Prediction of Neurological Enzyme Targets for Known and New Compounds with a Model using Galvez's Topological Indices

Francisco J. Prado-Prado[1*], Humberto González-Díaz [2,3] , Eduardo Sobarzo-Sanchez [4], Isela García-Pintos [5]

[1] *Biomedical Sciences Department, Health Sciences Division,*
University of Quintana Roo UQROO, 77039, Mexico
[2] *Department of Organic Chemistry II, Faculty of Science and Technology,*
University of the Basque Country UPV/EHU, 48940, Leioa, Spain
[3] *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*
[4] *Department of Organic Chemistry, Faculty of Pharmacy, USC, 15782, Spain*
[5] *Department of Organic Chemistry, University of Vigo, 36200, Spain*

Abstract. Alzheimer's Disease (AD), Parkinson, and other neurodegenerative diseases are a major health problem nowadays. However, in many cases, current therapies are merely palliative and only temporarily slow cognitive decline. In this sense, the discovery of new drugs for the treatment of neurodegenerative diseases is a goal of the major importance. Public databases, like ChEMBL, contain a large amount of data about multiplexing assays of inhibitors of a group of enzymes with special relevance in central nervous system. Mono Amino Oxidases (MAOs), Acetyl Cholinesterase (AChE), Glycogen Synthase Kinase-3 (GSK-3), AChE (AChE), and 5α-reductases (5αRs). This data conform an important information source for the application of multi-target computational models. However, almost all the computational models known focus in only one target. In this work, we developed linear multi-target QSAR models (mt-QSAR) for inhibitors of 8 different enzymes promising in the treatment of different neurodegenerative diseases. In so doing, we combined by the first time the software DRAGON with Moving Average parameters with this objective. The best DRAGON model found predict with very high accuracy, specificity, and sensitivity >90% a very large data set >10000 cases in training and validation series.

Keywords: Neurodegenerative diseases, multi-target enzyme inhibitors; QSAR, Box and Jenkins moving averages; Galvez's charge transfer indices; Topological indices; chemical graph theory.

*corresponding author:
PRADO-PRAO, F. (fenol1@hotmail.com), *Biomedical Sciences Department, Health Science Division,*
University of Quintana Roo, Chetumal, 77039, México

1. INTRODUCTION

The discovery of new compounds for the treatment of neurodegenerative diseases is a goal of the major importance for medicinal chemistry and biopharmaceutical industry. In fact, neurodegenerative diseases have a high negative impact in personal and public health. For instance, Alzheimer´s disease [1] 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 [2, 3]. 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 [4]. A group of neuronal enzymes stands out between the potential targets of drugs useful in the treatment of these diseases. In fact, the functions of neuronal enzymes and its implication in various human diseases have triggered an active search for potent and selective neuronal enzyme inhibitors [5] in the last years. Mono Amino Oxidases (MAOs), Acetyl Cholinesterase (AChE), Glycogen Synthase Kinase-3 (GSK-3), Acetyl Cholinesterase (AChE), and 5α-reductases (5αRs); stand out among the more promising enzymes in this sense. GSK-3 has two isoforms, GSK-3α and GSK-3β [6]. In particular, GSK-3β is well known to play critical roles in oxidative stress-induced neurodegenerative diseases such as AD [7, 8]. A more comprehensive understanding of the mechanistic basis for GSK-3 isoform-specific functions could lead to the development of isoform-specific inhibitors [9]. MAOs are important

flavoenzymes with two isoforms (MAO-A and MAO-B); which are responsible for the oxidative deamination of neurotransmitters and dietary amines and thus involved in neurodegenerative diseases. MAO-A has a higher affinity for serotonin and norepinephrine whereas MAO-B preferentially deaminates phenylethylamine and benzylamine. Therefore, selective MAO-A inhibitors (*i.e.* chlorgyline) are used in the treatment of neurological disorders such as depression, whereas the MAO-B inhibitors (*i.e.* selegiline) are useful in the treatment of AD and Parkinson's. All of these aspects have led to an intensive search for novel MAO inhibitors (MAOIs) and this effort has increased considerably in recent years. On the other hand, AChE activity is associated with neuritic plaques (NPs) and neurofibrillary tangles (NFTs) in AD brains. AChE inhibitors (AChIs) are promising drugs in clinical practice for the treatment of AD. Recently, Bond et al reviewed the effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of AD [10]. Last, 5αRs are a family of isozymes expressed in a seveal tissues including the central nervous system. Very recently, Traish [11] performed a comprehensive literature search from 1970-2011 via PubMed and summarized the relevant information with special emphasis on the central nervous systems.

Thus, the rational design of new enzyme inhibitors for the treatment of neurodegenerative disease constitutes a major goal. In our opinion, Quantitative Structure-Activity Relationships (QSAR) techniques may be of help in this sense. Regrettably, almost current QSAR techniques are able to predict new outcomes only for one specific assay. In our opinion, we can evade this problem developing new Multi-target/Multiplexing QSAR models (mt-QSAR/mx-QSAR). These methods are especially powerful when we need to process very large collections of compounds assayed against multiple molecular or cellular targets in different assay conditions (m_j) as is the case of CHEMBL [12, 13]. This step may be of the major relevance for the future of QSAR. Notably, mt-QSAR models are able to predict the results of the assay of different drugs for multiple targets. However, mt-QSAR models are unable to predict different results for a given series of targets when we change the set of specific assay conditions for each target. Fortunately, the new class of mx-QSAR models applies not only to different targets but also to different multiplexing assay conditions (m_j) for all targets. Specifically, we have reported the first mx-QSAR model for multiplexing assays of anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors *in vitro*, *in vivo*, and in different cellular lines [14].

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 [15, 16]. We can use different topological indices (TIs) of molecular graphs (G) to speed up the process of codification of the molecular structure of drugs in QSAR studies. In particular, Galvez's charge transfer indices and their different variants (G-like indices) have been demonstrated to be very useful in QSAR/QSPR studies of small molecules. For instance, the HIV-1 RT inhibitory activity of thiazolidinones has been analyzed with different TIs and G-like values together with other TIs have been found to be accurate predicting the activity of these compounds [17]. G-like indices also overtook 11 kinds of molecular descriptors in one study of inhibitors of flavonoids aldose reductase enzyme [18]. For a detailed revision of many QSAR models using TIs (including G_k indices) see the recent in-depth review published in Chemical Reviews by Galvez's group [19]. In addition, G-like indices have been found to be also useful for study complex bio-molecular networks and not only chemical structure. For instance, González-Díaz *et al.* [20] used G-like indices to study different classes of networks found in drug research, nature, technology, and social-legal sciences. Important QSAR studies of the enzymes under study in this work have been published before. For instance, Santana and Uriarte *et al.* [21-23] have published several QSAR and experimental studies of new MAO inhibitors. However, there are not mt-QSAR models based on G-like indices for all these enzymes together. In this work, we used by the first time molecular descriptors calculated developed by Galvez *et al.* to develop one mt-QSAR model for inhibitors of 8 different enzymes. The best model found use as input we used moving average operators to predict with very high accuracy, specificity, and sensitivity >90% a very large data set >10,000 cases in training and validation series.

2. MATERIALS AND METHODS
2.1 Computational methods

Regrettably, almost current Quantitative Structure-Activity Relationships (QSAR) techniques are able to predict new outcomes only for one specific assay. In our opinion, we can evade this problem developing new Multi-target/Multiplexing QSAR models (mt-QSAR/mx-QSAR). These methods are especially powerful when we need to process very large collections of compounds assayed against multiple molecular or cellular targets in the different j^{th} assay conditions (c_j); as is the case of CHEMBL [12, 13]. This step may be of the major relevance for the future of QSAR. Notably, mt-QSAR models are able to predict the results of the assay of different drugs for multiple targets. However, mt-QSAR models are unable to predict different results for a given series of targets when we change the set of specific assay conditions for each target. Fortunately, the new class of mx-QSAR models applies not only to different targets but also to different multiplexing assay conditions (c_j) for all targets. Specifically, we have reported the first mx-QSAR model for multiplexing assays of anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors *in vitro*, *in vivo*, and in different cellular lines [14]. In a first step, we need to calculate the molecular descriptors using D_i of a given i^{th} compound using one or more software for generation of molecular descriptors. In a second step, we expand the raw dataset of molecular descriptors adding new variables $\Delta D_{ij} = D_i - <D_{ij}>$. Next, we upload this preprocessed data to one Statistics or Machine Learning software to seek the model. The linear mt-QSAR model based on moving averages and LDA has the following general form:

$$S_i\left(m_j\right) = a_0 + \sum_i^q b_i \cdot D_i + \sum_{q+1}^r c_i \cdot \left\langle D_{ij}\right\rangle + \sum_{r+1}^s d_i \cdot \Delta D_{ij} \qquad (1)$$

$$a_0 + \sum_i^q b_i \cdot D_i + \sum_{q+1}^r c_i \cdot \left\langle D_{ij}\right\rangle + \sum_{r+1}^s d_i \cdot \left(D_i - \left\langle D_{ij}\right\rangle\right)$$

Where, S_{ij} is a numerical score of the biological activity of the ith compound measured under the jth assay defined by the set of conditions c_j. In these models, the average $<D_{ij}> = <D_i(c_j)>$, used to calculate ΔD_{ij} values, is the average of the D_i for different compounds and do not runs over a time domain but over a set of molecular descriptors that obey a given boundary condition c_j. These deviation-like parameters ΔD_{ij} are inspired in the idea of moving averages used in time series analysis [24]. The idea of using moving average operators comes from the seminar works on time series analysis published by Box and Jenkins [25]. More recently, González-Díaz *et al.* [26, 27] have used moving average operators to construct mt-QSAR models. See also the excellent works published by Speck-Planche and Cordeiro *et al.* [28-33].

3. RESULTS AND DISCUSSION
3.1. Multi-target DRAGON model of drug-neuroenzyme interaction

The outcome of multiplexing neural enzyme inhibition assays depend both on drug structure and the set of assay conditions selected (m_j) [34]. In this work, we report the first mx-QSAR model capable of predict whether a drug with a determined molecular structure may give or not a positive result in different multiplexing assay conditions m_j. These models are expected to give different classification probabilities of the compound for different: organisms (o_t), biological assays (a_u), molecular or cellular targets (t_e). The best mt-QSAR model found using DRAGON was the following:

$$S_i\left(m_j\right) = -1.17 \cdot \left\langle D_1\left(e\right)\right\rangle - 7.26 \cdot \left\langle D_2\left(a\right)\right\rangle - 0.67 \cdot \Delta D_3\left(a\right) \qquad (2)$$
$$+ 0.07 \cdot \Delta D_4\left(a\right) - 0.29 \cdot \Delta D_5\left(a\right) - 2.63 \cdot \Delta D_6\left(a\right)$$
$$+ 2.42 \cdot \Delta D_7\left(a\right) + 0.76 \cdot \Delta D_8\left(o\right) + 5.87 \cdot \Delta D_9\left(o\right) + 6.9856$$
$$N = 3683 \quad R_c = 0.80 \quad \chi^2 = 18712.55 \quad p-level < 0.001$$

$S(m_j) = S(d_i, a_u, o_t, t_e)$ is a real-valued variable that scores the propensity of the drug to be active in multiplex pharmacological assays of the drug d_i carried out on the conditions selected $m_{j =>}$ assay = a_u, organism = o_t, and target enzyme t_e. The statistical parameters for the above equation in training are: Number of cases used to train the model (N), Canonical Regression Coefficient (Rc), Sensitivity (Sn), Specificity (Sp), and Accuracy (Ac) [24]. The probability cut-off for this LDA model is $^i p_1(m_j) > 0.5 => C_i(m_j) = 1$. It means that the ith drug (d_i) predicted by the model with probability > 0.5 are expected to inhibit the enzyme present in the jth assays carry

3

out under the given set of conditions m_j. In **Table 1**, we explain in detail the different terms of this equation. Online supplementary material files contain detailed lists of results for all cases analyzed.

Table 1 comes about here

This linear equation presented good results both in training and external validation series with overall Accuracy in training series above 90% (see **Table 2**). According to previous reports [35-43] values accuracy higher than 75% are acceptable for LDA-QSAR models. The reader should be aware that N here is not number of compounds but number of statistical cases. One compound may lead to 1 or more statistical cases because it may give different outcomes for alternative biological assays carried out in different organisms and used different enzymes as targets [44].

Table 2 comes about here

3.1.1 Prediction of interaction with other neuro-enzymes

An additional use of mt-DRAGON model was to carry out the "*in silico*" or virtual screening of the new interactions with respect to all other enzymes used in this model. It may help to found new interactions for these drugs or discard possible toxicological effects depending on the other interactions predicted and/or discarded for these compounds. This type of experiment is of the major importance due to the cost in terms of animal sacrifice, time, materials and human resources of the experimental assay of all compounds against all these targets, see recent reviews by Duardo-Sanchez *et al.* [45-48].

Using this model we can predict the different relationships between the drug-protein interactions [49, 50]. We can reach this goal because the model predicts all neuro enzymes as non-active or moderate activity with respect to all compounds that are used in the model. Another important goal, is that the model predicted all the before results on *Homo sapiens* organism. In **Table 3,** we depict all the labels of the experimental parameters, target enzymes, and organism.

Table 3 comes about here

CONCLUSION

The functions of neural enzymes and its implication in various human diseases have triggered an active search for potent and selective neural enzyme inhibitors. Theoretical mt-QSAR models based on LDA and DRAGON descriptors may become a useful tool in this sense. 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 models with large data bases, and this helps us to make the results more reliable.

REFERENCES

[1]   Olson RE. Secretase inhibitors as therapeutics for Alzheimer´s disease. *Annu Rep Med Chem*, **2000**; 35: 31-40.

[2]   Deguchi K, Tsuru K, Hayashi T, Takaishi M, Nagahara M, Nagotani S. Insights into oxidative stress and potential novel therapeutic targets for Alzheimer disease. *J Cereb Blood Flow Metab*, **2006**; 26: 1263-1273.

[3]   Stabenfeldt SE, García AJ, LaPlaca MC. Thermoreversible laminin-functionalized hydrogel for neural tissue engineering. *J Biomed Mater Res A*, **2006**; 77: 718-725.

[4]   Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **2002**; 297: 353-356.

[5]   Huang J, Chen Y-j, Bian W-h, Yu J, Zhao Y-w, Liu X-y. Unilateral amyloid-β25-35 injection into the rat amygdala increases the expressions of aberrant tau phosphorylation kinases *Chinese Medical Journal*, **2010**; 123: 1311-1314.

[6]   Woodgett JR. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *Embo J*, **1990**; 9: 2431-2438.

[7]   Caia F, Wanga F, Lina F-K, Liua C, Maa L-Q, Liua J, Wua W-N, Wanga W, Wanga J-H, Chen J-G. Redox modulation of long-term potentiation in the hippocampus via regulation of the glycogen synthase kinase-3β pathway. *Free Radical Biology and Medicine*, **2008**; 45: 964-970.

[8]   Maiese K, Chong ZZ. Insights into oxidative stress and potential novel therapeutic targets for Alzheimer disease. *Restorative Neurology and Neuroscience*, **2004**; 22: 87-104.

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

[10]   Bond M, Rogers G, Peters J, Anderson R, Hoyle M, Miners A, Moxham T, Davis S, Thokala P, Wailoo A, Jeffreys M, Hyde C. The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of Technology Appraisal No. 111): a systematic review and economic model. *Health Technol Assess*, **2012**; 16: 1-470.

[11]   Traish AM. 5alpha-Reductases in Human Physiology: An Unfolding Story. In: ed.^eds., Endocr Pract, 2012; pp. 1-38.

[12]   Riera-Fernandez I, Martin-Romalde R, Prado-Prado FJ, Escobar M, Munteanu CR, Concu R, Duardo-Sanchez A, Gonzalez-Diaz H. From QSAR models of Drugs to Complex Networks: State-of-Art Review and Introduction of New Markov-Spectral Moments Indices. *Current Topics in Medicinal Chemistry*, **2012**.

[13]   Prado-Prado F, Garcia-Mera X, Escobar M, Sobarzo-Sanchez E, Yanez M, Riera-Fernandez P, Gonzalez-Diaz H. 2D MI-DRAGON: a new predictor for protein-ligands interactions and theoretic-experimental studies of US FDA drug-target network, oxoisoaporphine inhibitors for MAO-A and human parasite proteins. *European Journal of Medicinal Chemistry*, **2011**; 46: 5838-51.

[14]   Garcia I, Fall Y, Gomez G, Gonzalez-Diaz H. First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different cellular lines. *Molecular Diversity*, **2011**; 15: 561-567.

[15]   Nunez MB, Maguna FP, Okulik NB, Castro EA. QSAR modeling of the MAO inhibitory activity of xanthones derivatives. *Bioorg Med Chem Lett*, **2004**; 14: 5611-5617.

[16]   Todeschini R, Consonni V. Handbook of Molecular Descriptors. *Wiley VCH*, **2000**.

[17]   Prabhakar YS, Rawal RK, Gupta MK, Solomon VR, Katti SB. Topological descriptors in modeling the HIV inhibitory activity of 2-aryl-3-pyridyl-thiazolidin-4-ones. *Comb Chem High Throughput Screen*, **2005**; 8: 431-7.

[18]   Fernandez M, Caballero J, Helguera AM, Castro EA, Gonzalez MP. Quantitative structure-activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds. *Bioorg Med Chem*, **2005**; 13: 3269-77.

[19]   Garcia-Domenech R, Galvez J, de Julian-Ortiz JV, Pogliani L. Some new trends in chemical graph theory. *Chem Rev*, **2008**; 108: 1127-69.

[20]   Riera-Fernandez P, Munteanu CR, Dorado J, Martin-Romalde R, Duardo-Sanchez A, Gonzalez-Diaz H. From chemical graphs in computer-aided drug design to general Markov-Galvez indices of drug-target, proteome, drug-parasitic disease, technological, and social-legal networks. *Curr Comput Aided Drug Des*, **2011**; 7: 315-37.

[21]   Molina E, Sobarzo-Sanchez E, Speck-Planche A, Matos MJ, Uriarte E, Santana L, Yanez M, Orallo F. Monoamino oxidase a: an interesting pharmacological target for the development of multi-target QSAR. *Mini Rev Med Chem*, **2012**; 12: 947-58.

[22]   Santana L, Gonzalez-Diaz H, Quezada E, Uriarte E, Yanez M, Vina D, Orallo F. Quantitative Structure-Activity Relationship and Complex Network Approach to Monoamine Oxidase A and B Inhibitors. *Journal of Medicinal Chemistry*, **2008**; 51: 6740-6751.

[23]   Santana L, Uriarte E, Gonzalez-Diaz 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: 1149-56.

[24]   Hill T, Lewicki P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining. StatSoft: Tulsa 2006

[25]   Box GEP, Jenkins GM. Time series analysis. Holden-Day 1970.

[26]   Luan F, Cordeiro MN, Alonso N, Garcia-Mera X, Caamano O, Romero-Duran FJ, Yanez M, Gonzalez-Diaz H. TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. *Bioorg Med Chem*, **2013**; 21: 1870-9.

[27]   Tenorio-Borroto E, Garcia-Mera X, Penuelas-Rivas CG, Vasquez-Chagoyan JC, Prado-Prado FJ, Castanedo N, Gonzalez-Diaz H. Entropy Model For Multiplex Drug-Target Interaction Endpoints Of Drug Immunotoxicity. *Curr Top Med Chem*, **2013**.

[28]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. A ligand-based approach for the in silico discovery of multi-target inhibitors for proteins associated with HIV infection. *Mol Biosyst*, **2012**; 8: 2188-96.

[29]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Chemoinformatics in anti-cancer chemotherapy: Multi-target QSAR model for the in silico discovery of anti-breast cancer agents. In: ed.^eds., Eur J Pharm Sci. 2012 Elsevier B.V: Netherlands, 2012; pp. 273-9.

[30]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. In Silico Discovery and Virtual Screening of Multi-Target Inhibitors for Proteins in Mycobacterium tuberculosis. In: ed.^eds., Comb Chem High Throughput Screen: Netherlands, 2012; pp. 666-73.

[31]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Predicting multiple ecotoxicological profiles in agrochemical fungicides: a multi-species chemoinformatic approach. In: ed.^eds., Ecotoxicol Environ Saf. 2012 Elsevier Inc: United States, 2012; pp. 308-13.

[32]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. In: ed.^eds., Bioorg Med Chem. 2012 Elsevier Ltd: England, 2012; pp. 4848-55.

[33]   Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. *Bioorganic and Medicinal Chemistry*, **2011**; 19: 6239-44.

[34]   Gerets HH, Dhalluin S, Atienzar FA. Multiplexing cell viability assays. *Methods in Molecular Biology*, **2011**; 740: 91-101.

[35]   Patankar SJ, Jurs PC. Classification of inhibitors of protein tyrosine phosphatase 1B using molecular structure based descriptors. *J Chem Inf Comput Sci*, **2003**; 43: 885-99.

[36]   Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, 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.

[37]   Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castanedo N, Ibarra-Velarde F, Huesca-Guillen A, Sanchez AM, Torrens F, Castro EA. Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic. *Bioorg Med Chem*, **2005**; 13: 1005-20.

[38]   Marrero-Ponce Y, Machado-Tugores Y, Pereira DM, Escario JA, Barrio AG, Nogal-Ruiz JJ, Ochoa C, Aran VJ, Martinez-Fernandez AR, Sanchez RN, Montero-Torres A, Torrens F, Meneses-Marcel A. A computer-based approach to the rational discovery of new trichomonacidal drugs by atom-type linear indices. *Curr Drug Discov Technol*, **2005**; 2: 245-65.

[39]   Casanola-Martin GM, Marrero-Ponce Y, Khan MT, Ather A, Sultan S, Torrens F, Rotondo R. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorg Med Chem*, **2007**; 15: 1483-503.

[40]   Casanola-Martin GM, Marrero-Ponce Y, Tareq Hassan Khan M, Torrens F, Perez-Gimenez F, Rescigno A. Atom- and bond-based 2D TOMOCOMD-CARDD approach and ligand-based virtual screening for the drug discovery of new tyrosinase inhibitors. *J Biomol Screen*, **2008**; 13: 1014-24.

[41]   Casanola-Martin GM, Marrero-Ponce Y, Khan MT, Khan SB, Torrens F, Perez-Jimenez F, Rescigno A, Abad C. Bond-based 2D quadratic fingerprints in QSAR studies: virtual and in vitro tyrosinase inhibitory activity elucidation. *Chem Biol Drug Des*, **2010**; 76: 538-45.

[42]   Rodriguez-Soca Y, Munteanu CR, Dorado J, Pazos A, Prado-Prado FJ, Gonzalez-Diaz H. Trypano-PPI: A Web Server for Prediction of Unique Targets in Trypanosome Proteome by using Electrostatic Parameters of Protein-protein Interactions. *Journal of Proteome Research*, **2010**; 9: 1182-1190.

[43]   Gonzalez-Diaz H, Muino L, Anadon AM, Romaris F, Prado-Prado FJ, Munteanu CR, Dorado J, Pazos Sierra A, Mezo M, Gonzalez-Warleta M, Garate T, Ubeira FM. MISS-Prot: web server for self/non-self discrimination of protein residue networks in parasites; theory and experiments in Fasciola peptides and Anisakis allergens. *Molecular Biosystems*, **2011**; 7: 1938-1955.

[44]   Martinez-Romero M, Vazquez-Naya JM, Rabunal JR, Pita-Fernandez S, Macenlle R, Castro-Alvarino J, Lopez-Roses L, Ulla JL, Martinez-Calvo AV, Vazquez S, Pereira J, Porto-Pazos AB, Dorado J, Pazos A, Munteanu CR. Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. *Curr Drug Metab*, **2010**; 11: 347-68.

[45]   Duardo-Sanchez A, Patlewicz G, Lopez-Diaz A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. *Curr Top Med Chem*, **2008**; 8: 1666-75.

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

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

[48]   Gonzalez-Diaz H, Romaris F, Duardo-Sanchez A, Perez-Montoto LG, Prado-Prado F, Patlewicz G, Ubeira FM. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Curr Pharm Des*, **2010**; 16: 2737-64.

[49]   Gonzalez-Diaz H, Prado-Prado F, Sobarzo-Sanchez E, Haddad M, Maurel Chevalley S, Valentin A, Quetin-Leclercq J, Dea-Ayuela MA, Teresa Gomez-Munos M, Munteanu CR, Jose Torres-Labandeira J, Garcia-Mera X, Tapia RA, Ubeira FM. NL MIND-BEST: a web server for ligands and proteins discovery--theoretic-experimental study of proteins of Giardia lamblia and new compounds active against Plasmodium falciparum. *J Theor Biol*, **2011**; 276: 229-49.

[50]   Gonzalez-Diaz H, Prado-Prado F, Garcia-Mera X, Alonso N, Abeijon P, Caamano O, Yanez M, Munteanu CR, Pazos A, Dea-Ayuela MA, Gomez-Munoz MT, Garijo MM, Sansano J, Ubeira FM. MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from Trichomonas gallinae. *J Proteome Res*, **2011**; 10: 1698-718.

FIGURES AND TABLES PRESENTS IN THIS MANUSCRIPT

Table 1. Details of the DRAGON mt-QSAR model for neural enzyme inhibitors

Step	Coeff.	Value	Variable	Expansion[a]	Symbols[b]
0	a_0	6.98560	-	-	-
1	c_1	-1.16744	$\langle D_1(e) \rangle$	-	D_1 = GGI2
2	c_2	-7.25512	$\langle D_2(a) \rangle$	-	D_2 = JGI1
3	d_3	0.67000	$\Delta D_3(e)$	$= D_3 - \langle D_3(e) \rangle$	D_3 = GGI4
4	d_4	0.06642	$\Delta D_4(a)$	$= D_5 - \langle D_5(a) \rangle$	D_4 = GGI1
5	d_5	-0.28877	$\Delta D_5(a)$	$= D_5 - \langle D_5(a) \rangle$	D_5 = GGI3
6	d_6	-2.63365	$\Delta D_6(a)$	$= D_6 - \langle D_6(a) \rangle$	D_6 = GGI10
7	d_4	2.42299	$\Delta D_7(a)$	$= D_7 - \langle D_7(a) \rangle$	D_7 = JGT
8	d_8	0.75944	$\Delta D_8(o)$	$= D_8 - \langle D_8(o) \rangle$	D_8 = GGI7
9	d_9	5.87857	$\Delta D_9(o)$	$= D_9 - \langle D_9(o) \rangle$	D_9 = JGT

[a] $\langle D_i(m_j) \rangle = SUM(D_i)/n_1(m_j)$, see materials and methods; [b] Symbols of G-like descriptors calculated by the software DRAGON.

Table 2. Results of different DRAGON multi-target classification models

Drug-Enzyme interaction	Train			Statistical Parameter
	Active	Non-Active	%	
Active	9231	1020	90.0	Sn
Non-Active	771	8289	91.5	Sp
			90.7	Ac
Drug-Enzyme interaction	Validation			Statistical Parameter
	%	%	%	
Active	3145	324	90.7	Sn
Non-Active	245	2775	91.9	Sp
			91.2	Ac

[a] Sensitivity = Sn = Positive Correct/Positive Total; Specificity = Sp = Negative Correct / Negative Total; Accuracy = Ac = Total Correct / Overall Total

Table 3. Labels of the experimental parameters, target enzymes, and organisms

Label	Target	Assay	Standard (Unit)
1	AChE	12	cKi (nM)
2	Gamma-secretase subunit APH-1A	13	EC_{50} (nM)
3	Gamma-secretase subunit APH-1B	14	ED_{50} (nM)
4	Gamma-secretase subunit PEN-2	16	IC50 (microM)
5	Glycogen synthase kinase-3 alpha	17	IC50 (nM)
6	Glycogen synthase kinase-3 beta	20	Inhibition (%)
7	Monoamine oxidase A	26	Inhibition ((nM of H202) min-1 (ug of protein)-1)
8	Monoamine oxidase B	27	Inhibition (ug)
Label	Organism	30	Ki (nM)
1	Bos taurus	34	Ki
2	Electrophorus electricus	37	Log IC50
3	Homo sapiens	39	Log Ki (M)
4	Mus musculus	40	Log Ki
5	Rattus norvegicus	42	Relative activity
6	Torpedo californica	43	Selectivity (nM)
Assay	Standard (Unit)	44	Selectivity
1	Activity (%)	48	Selectivity coefficient
4	Activity (nM)	49	Selectivity index
5	Activity (nM min-1 mg-1)	50	Selectivity ratio
7	Activity (ug)	53	SI
9	Activity	55	Specific activity (%)
10	A-selectivity	56	Specific activity (nM Pi min-1 (mg of protein)-1)