



Conference Proceedings Paper – Entropy

Multi Activity QSAR Models for Anti- Parasite Drugs Using Markov Entropy Indices

Humberto González-Díaz ¹, María Fernanda Frías-Fraire ², Alejandra Aguirre-Crespo ²,
Francisco Javier Aguirre-Crespo ² and Francisco Javier Prado Prado ² *

¹ Organic Chemistry II, University of the Basque Country UPV/EHU, Spain.

² Biomedical Sciences Department, Health Sciences Division, University of Quintana Roo, UQROO, 77039, Mexico.

* Author to whom correspondence should be addressed; E-Mail: fenol1@hotmail.com.

Published: 13 November 2015

Abstract: There are many parasite species with very different antiparasite drugs susceptibility. Computational methods in biology and chemistry prediction of the biological activity based on Quantitative Structure-Activity Relationships (QSAR) substantially increases the potentialities of this kind of networks avoiding time and resources consuming experiments. Unfortunately, almost QSAR models are unspecific or predict activity against only one species. To solve this problem we developed here a multi-species QSAR classification model (ms-QSAR). In so doing, we use Markov Chains theory to calculate new multi-target spectral moments to fit a QSAR model that predict by the first time a ms-QSAR model for 500 drugs tested in the literature against 16 parasite species and other 207 drugs no tested in the literature using entropy type indices. The data was processed by Artificial Neural Network (ANN) classifying drugs as active or non-active against the different tested parasite species. The best ANN found was MLP 23:23-18-1:1. Overall model classification accuracy was 85.65% (211/244 cases) in training. Validation of the model was carried out by means of external predicting series. In this serie, the model classified correctly 81.85% (275/357 cases).

Keywords: ms-QSAR; anti-parasite; entropy indices; Artificial Neural Network

1. Introduction

There is a high interest on the search of rational approaches for antiparasite drugs discovery. In this way, theoretical studies as quantitative-structure-activity-relationships (QSAR) may play an important role. Disappointingly, QSAR studies are generally based on databases considering only structurally parent compounds acting against one single microbial species (1-7). As a consequence, to predict the antiparasite activity for a given series of compounds one have to use/seek as many QSAR models as microbial species drugs susceptibility is desirable to predict. It is very important the report of one single unified equation to calculate the probability of activity of a given drug against different antiparasite species. The basic concept of parasite infection is that parasites are a very primitive life form and have been on this planet for much longer than man has. They have, therefore, been very successful at what they do because they are not an endangered species in any place we know they exist (8, 9). In particular, the spread and present distribution of many parasites throughout the world has largely been the result of human activities, and the advent of AIDS has added a new chapter to the history of parasitology. Infections caused by parasites have increased dramatically during the past decades and there are one of the most important infectious in the world, the most important is the infection caused by *Plasmodium spp.*, millions of people have been infected each year, and millions die each year.

Up today there are near to 5000 molecular descriptors that may be in principle generalized and used to solve the former problem. In addition other QSAR approaches have been introduced recently with demonstrated utility in medicinal chemistry. In any case, no one of these indices have been extended yet to encode additional information to chemical structure. Our group has introduced elsewhere one Markov model (MM) encoding molecular backbones information, with several applications in bioorganic medicinal chemistry. The method was named the MARCH-INSIDE approach, MARKovian CHEmicals IN SILICO Design. It allowed us introducing matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties. Specifically, the stochastic spectral moments introduced by our group have been largely used for small molecules QSAR problems including design of fluckicidal, anticancer and antihypertensive drugs. Applications to macromolecules have been restricted to the field of RNA without applications to proteins (10). The entropy like molecular descriptors has demonstrated flexibility in many bioorganic and medicinal chemistry problems such as: estimation of anticoccidial activity, modeling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity (11-13).

In recent studies the MARCH-INSIDE method has been extended to encompass molecular environment interesting information in addition to molecular structure. This new interpretation allows calculating molecular thermodynamic entropy for many physicochemical and biological processes. This approach is able to take into consideration for instance not only the molecular structure of the drug but the entropy of its interaction with the specific parasite organism the drug has to eliminate too. The present study develops a single linear equation based on these previous ideas to predict the antiparasite activity of drugs against different species.

2. Methods

2.1. Markov model for drug-target step-by-step interaction

By using, Chapman-Kolmogorov equations we can calculate multi-target ${}^k C_{\theta,s}(j)$ values referred to atoms (nodes) in molecular graphs. As was mentioned above multi-target here means that we obtain different ${}^k C_{\theta,s}(j)$ values for the same atom in the same molecule when the molecular target (bacteria, virus, parasite, receptor, enzyme, etc.) change. First, we have to calculate the absolute probabilities ${}^s p_k(j)$ for the interaction in many step of different j-th atoms with the specific target. Here targets are only different microbial species (s). In this sense we insert the superscript s in the symbol of the centrality. These values can be determined as the elements of the vectors ${}^k \boldsymbol{\pi}(s)$. These vectors are elements of a Markov chain based on the stochastic matrix ${}^1 \boldsymbol{\Pi}$, which describes probabilities of interaction ${}^s p_1(i,j)$ of the j-th atom given that previously other i-th atom has interacted with the target.

$${}^k \boldsymbol{\pi}_{s=0} \boldsymbol{\pi}(s) \cdot [{}^1 \boldsymbol{\Pi}(s)]^k = [{}^s p_0(1), {}^s p_0(2), {}^s p_0(3), \dots, {}^s p_0(n)] \cdot \begin{bmatrix} {}^s p_1(1,1) & {}^s p_1(1,2) & \dots & \dots & {}^s p_1(1,n) \\ {}^s p_1(2,1) & {}^s p_1(s) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ {}^s p_1(n,1) & \dots & \dots & \dots & {}^s p_1(n,n) \end{bmatrix}^k \quad (1) \quad \mathbf{T}$$

he specificity for one target is given using target specific weights in the definition of the elements of the matrix ${}^1 \boldsymbol{\Pi}$. The theoretic foundations of the method have been given in previous works, so we do not detail it here but refer the reader to these works (14, 15). After that, the entropy centrality is very ease to calculate applying the Shannon's formula to each element ${}^s p_k(j)$ of the vectors ${}^k \boldsymbol{\pi}(s)$ and obtain the entropy centrality measures ${}^k C_{\theta,s}(j)$. As in the example 1 we can sum the ${}^k C_{\theta,s}(j)$ values for specific atom sets (AS), or the same groups of nodes, to create local molecular descriptors for the drug-target interaction. Herein the AS used were: halogens (X), insaturated carbons (C_{ins}), saturated carbons (C_{sat}), heteroatoms (Het), and hydrogens bound to heteroatoms (H-Het). The corresponding symbols of the local entropy centrality for these AS are: ${}^k C_{\theta,s}(X)$, ${}^k C_{\theta,s}(C_{ins})$, ${}^k C_{\theta,s}(C_{sat})$, ${}^k C_{\theta,s}(Het)$, ${}^k C_{\theta,s}(H-Het)$ and ${}^k C_{\theta,s}(T)$. In this study, we calculated the first six classes of entropy centrality ($k = 0$ to 5) for the 5 AS in total $6 \cdot 5 = 30$ molecular local centralities for each drug (15). At following, we give the formula for both the transition probabilities (elements of the matrix) and the atoms set entropy centrality measures. This methodology has been successfully tested previously, see the works of Gonzalez-Diaz, H. *et. al.* (16-25).

$${}^s p_0(j) = \frac{{}^s w_j}{\sum_{k=1}^n {}^s w_k} \quad (2) \quad {}^s p_1(i, j) = \frac{\delta_{ij} \cdot {}^s w_j}{\sum_{k=1}^n \delta_{ik} \cdot {}^s w_k} \quad (3)$$

$${}^k C_{\theta,s}(AS) = - \sum_{j \in AS} {}^k C_{\theta,s}(j) = - \sum_{j \in AS} {}^s p_k(j) \log [{}^s p_k(j)] \quad (4)$$

2.2. ANN models

The ANN models are non-linear models useful to predict the biological activity of a large data set of molecules. This technique is an alternative to linear methods such as LDA (26, 27). **Figure 1** depicts the network 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 (28-30).

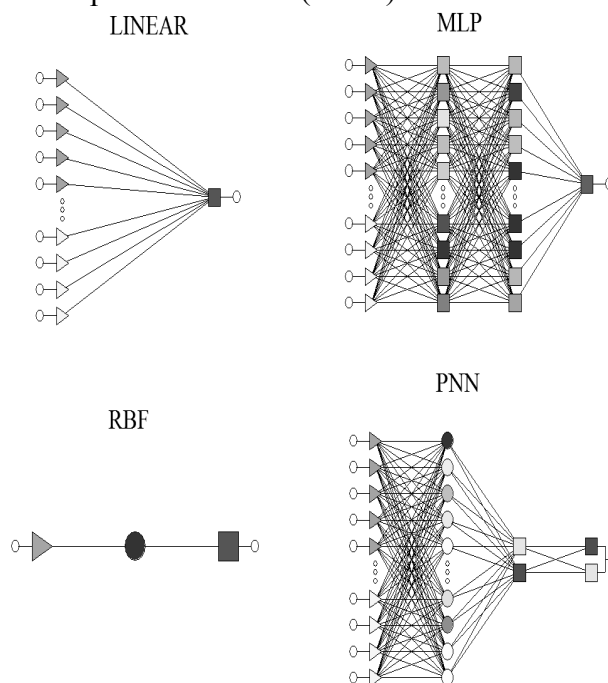


Figure 1. Depicts the networks maps for some of the ANN models used in this manuscript.

2.3. Data set

The data set was conformed by a set of marketed and/or very recently reported antiparasite drugs which reported MIC₅₀ against different virus. The data set was conformed by 500 different drugs experimentally tested against some species of a list of 16. The three data sets used were as follows training series: 115 active compounds plus 129 non-active compounds (244 in total); predicting series: 114 + 243 = 357 in total.

3. Results and discussion

The data was processed by Artificial Neural Network (ANN) classifying drugs as active or non-active against the different tested parasite species. The best ANN found was MLP 23:23-18-1:1. Overall model classification accuracy was 85.65% (211/244 cases) in training. Validation of the model was carried out by means of external predicting series. In this serie, the model classified correctly 81.85% (275/357 cases). We compare different types of networks to obtain a better model; **Table 1** shows the classification matrix of the different networks. MLP 23:23-18-1:1 was taken as the main network because it presented a wider range of variables, 23 inputs in the first layer and 23 neurons in the second layer, and two sets of cases (Training and Validation). Another tested networks found were MLP 8:8-10-1:1 presented high accuracy but only classified protein variables, PNN 190:190-14891-2-2:1 had a very low percentage of PP leading to possible errors in the model although its accuracy was very good, and a RBF 1:1-1-1:1 with a bad accuracy and presents only one variable leading possible

error in the model, see **Table 1**. We depict the ROC-curve for MLP 23:23-18-1:1 to show how reliable was the network model developed, see **Figure 1**. The network found was LNN and it showed training performance higher than 92.8%. The summary of results is shown in **Table 1**. After direct inspection of the results reported in **Table 1** for ANN methods, we can conclude that a complex ANN method is better to predict the activity than LDA.

Table 1. Comparison of ANN classification models.

Model Profile	DTP	nDTP	%	Stat. Par.	DTP	nDTP	%
PNN 42: 42-14891-2-2:1	128	1	99.22	DTP	127	2	98.45
	35	80	69.57	nDTP	106	122	53.51
			84.40	Accuracy			75.98
LNN 8: 8-10-1:1	120	9	93.02	DTP	114	15	88.37
	45	70	60.87	nDTP	90	138	60.53
			76.95	Accuracy			74.45
RBF 1: 1-1-1:1	129	0	100.00	DTP	129	0	100.00
	32	83	72.17	nDTP	81	147	64.47
			86.09	Accuracy			82.24
MLP 23:23-18-1:1	129	11	89	DTP	106	23	82.17
	20	95	82.23	nDTP	42	186	81.57
			85.65	Accuracy			81.85

We depict the ROC-curve for MLP 23:23-18-1:1 to show how reliable was the network model developed. In **Figure 2**, we depict the ROC-curve (31, 32) for MLP tested. Notably, the model presented an area under curve higher than 0.5 (the value for a random classifier). If the data are in 0.5 (black line in Figure 2), it means that our new data model predicts 50%. But the ROC-curve of our model is close to one, which means that our model predicts correctly. The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before (33); see, for instance, the work of Fernandez and Caballero (34). The same is true about the ANNs tested, where is illustrated ROC-curves of ANN MLP with an area higher than 0.98. We processed our data with ANNs looking for a better model. In general, the ANN MLP tested was statically significant (27).

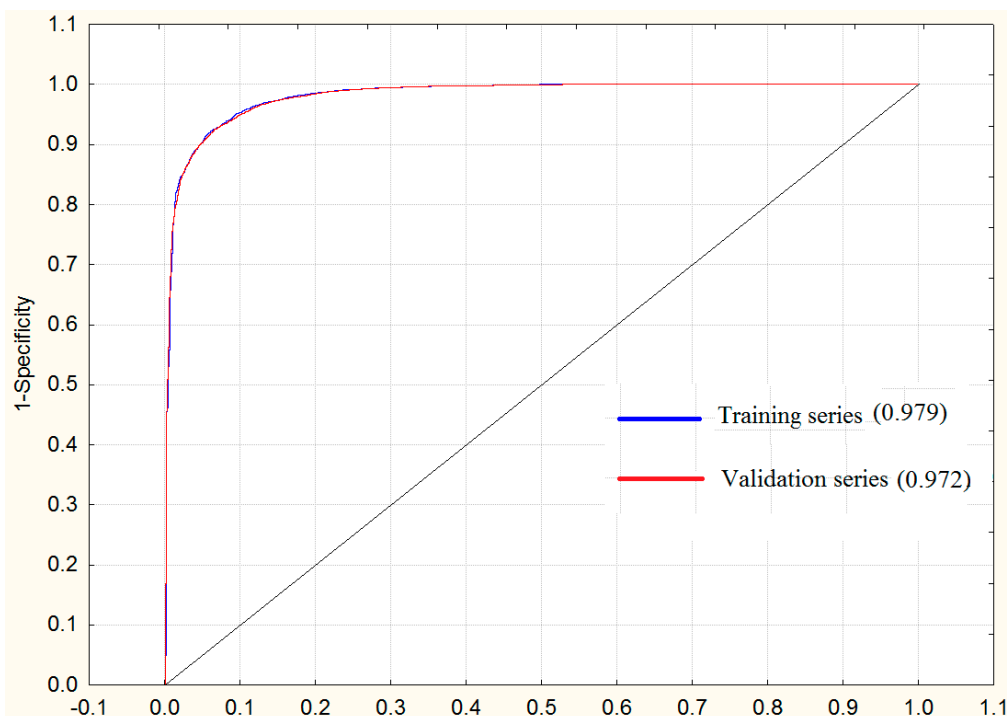


Figure 2. ROC curve for training and prediction of LNN network

Comparison with previous ML models.

The ANN model shows excellent results with a relatively small number of parameters (only 23) with respect to some previously published Machine-Learning (ML) models. To assess the importance of this result, we compared these ML models with other models used to address the same problem. For an example, have been reported a notably more complicated ML model, which included a non-linear SVM model, a large number of parameters as well as many class-to-class trials rather than the single QSAR equation used in this work (35, 36). All the other models included less than (20) input parameters or unknown parameters and some with 1000 or more (5000+) cases, and non-linear techniques such as Support vector Machine (SVM) and others (37-41). Our model is notably simpler and is the only developed in 3D parameters (proteins). However, some of these other models have low accuracy, or use ROC curve or Correlation coefficient as good classification at which makes the task more difficult (42-48), see **Table 2** for details.

Table 2. Comparison of our LDA model with ML other models.

Drugs			Targets		MODEL						
I ^a	D ^b	Type ^c	I ^a	D ^b	Method ^d	N ^e	Nv ^f	Tech. ^g	Ac ^h	P ⁱ	Ref.
θ	2	PS	2D	3	ms-QSAR	601	23	ANN	85.65	%	*
θ	3	Prot.	GT	3	MI-DRA	4485	37	ANN	87.0	%	(49)
χ	2	Prot.	ξ	3	MI	6098	15	LDA	86.3	%	(50)
χ	2	Prot.	ξ	3	MI	6098	20	ANN	90.4	%	(51)
θ	2	Prot.	TI	3	MI-DRA	2337	21	ANN	91.06	%	(52)
CTD	2	Prot.	σ	-	DTL	5367	2	SVM	94.4	%	(35)
PLB	2	Prot.	PLC	-	Relieff	1300	-	LSVM	89.9	R	(36)
S	2	MR	-	2	BLM	-	-	KRM	95.5	ROC	(37)
ΔG	2	Prot.	ΔG	-	PROPKA	26	-	-	-	-	(38)
PP	2	Prot.	PP	3	-	372	-	RSVM	84.0	%	(39)
ΔG	2	Prot.	ΔG	-	GBSA	2331	2	-	-	-	(40)
PP	2	MR	PP	1	MI	2559	10	LDA	72.3	%	(41)

^a Structural Indices: Drugs Total or Group Electronegativity (χ), Protein Total or Region Electrostatic Potential (ξ), Topological indices (TI), Getaway parameters (GT), Similarity (S), 2D-Dragon descriptors (2D), Gibbs Free energy (ΔG), Physicochemicals properties (PP), Composition (C), Transition (T), Distribution (D), Kernel parameter (σ), Protein sequence, Ligand-Structure And Binding-Pocket (PLB), Molecular Operating Environment (MOE), Surfex-Dock parameters (SD);. ^b Dimension of the structures parameters: 1D-structure, 2D-Structure and 3D structure; ^c Types of drug targets: prot. (all types of proteins), MR (Membrane Receptor protein), Enz. (Enzymes), PS (Parasite species). ^d Method: MARCH-INSIDE (MI), Bipartite Local Model (BLM), Drug-Target Likeness (DTL), pKa values for protein ligand complexes (PROPKA), comparative interaction fingerprint analysis (CoIFA), Negative Training Data (NTD), Ligand-Based drug design (LBDD), Quantum Mechanics (QM), Molecular Mechanics + Generalized Born Surface Area theory (GBSA). ^e N- Number of cases used to seek the model. ^f Nv- Number of variables in the model. ^g Statistical method used to seek the model: LDA-Linear Discriminant Analysis, SVM-Support Vector Machine, Kernel Regression-Based Method (KRM), Multi Layer Perceptron (MLP), Stepwise Multiple Linear Regression (SMLR), Genetic Function Approximation (GFA), k -Nearest Neighbors (k NN), Predictive Index (PI), Partial-Least-Square (PLS), Least-Square + Support-Vector Machine (LSVM), Radial-Basis Function + Support-Vector Machine (RSVM), Principal Component Analysis (PCA). ^h Accuracy (Ac). ⁱ Paramater used in the accuracy. * This work.

4. Conclusion

Using the MI approach, it is possible to seek for an ms-QSAR classifier to predict the probability of drugs with antiparasite activity of more than 16 different parasite species. The model can be used as a tool for preliminary screening of drugs without relaying upon geometrical optimization of drug, receptor, and drug-receptor complex structure and avoiding receptor alignment as well. We compared our model with different models already published, and concluded that a specific model of antiportozaal mt-QSAR is much more accurate than other models that only cover a single target or mt-QSAR models are seeking information from many targets, but without being specific anti-protozoa. That is why the need to develop new methodologies specific mt-QSAR on antiprotozoal, to find and design better drugs.

References

1. Fowler JH, Jeon S. The authority of Supreme Court precedent. *Social Networks*. 2008; 30:16-30.
2. Mason O, Verwoerd M. Graph theory and networks in Biology. *IET systems biology*. 2007 Mar;1(2):89-119.
3. Newman ME. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001 Jul;64(1 Pt 2):016132.
4. Newman ME. Scientific collaboration networks. I. Network construction and fundamental results. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001 Jul;64(1 Pt 2):016131.
5. Ohn JH, Kim J, Kim JH. Social network analysis of gene expression data. *AMIA Annu Symp Proc*. 2003:958.
6. De P, Singh AE, Wong T, Yacoub W, Jolly AM. Sexual network analysis of a gonorrhoea outbreak. *Sex Transm Infect*. 2004 Aug;80(4):280-5.
7. Johnson JC, Orbach MK. Perceiving the political landscape: ego biases in cognitive political networks. *Social Networks* 2002;24 291-310.
8. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics*. 2008;8:750-78.
9. 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(10):1025-39.
10. 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 MAY 2011;15(2):561-7.
11. Nunez MB, Maguna FP, Okulik NB, Castro EA. QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg Med Chem Lett*. 2004 Nov 15;14(22):5611-7.
12. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Wiley VCH. 2000.
13. 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 Aug;8(5):431-7.
14. Prado-Prado F, González-Díaz 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.
15. González-Díaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species. *Bioorg Med Chem*. 2006 Jun 5;14 5973–80.
16. Gonzalez-Diaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: predicting antifungal activity against different species. *Bioorganic & medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2006 Sep 1;14(17):5973-80.
17. Prado-Prado FJ, Gonzalez-Diaz H, Santana L, Uriarte E. Unified QSAR approach to antimicrobials. Part 2: predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorganic & medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2007 Jan 15;15(2):897-902.

18. Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Current topics in medicinal chemistry*. [Research Support, Non-U.S. Gov't Review]. 2008;8(18):1676-90.
19. Gonzalez-Diaz H, Prado-Prado FJ. Unified QSAR and network-based computational chemistry approach to antimicrobials, part 1: multispecies activity models for antifungals. *Journal of computational chemistry*. [Research Support, Non-U.S. Gov't]. 2008 Mar;29(4):656-67.
20. Prado-Prado FJ, Gonzalez-Diaz H, de la Vega OM, Ubeira FM, Chou KC. 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. *Bioorganic & medicinal chemistry*. [Comparative Study Research Support, Non-U.S. Gov't]. 2008 Jun 1;16(11):5871-80.
21. Prado-Prado FJ, Borges F, Perez-Montoto LG, Gonzalez-Diaz H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *European journal of medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2009 Oct;44(10):4051-6.
22. Prado-Prado FJ, Borges F, Uriarte E, Perez-Montoto LG, Gonzalez-Diaz H. Multi-target spectral moment: QSAR for antiviral drugs vs. different viral species. *Analytica chimica acta*. [Research Support, Non-U.S. Gov't]. 2009 Oct 5;651(2):159-64.
23. Prado-Prado FJ, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou KC, 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. *Bioorganic & medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2009 Jan 15;17(2):569-75.
24. Prado-Prado FJ, Uriarte E, Borges F, Gonzalez-Diaz H. Multi-target spectral moments for QSAR and Complex Networks study of antibacterial drugs. *European journal of medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2009 Nov;44(11):4516-21.
25. Prado-Prado FJ, Garcia-Mera X, Gonzalez-Diaz H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorganic & medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2010 Mar 15;18(6):2225-31.
26. Concu R, Podda G, Ubeira FM, Gonzalez-Diaz H. Review of QSAR models for enzyme classes of drug targets: Theoretical background and applications in parasites, hosts, and other organisms. *Curr Pharm Des*. 2010;16(24):2710-23.
27. Roy K, Mandal AS. Development of linear and nonlinear predictive QSAR models and their external validation using molecular similarity principle for anti-HIV indolyl aryl sulfones. *Journal of enzyme inhibition and medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2008 Dec;23(6):980-95.
28. Roy K, Mandal AS. Predictive QSAR modeling of CCR5 antagonist piperidine derivatives using chemometric tools. *Journal of enzyme inhibition and medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2009 Feb;24(1):205-23.
29. Roy K, Pratim Roy P. Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *European journal of medicinal chemistry*. [Comparative Study Research Support, Non-U.S. Gov't]. 2009 Jul;44(7):2913-22.

30. Patra JC, Singh O. Artificial neural networks-based approach to design ARIs using QSAR for diabetes mellitus. *Journal of computational chemistry*. 2009 Nov 30;30(15):2494-508.
31. James A, Hanley BJM. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
32. Gonzalez-Diaz H, Bonet I, Teran C, De Clercq E, Bello R, Garcia MM, et al. ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. *European journal of medicinal chemistry*. 2007 May;42(5):580-5.
33. Agüero-Chapin G, Varona-Santos J, de la Riva GA, Antunes A, Gonzalez-Villa T, Uriarte E, et al. Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from *Coffea arabica* and prediction of a new sequence. *Journal of proteome research*. 2009 Apr;8(4):2122-8.
34. Fernandez M, Caballero J, Tundidor-Camba A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. *Bioorganic & medicinal chemistry*. 2006 Jun 15;14(12):4137-50.
35. Chen X, Fang Y, Yao L, Chen Y, Xu H. Does drug-target have a likeness? *Methods Inf Med*. 2007;46(3):360-6.
36. Li J, Lei B, Liu H, Li S, Yao X, Liu M, et al. QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling. *J Comput Chem*. 2008 Dec;29(16):2636-47.
37. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*. 2009 Sep 15;25(18):2397-403.
38. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*. 2008 Nov 15;73(3):765-83.
39. Li Q, Lai L. Prediction of potential drug targets based on simple sequence properties. *BMC Bioinformatics*. 2007;8:353.
40. Yang CY, Sun H, Chen J, Nikolovska-Coleska Z, Wang S. Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. *J Am Chem Soc*. 2009 Sep 30;131(38):13709-21.
41. Vina D, Uriarte E, Orallo F, Gonzalez-Diaz H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol Pharm*. 2009 May-Jun;6(3):825-35.
42. Mascarenhas NM, Ghoshal N. Combined ligand and structure based approaches for narrowing on the essential physicochemical characteristics for CDK4 inhibition. *Journal of chemical information and modeling*. [Research Support, Non-U.S. Gov't]. 2008 Jul;48(7):1325-36.
43. Vijayan RS, Bera I, Prabu M, Saha S, Ghoshal N. Combinatorial library enumeration and lead hopping using comparative interaction fingerprint analysis and classical 2D QSAR methods for seeking novel GABA(A) alpha(3) modulators. *Journal of chemical information and modeling*. [Comparative Study Research Support, Non-U.S. Gov't]. 2009 Nov;49(11):2498-511.
44. Pham TA, Jain AN. Parameter estimation for scoring protein-ligand interactions using negative training data. *Journal of medicinal chemistry*. [Research Support, N.I.H., Extramural]. 2006 Oct 5;49(20):5856-68.

45. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, et al. The role of quantum mechanics in structure-based drug design. *Drug discovery today*. [Review]. 2007 Sep;12(17-18):725-31.
46. Zhang S, Golbraikh A, Tropsha A. Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces. *Journal of medicinal chemistry*. [Research Support, N.I.H., Extramural]. 2006 May 4;49(9):2713-24.
47. Michel J, Verdonk ML, Essex JW. Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? *Journal of medicinal chemistry*. [Research Support, Non-U.S. Gov't]. 2006 Dec 14;49(25):7427-39.
48. Naumann T, Matter H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *Journal of medicinal chemistry*. 2002 Jun 6;45(12):2366-78.
49. Prado-Prado F, Garcia-Mera X, Escobar M, Alonso N, Caamano O, Yanez M, et al. 3D MI-DRAGON: new model for the reconstruction of US FDA drug- target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. *Curr Top Med Chem*. 2012;12(16):1843-65.
50. Gonzalez-Diaz H, Prado-Prado F, Garcia-Mera X, Alonso N, Abeijon P, Caamano O, et al. 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*. *Journal of proteome research*. [Research Support, Non-U.S. Gov't]. 2011 Apr 1;10(4):1698-718.
51. Gonzalez-Diaz H, Prado-Prado F, Sobarzo-Sanchez E, Haddad M, Maurel Chevalley S, Valentin A, et al. 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*. *Journal of theoretical biology*. [Research Support, Non-U.S. Gov't]. 2011 May 7;276(1):229-49.
52. Prado-Prado F, Garcia-Mera X, Escobar M, Sobarzo-Sanchez E, Yanez M, Riera-Fernandez P, et al. 2D MI-DRAGON: a new predictor for protein-ligands interactions and theoretic-experimental studies of US FDA drug-target network, oxoisoaporphine inhibitors for MAO-A and human parasite proteins. *Eur J Med Chem*. 2011 Dec;46(12):5838-51.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).