz, H.; de la Vega, O.M.; Ubeira, F.M.; Chou, K.C. 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.* **2008**, *16*, 5871-5880.

59. 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. *Bioorg. Med. Chem.* **2007**, *15*, 897-902.

60. 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. *Chem. Med. Chem.* **2007**, *2*, 449-478.

61. Marrero-Ponce, Y.; Meneses-Marcel, A.; Castillo-Garit, J.A.; Machado-Tugores, Y.; Escario, J.A.; Barrio, A.G.; Pereira, D.M.; Nogal-Ruiz, J.J.; Aran, V.J.; Martinez-Fernandez, A.R.; Torrens, F.; Rotondo, R.; Ibarra-Velarde, F.; Alvarado, Y.J. Predicting antitrichomonal activity: A computational screening using atom-based bilinear indices and experimental proofs. *Bioorg. Med. Chem.* **2006**, *14*, 6502-6524.

62. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Pereira, D.M.; Escario, J.A.; Nogal-Ruiz, J.J.; Ochoa, C.; Aran, V.J.; Martinez-Fernandez, A.R.; *et al.* A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: Outcomes of in silico studies supported by experimental results. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3838-3843.

63. Marrero-Ponce, Y.; Diaz, H.G.; Zaldivar, V.R.; Torrens, F.; Castro, E.A. 3D-chiral quadratic indices of the 'molecular pseudograph's atom adjacency matrix' and their application to central chirality codification: Classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Bioorg. Med. Chem.* **2004**, *12*, 5331-5342.

64. 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.

65. 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.

66. Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1194-1203.

67. Cronin, M.T.; Aptula, A.O.; Dearden, J.C.; Duffy, J.C.; Netzeva, T.I.; Patel, H.; Rowe, P.H.; Schultz, T.W.; Worth, A.P.; Voutzoulidis, K.; Schuurmann, G. Structure-based classification of antibacterial activity. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 869-878.

68. 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. *J. Med. Chem.* **2006**, *49*, 1149-1156.

69. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. Standardized Multiple Regression Model. In *Applied Linear Statistical Models*, 5th ed.; McGraw Hill: New York, NY, USA, 2005; pp. 271-277.

70. Budavari, S. *The Merck Index*, 12th ed.; Merck & Co, Inc: Whitehouse Station, NJ, USA, 1996.

Sample Availability: Not available.