Multi-Target Spectral Moment QSAR *vs.* ANN for antiparasitic drugs against different parasite species.

Francisco J. Prado-Prado [1,*], Xerardo García-Mera [1], and Humberto González-Díaz [2*].
[1] *Department of Organic Chemistry, Faculty of Pharmacy*
University of Santiago de Compostela, 15782 Santiago de Compostela, Spain.
[2] *Department of Microbiology & Parasitology*
University of Santiago de Compostela, 15782 Santiago de Compostela, Spain.

Abstract: There are many of pathogen parasite species with different susceptibility profile to antiparasitic drugs. Unfortunately, almost QSAR models predict the biological activity of drugs against only one parasite species. Consequently, predicting the probability with which a drug is active against different species with a single unify model is a goal of the major importance. 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 mt-QSAR model for 500 drugs tested in the literature against 16 parasite species and other 207 drugs no tested in the literature using spectral moments. The data was processed by Linear Discriminant Analysis (LDA) classifying drugs as active or non-active against the different tested parasite species. The model correctly classifies 311 out of 358 active compounds (86.9%) and 2328 out of 2577 non-active compounds (90.3%) in training series. Overall training performance was 89.9%. Validation of the model was carried out by means of external predicting series. In these series the model classified correctly 157 out 190, 82.6% of antiparasitic compounds and 1151 out of 1277 non-active compounds (90.1%). Overall predictability performance was 89.2%. In addition we developed four types of non Linear Artificial Neural Networks (ANN) and we compared with the mt-QSAR model. The improved ANN model had an overall training performance was 87%. The present work report the first attempts to calculate within a unify framework probabilities of antiparasitic action of drugs against different parasite species based on spectral moment analysis.

Keywords*:* mt-QSAR, Markov model, antiparasitic drugs, Linear Discriminant Analysis, Artificial Neural Network.
***corresponding author:** gonzalezdiazh@yahoo.es or humbertogd@uclv.edu.cu

1. Introduction

A parasite is an organism that lives on or inside another organism (the host) and causes harm to the host. Parasites that enter through the mouth are swallowed and can remain in the intestine or burrow through the intestinal wall and invade other organs. Parasites that enter through the skin bore directly through the skin or are introduced through the bites of infected insects (the vector) like *Plasmodium falciparum* that produce malaria infection[1]. In this sense, quantitative structure-activity relationships (QSAR) studies may play an important role; QSAR are used like predictive tools for the molecular development[2, 3].

As a result, there is an increasing interest on the development of rational approaches for antiparasitic drugs discovery. In this sense, a very important role may be played by computer-aided drug design techniques based on multi-target Quantitative Structure-Activity Relationships (mt-QSAR) studies. It means that they are models connecting the structure of drugs with the biological activity against different targets (microbial species in the case of antimicrobial drugs) [4, 5]. This kind of study may also help in a Multi Objective Optimization (MOOP) of desired properties or activity of drugs against different targets; see for instance the recent works carried out by Cruz-Monteagudo in the topic [6, 7]. In principle, up to date there are over 1600 molecular descriptors that may be generalized and used to solve the former problem [8, 9]. Many of these indices are known as Topological Indices (TIs) or simply invariants of a molecular graph, whose vertices are atoms weighed with physicochemical properties (mass, polarity, electro negativity, or charge) [10]. Unfortunately, most of the QSAR studies reported up-to-date are based on molecular descriptors

and databases of structurally parent compounds applicable to only one single parasite species. Therefore, it is of major interest the development of one single unified equation explaining antiparasitic activity of structurally heterogeneous series of compounds against as many species as possible [11]. In fact, other mt-QSAR approaches, with demonstrated usefulness, have been introduced recently in medicinal chemistry [12]. We introduced a Markov Model encoding molecular backbones information. The method was named the MARCH-INSIDE, MARkovian CHemicals IN SIlico Design [13]. It allowed us introducing matrix invariants such as stochastic entropies, potentials, and spectral moments for the study of molecular properties [14]. Specifically, the stochastic spectral moments introduced by our group have been largely used for small molecules mt-QSAR problems including the design of fluckicidal, anticancer and antihypertensive drugs [15]. Applications to macromolecules have been restricted to the field of RNA without applications to proteins [16-19]. In three recent reviews, we discussed the multiple applications of MARCH-INISDE to classic QSAR, macromolecular QSAR, and specially mt-QSAR [20, 21]. However, we have never used before stochastic spectral moments to develop an mt-QSAR for antiparasitic drugs.

The main steps involved in developing an mt-QSAR model are (a) selection of the data set, (b) calculation of molecular descriptors, (c) fitting the statistical model, and (d) validation of the model. Numerous different molecular descriptors have been reported to encode chemical structure in QSAR studies. Furthermore, there are multiple chemo metric approaches that can in principle be selected for step. Multiple linear regression (MLR), linear discriminant analysis (LDA), partial least squares (PLS), and different kinds of artificial neural networks (ANN) can be used to relate molecular structure (represented by molecular descriptors) with biological properties. The ANNs are particularly useful in QSAR studies in which the linear models fit poorly due to high data complexity[22, 23].

There are different kinds of ANNs, and these include multilayer perceptron (MLP), radial basis functions (RBF), and PNNs this ANN is a variant of RBF systems. In particular, PNN is a type of neural network that uses a kernel-based approximation to form an estimate of the probability density functions of classes in a classification problem[24]. In this work, we develop, for the first time, a single linear equation based on these previous ideas to predict the antiparasite activity of drugs against different species and compare a model developed by LDA analysis with several different models developed by ANN, looking for a better model.

2. Methods

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

Let us consider a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor at the time of triggering the pharmacological effect. For the sake of simplicity, from now on, we are going to consider a general structure-less molecular receptor or drug-target, understanding by structure-less receptor a receptor whose chemical structure is not taken into consideration. In our model, we approach this problem considering the free energy ${}^{k}g_{ij}(s)$ of interaction between an atom in the drug and the drug receptor after k-steps or previous interactions. We state that ${}^{k}g_{ij}(s)$ is also a state function and the symbol g points precisely to Gibbs energy. S indicates that this energy depends on the specific drug target in different microbial species. Afterwards, the interaction has to define the free energy of interaction ${}^{k}g_{ij}(s)$ between the j-th atom and the receptor for a specific microbial species (s) given that i-th atom has been interacted at a previous time t_k. So, one can suppose that, atoms begin binding to this receptor in discrete intervals of time t_k. However, there are several alternative ways in which such step-by-step binding process may occur. In this picture, the free energy ${}^{1}g_{ij}(s)$ can be defined by analogy as dependent on a constant for the atom-target interaction $\Gamma_{ij}(s)$ [15, 25]:

$$ {}^{1}g_{ij}(s) = -R \cdot T \cdot \log {}^{1}\Gamma_{ij}(s) \qquad (1) $$

The present approach to antimicrobial-receptor interaction has two main drawbacks. The first is the difficulty of defining the constants. In this work, we solve the first question by estimating the use of the ratio of occurrence $n_j(s)$ of the j-th atom on active molecules against a given species (frequency of effective

interactions) with respect to the number of atoms of the j-th class in the molecules tested against the same species $n_T(s)$. Consequently, one of the most important steps is the change on the value of the atomic weights used ${}^{k}g_{ij}(s)$ for different pathogen species. Regarding ${}^{1}\Gamma_{ij}(s)$, we must take into account that once the j-th atoms have interacted, the preferred candidates for the next interaction are those i-th atoms bound to j by a chemical bond [15]:

$$ {}^{1}\Gamma_{ij}(s) = \left(\alpha_{ij} \cdot \frac{n_j(s)}{n_T(s)} + 1 \right) = e^{\frac{{}^{1}g_{ij}(s)}{R \cdot T}} \qquad (2) $$

Where, α_{ij} are the elements of the atom adjacency matrix, $n_j(s)$, $n_T(s)$, and ${}^{1}g_{ij}(s)$ have been defined in the paragraph above, R is the universal gases constant, and T the absolute temperature. The number 1 is added to avoid forbidden negative values as inputs for the logarithmic function. The second problem relates to the description of the interaction process at higher times $t_k > t_1$. Therefore, Markov Chain theory enables a simple calculation of the probabilities with which the drug-receptor interaction takes place in the time until the antimicrobial effect is achieved. As depicted in **Figure 1**, this model deals with the calculation of the probabilities (${}^{k}p_{ij}$) with which any arbitrary molecular atom j-th binds to the structure-less molecular receptor given that other atom i-th has been bound before; along discrete time periods t_k (k = 1, 2, 3, …); (k = 1 in grey), (k = 2 in blue) and (k = 3 in red) throughout the chemical bonding system. The procedure described here considers the atoms of the molecule as states of the Markov Chain. The method arranges all the ${}^{1}g_{ij}(s)$ free energies of interaction as a squared table of n x n dimension. After normalization of the matrix we can built up the corresponding stochastic matrix **${}^{1}\Pi(s)$**, which has the elements ${}^{1}\pi_{ij}(s)$ respectively. The matrix is called the 1-step drug-target interaction stochastic matrix. **${}^{1}\Pi(s)$** is built too as a squared table of order n, where n represents the number of atoms in the molecule. The elements ${}^{1}\pi_{ij}(s)$ of the 1-step drug-target interaction stochastic matrix are the binding probabilities with which a j-th atom binds to a structure-less molecular receptor given that other i-th atoms have been interacted before at a time $t_1 = 1$[15]:

$$ {}^{1}\pi_{ij}(s) = \frac{{}^{1}g_{ij}(s)}{\sum_{a=1}^{n} {}^{1}g_{ia}(s)} = \frac{\alpha_{ij} \cdot (-RT) \cdot \log\left(\frac{n_j(s)}{n(s)} + 1 \right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot (-RT) \cdot \log\left(\frac{n_a(s)}{n_T(s)} + 1 \right)} = \frac{\alpha_{ij} \cdot \log\left(\frac{n_j(s)}{n_T} + 1 \right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1 \right)} \qquad (3) $$

Such a model is stochastic *per se* (probabilistic step-by-step atom-receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system). One can calculate the atomic spectral moments ${}^{k}\mu_s(j) = {}^{k}\pi_{jj}(s)$, values on the main diagonal (i = j) of the matrix, in order to numerically characterize the propensity with which a specific atom interacts several times with a drug receptor. In addition, the ${}^{k}\mu_s(j)$ can be summed for specific atom sets (AS) to create local molecular descriptors ${}^{k}\mu_s(AS)$ for the drug-target interaction. Herein the AS used were as follows: halogens (X), unsaturated carbons (C_{unst}), saturated carbons (C_{sat}), heteroatom (Het), and hydrogen bound to heteroatom (H-Het). The corresponding symbols of the local spectral moments for these AS are: $\mu_k(X,s)$, $\mu_k(C_{unst},s)$, $\mu_k(C_{sat},s)$, $\mu_k(Het,s)$, $\mu_k(H\text{-}Het,s)$. Finally, the sum of all atoms (it means that AS = Total contains all atoms) is useful as a total molecular descriptor.

$$ {}^{k}\mu_s(AS) = Tr\left(\left[{}^{1}\Pi(s) \right]^{k} \right)_{j \in AS} = Tr \left(\begin{bmatrix} {}^{1}\pi_{11}(s) & {}^{1}\pi_{12}(s) & . & . & {}^{1}\pi_{1n}(s) \\ {}^{1}\pi_{21}(s) & {}^{1}\pi_{22}(s) & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ {}^{1}\pi_{n1}(s) & . & . & . & {}^{1}\pi_{nn}(s) \end{bmatrix}^{k} \right)_{j \in AS} = \sum_{i=j \in AS}^{n} \pi_{ij}(s) \quad (4) $$

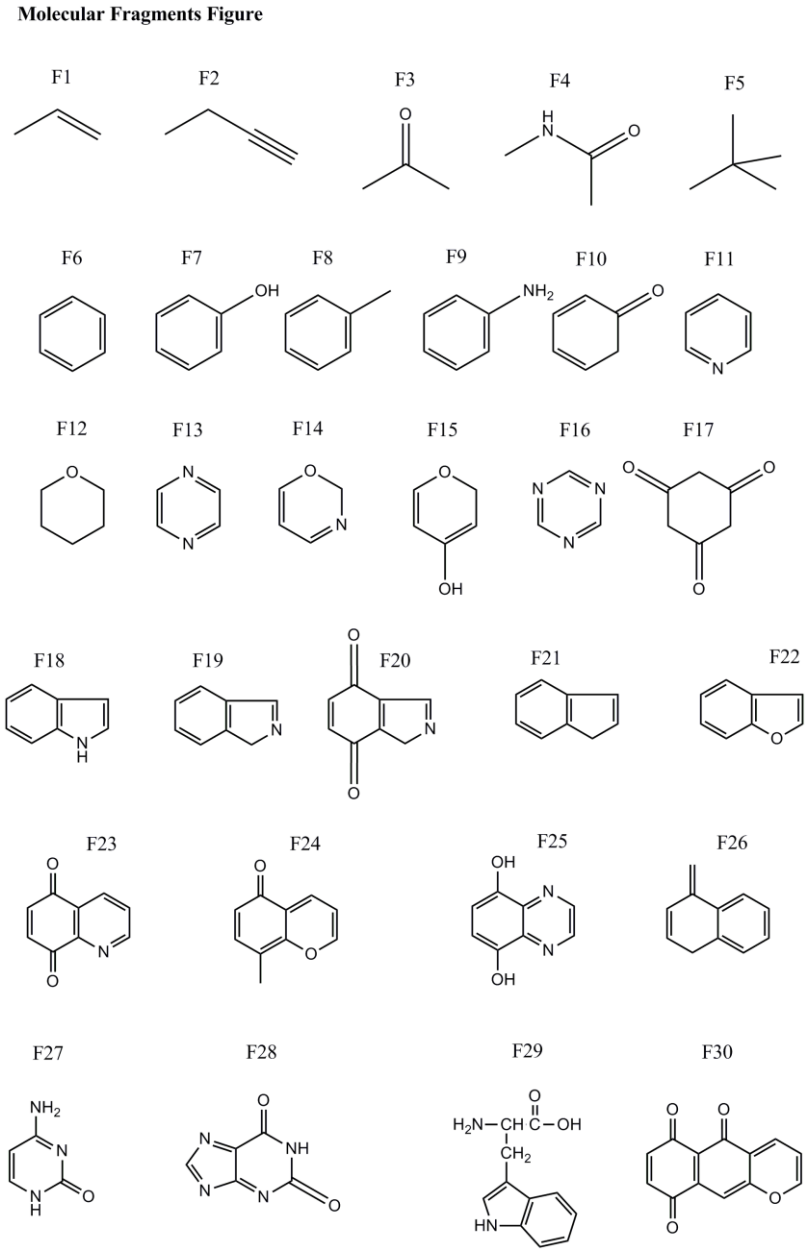

Figure 1. Molecular fragments

2.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR with the general formula:

$$Actv = a_0 \cdot {}^0\mu_s(AS) + a_1 \cdot {}^1\mu_s(AS) + a_2 \cdot {}^2\mu_s(AS) + a_3 \cdot {}^3\mu_s(AS)_s ..... + a_k \cdot {}^k\mu_s(AS) + b_0 \qquad (5)$$

Here, ${}^k\mu_s(AS)$ are the spectral moments described above that act as molecule-target interaction descriptors specific for each drug-microbial specie pair. We selected Linear Discriminant Analysis (LDA) to fit the classification functions. The model deals with the classification of a set of compounds as active or not-active against different microbial species [26]. A dummy variable (Actv) was used to codify the antimicrobial activity. This variable indicates either the presence (Actv = 1) or absence (Actv = –1) of antimicrobial activity of the drug against specific species. In equation (5), a_k represents the coefficients of the classification function, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package [27]. Forward stepwise strategy was set as the one used for variable selection[26]. The quality of LDA models was determined by examining Wilk's U statistic, Fisher ratio (F),

and the p-level (p). We also inspected the Accuracy, Sensitivity, and Specificity of the method. The Validation of the model was corroborated by external validation series [26].

2.3. Data set

The data set was formed by a set of marketed and/or very recently reported antiparasitic drugs which low reported low MIC_{50} < 10 μM against different parasite species. The data set was conformed to more than 500 drugs experimentally tested against some species of a list of 16 and other 207 drugs no tested in the literature. Not all drugs were tested in the literature against all listed species so we were able to collect 4403 cases (drug/species pairs) instead of 707 x 16 cases.

3. Results and discussion

One of the main advantages of the present stochastic approach is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MM. The generalized parameters fit on a more clearly physicochemical sense with respect to our previous ones [14, 28, 29]. More specifically, this work introduces for the first time a single linear QSAR equation model to predict the antiparasitic activity of drugs against different parasite species. The best model found was:

$$actv = 1.49^{1}\cdot\mu_s(C_{unst}) + 1.12^{5}\cdot\mu_s(C_{unst}) + 1.92^{3}\cdot\mu_s(C_{sat}) + 0.53^{4}\cdot\mu_s(X) + 1.71^{1}\cdot\mu_s(H-Het) - 0.97^{2}\cdot\mu_s(H-Het) - 5.21 \quad (6)$$

$$\lambda = 0.52 \qquad \chi^2 = 1904.6 \qquad p < 0.001$$

In the model, the coefficient λ is the Wilk's statistic; statistic for the overall discrimination, χ^2 is the Chi-square, and p the error level. In this equation, ${}^{k}\mu_s$ were calculated for the total (T) of atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic for instance: heteroatom (*Het*) and hydrogen bound to heteroatom (*H-Het*) and saturated Carbon atoms (C_{sat}) and Halogen atoms (X). The model correctly classifies 311 out of 358 active compounds (86.9%) and 2328 out of 2577 non-active compounds (90.3%) in training series. Overall training performance was 89.9%. Validation of the model was carried out by means of external predicting series. In these series the model classified correctly 157 out 190, 82.6% of antiparasitic compounds and 1151 out of 1277 non-active compounds (90.1%). Overall predictability performance was 89.2%, **see Table 1**. These results were received very fine by the authors that developed LDA and also non-linear QSAR classification models [30-46].

Parameter	%	Classes	Non-active	Antiparasitic
		Analysis		
Sensitivity	90.3	Non-active	**2328**	249
Specificity	86.9	Antiparasitic	47	**311**
Accuracy	89.9			
		Validation		
Sensitivity	90.13	Non-active	**1151**	126
Specificty	82.63	Antiparasitic	33	**157**
Accuracy	89.16			

The positive cases are in black

Table 1. Results of the model. analysis. validation.

The characteristic most interesting of the present model is that the ${}^{k}\mu_s$ used as molecular descriptors depend both on the molecular structure of the drug and the parasite species against which the drug must act. The codification of the molecular structure is basically due to the use of the adjacent factor α_{ij} to encode atom-atom bonding, molecular connectivity. The other aspect that allows encoding molecular structural changes is that the spectral moment ${}^{k}\mu_s$ are atom-class specific. This property is related to the definition of the ${}^{k}\mu_s$. The values of these species, specific atomic standard free energies reported herein for the first time, are given in **Table 2** for some atoms and more than 16 species. For example, one change in the molecular structure of, e.g. F by O, necessarily implies a change in the moments of interaction. Moreover, the most interesting fact is that ${}^{k}\mu_s$ are the molecular descriptors reported for antimicrobial mt-QSAR studies able to distinguish among a large number of parasites species.

PARASITE SPECIES	C	H	N	O	S	Cl	Br	I	F	P
Cryptosporidium parvum	0.2	0.2	0.19	0.18	0.09	0.25	0.06	0	0.27	0
Entamoeba histolytica	0.15	0.15	0.13	0.14	0	0.15	0.3	0.3	0	0.14
leishmania amazonesis	0.18	0.18	0.18	0.18	0	0.18	0	0	0	0
Leishmania donovani	0.2	0.19	0.19	0.18	0.14	0.3	0	0	0.21	0
Leishmania infantum	0.19	0.2	0.18	0.19	0.15	0.16	0.26	0	0.18	0
Leishmania major	0.16	0.17	0.2	0.13	0	0.3	0	0	0	0
Leishmania mexicana	0.19	0.19	0.18	0.2	0	0	0	0	0	0
Leishmania species	0.19	0.18	0.18	0	0	0.24	0	0	0	0
Plasmodium falciparum	0.21	0.21	0.22	0.17	0.16	0.26	0.28	0.21	0.22	0.04
Pneumocystis carinii	0.29	0.28	0.29	0.27	0	0	0	0	0	0
Toxoplasma gondii	0.22	0.22	0.24	0.22	0.16	0.19	0	0	0.29	0
Trichomonas vaginalis	0.15	0.15	0.12	0.19	0.13	0.22	0	0	0	0
Trypanosoma brucei	0.16	0.18	0.15	0.27	0.3	0.16	0.18	0.3	0	0
Trypanosoma brucei brucei	0.17	0.17	0.16	0.19	0.21	0.22	0	0	0	0
Trypanosoma brucei rhodesiense	0.22	0.21	0.2	0.22	0.18	0	0	0	0	0.3
Trypanosoma cruzi	0.19	0.19	0.21	0.18	0.19	0.24	0.12	0.3	0.12	0

Table 2. Standard atomic free energy values for atom-receptor interactions.

One application of QSAR is the calculation of the contribution of different molecular fragments to the desired activity. Unfortunately, classic QSAR models may be used only to calculate unspecific fragment contributions or contributions for only one-specie or drug target [47, 48]. In this sense, one important application of mt-QSAR is the calculation of molecular fragments contribution to activities or action against different drug targets [4]. In this work we calculated contributions of different molecular fragments for activity against different parasitic species (drug targets). As a sort of example, we selected 30 molecular fragments against 4 species (*Leishmania spp., Entamoeba hystolitica, Plasmodium falciparum and Trypanosoma spp.*); see **Figure 1**. First, we calculated the specie-dependent atomic descriptors included in the QSAR equation for selected molecular fragments using the MARCH-INSIDE software. We calculated the scores of contributions of each fragment against the 4 parasite species studied by substiting the atomic descriptors into the QSAR equation using the Microsoft Excell application. Second, the contributions of each molecular fragment were standardized using the score of contribution less total average and divided by the standard deviation. These molecular fragment contributions can indicate the potential relation between molecular fragments with the activity against different parasitic species. For example, in **Table 3** we observed a positive contribution of aniline fragment F9 anti-parasite activity against both *Leishmania spp.* and *Plasmodium falciparum*. This result coincides with experimental outcomes presented by Patil *et al.*[49] confirming the dual anti-parasite activity of aryltriazolylhydroxamates (containing F9) against both group of parasites.

Molecular Fragment	*Leishmania spp.*	*Plasmodium falciparum*	*Entamoeba histolytica*	*Trypanosoma spp.*
F1	1.33	0.88	-0.87	0.51
F2	1.16	1.36	-1.61	0.42
F3	1.54	-0.14	-1.74	0.45
F4	1.26	-0.23	-1.37	-0.85
F5	-0.58	0.52	-1.01	-0.84
F6	-0.43	-0.13	-1.05	0.51
F7	-0.29	0.03	-1.02	-1.35
F8	-0.31	0.61	-1.02	0.42
F9	**0.81**	**0.88**	-0.66	0.46
F10	0.96	-0.73	-1.01	-0.28
F11	0.02	1.80	-1.01	-0.90
F12	1.70	-1.42	-1.49	0.31
F13	-0.73	-2.33	-1.49	-0.91
F14	-0.57	0.74	-1.49	-0.84
F15	-0.57	2.59	-0.76	-0.29
F16	-0.13	-0.57	-0.61	-0.28
F17	0.17	0.57	-0.90	-0.28
F18	1.01	1.14	-0.94	-0.27
F19	1.06	1.83	-0.35	0.26
F20	0.31	1.33	-0.63	0.18
F21	1.22	1.15	-0.48	0.68
F22	-0.59	0.65	-0.39	0.62
F23	-0.43	-0.78	-0.33	1.92
F24	1.20	2.58	0.25	0.51
F25	1.06	1.22	-1.14	0.62
F26	1.20	-0.62	-1.14	-0.03
F27	-1.22	0.67	-1.15	0.58
F28	1.22	0.55	0.84	-0.19
F29	-0.29	1.64	-1.73	0.77
F30	-0.14	-0.85	-1.66	0.18
max	1.70	2.59	0.84	1.92
min	-1.22	-2.33	-1.74	-1.35

Table 3. Molecular fragments contributions for activity.

3.1. Comparison of Linear (LDA) versus Nonlinear (ANN) Classifiers.

We processed our data with different Artificial Neural Networks (ANNs) looking for a better model. Four types of ANNs were used, namely, Probabilistic Neural Network (PNN), Radial Basic Function (RBF), Three Layers Perceptron (MLP-3), and Four Layer Perceptron (MLP-4). **Figure 2** depicts the networks maps for some of the ANN models tested. In general, at least one ANN of every type tested was statically significant. However, one must note that the profiles of each network indicate that these are

highly nonlinear and complicated models. For instance, PNN 6:6-2935-2-2:1 is an ANN with six inputs, six neurons in the first layer, 2 935 neurons in the second layer, two sets of cases (Training and Validation).

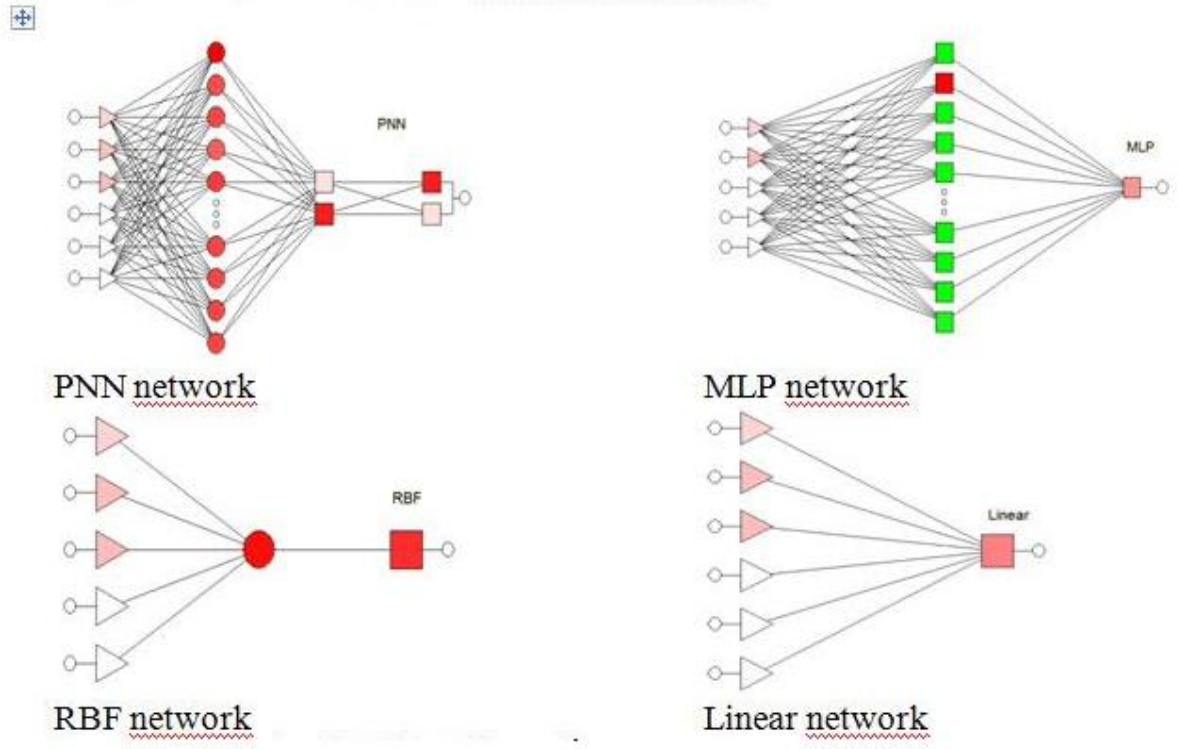

Figure 2. Representation of ANN´s models tested.

In **Figure 3**, we depict the ROC-curves for different ANNs tested. Notably, almost models presented and are under curve higher than 0.5. The all MLP models presented an area greater than 0.93 except 1. The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before; see, for instance, the work of Fernandez and Caballero[50]. The same is true about the other kinds of ANNs tested.

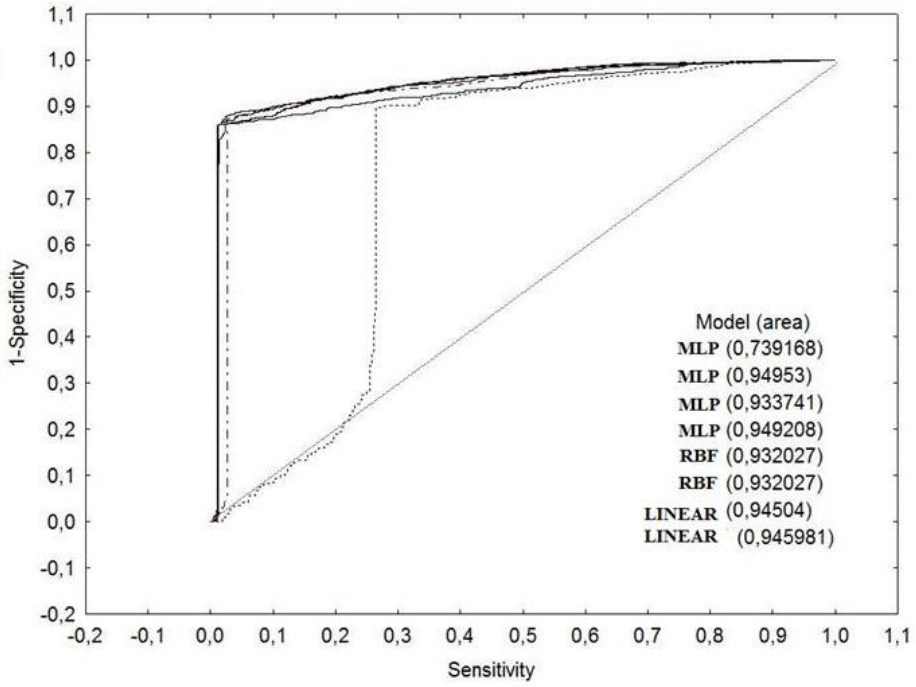

Figure 3. ROC curve of different ANN´s models tested.

The best network found was PNN and it showed training performance higher than 87%. The summary of results is depicted in **Table 4**. The other networks were MLP 6:6-10-6-1:1, RBF 5:5-1-1:1 and Linear

6:6-1:1. After direct inspection of the results reported in table 3 for ANN methods, we can conclude that there is no necessity to use a complex ANN method to obtain a better classification model in comparison with LDA method.

Profile	PNN 6:6-2935-2-2:1		MLP 6:6-10-6-1:1		RBF 5:5-1-1:1		Linear 6:6-1:1	
	1	0	1	0	1	0	1	0
Training								
n	358	2577	358	2577	358	2577	358	2577
Correct(%)	0	100	99	67	99	66	99	63
Total(%)	87.8		70.8		70.1		67.6	
Validation								
n	190	1277	190	1277	190	1277	190	1277
Correct(%)	0	100	98	63	98	65	98	70
Total(%)	87.0		67.6		68.9		74.0	

Table 4. Train and Validation Accuracy Results for ANN model.

Conclusion.

Using the MARCH-INSIDE approach is possible to seek a useful QSAR classifier for active/non-active drugs, which scores multi-species antiparasitic activity of chemicals. As a conclusion, and concerning the future research outlook, one can note that the present mt-QSAR methodology with a large data set improves the results significantly and allows us to obtain more realistic and correctly results. The mt-QSAR methodology may be able to predict the biological activity of drugs in more general situations than the traditional QSAR models which the most limitation is predict the biological activity of drugs against only one parasites species. We conclude that using ANN models do not improve the performance of the model using LDA analysis, actually only one network PNN achieves good performance, so the use of ANN is not essential to an improve mt-QSAR model.

Acknowledgments

The authors sincerely thank the king attention as well as useful comments from the editor Prof. A. Monge-Vega and two unknown referees. Prado-Prado, F. acknowledges the financial support from program Angeles Albariño and González-Díaz, H. acknowledges financial support of the Programme Isidro Parga Pondal both funded by the Xunta de Galicia and European Social Fund (ESF).

References.

1. Cooke, B. M.; Mohandas, N.; Coppel, R. L. *Semin Hematol* **2004,** *41*, 173.
2. Chou, K. C. *Current Medicinal Chemistry* **2004,** *11*, 2105.
3. Chou, K. C., D. Q. Wei, Q. S. Du, S. Sirois, and W. Z. Zhong. . *Current Medicinal Chemistry* **2006,** *13*, 3263.
4. Prado-Prado, F. J.; Gonzalez-Diaz, H.; de la Vega, O. M.; Ubeira, F. M.; Chou, K. C. *Bioorg Med Chem* **2008,** *16*, 5871.
5. Prado-Prado, F. J.; de la Vega, O. M.; Uriarte, E.; Ubeira, F. M.; Chou, K. C.; Gonzalez-Diaz, H. *Bioorg Med Chem* **2009,** *17*, 569.
6. Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N.; Cagide Fajin, J. L.; Morell, C.; Ruiz, R. M.; Canizares-Carmenate, Y.; Dominguez, E. R. *J Comb Chem* **2008,** *10*, 897.
7. Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N. *J Comput Chem* **2008,** *29*, 2445.
8. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*: Wiley-VCH, 2002.
9. Kubinyi, H. *J Cancer Res Clin Oncol* **1990,** *116*, 529.
10. Estrada, E.; Molina, E. *J Chem Inf Comput Sci* **2001,** *41*, 791.
11. González-Díaz, H.; Prado-Prado, F. *J Comput Chem* **2008,** *29*, 656.
12. Marrero-Ponce, Y.; Diaz, H. G.; Zaldivar, V. R.; Torrens, F.; Castro, E. A. *Bioorg Med Chem* **2004,** *12*, 5331.
13. González-Díaz, H.; Molina-Ruiz, R.; Hernandez, I., 2007.
14. González-Díaz, H.; Aguero, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castanedo, N. *Bioorg Med Chem Lett* **2005,** *15*, 551.
15. Prado-Prado, F.; González-Díaz, H.; Santana, L.; Uriarte, E. *Bioorg Med Chem* **2007,** *15*, 897.
16. González-Díaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E. *J Proteome Res* **2007,** *6*, 904.
17. González-Díaz, H.; Sanchez-Gonzalez, A.; Gonzalez-Diaz, Y. *J Inorg Biochem* **2006,** *100*, 1290.
18. González-Díaz, H.; Cabrera-Pérez, M. A.; Agüero-Chapín, G.; Cruz-Monteagudo, M.; Castañedo-Cancio, N.; del Río, M. A.; Uriarte, E. *Chemometrics and Intelligent Laboratory Systems* **2008,** *94*, 160.
19. Cruz-Monteagudo, M.; González-Díaz, H.; Agüero-Chapin, G.; Santana, L.; Borges, F.; Domínguez, R. E.; Podda, G.; Uriarte, E. *J Comput Chem* **2007,** *28*, 1909.
20. González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F. M.; Uriarte, E. *Proteomics* **2008,** *8*, 750.
21. Gonzalez-Diaz, H.; Prado-Prado, F.; Ubeira, F. M. *Curr Top Med Chem* **2008,** *8*, 1676.
22. Vanyur, R.; Heberger, K.; Jakus, J. *J Chem Inf Comput Sci* **2003,** *43*, 1829.
23. Vilar, S.; Santana, L.; Uriarte, E. *J Med Chem* **2006,** *49*, 1118.
24. Mosier, P. D.; Jurs, P. C. *J Chem Inf Comput Sci* **2002,** *42*, 1460.
25. González-Díaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. *Bioorg Med Chem* **2006,** *14* 5973.
26. Van Waterbeemd, H. In *Chemometric methods in molecular design;* Van Waterbeemd, H., Ed.; Wiley-VCH: New York, 1995; Vol. 2, 265.
27. StatSoft.Inc., 2002.
28. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg Med Chem* **2005,** *13*, 1119.
29. Cruz-Monteagudo, M.; González-Díaz, H. *Eur J Med Chem* **2005,** *40*, 1030.
30. Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. *J Chem Inf Comput Sci* **2002,** *42*, 1194.
31. Vilar, S.; Cozza, G.; Moro, S. *Curr Top Med Chem* **2008,** *8*, 1555.
32. Vilar, S.; Gonzalez-Diaz, H.; Santana, L.; Uriarte, E. *J Comput Chem* **2008,** *29*, 2613.
33. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Escobar, J.; Torrens, F.; Rotondo, R. *Chemosphere* **2008,** *73*, 415.
34. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. *J Mol Graph Model* **2007,** *26*, 32.

35. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sanchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg Med Chem* **2005,** *13*, 1005.
36. 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. *Bioorg Med Chem* **2006,** *14*, 6502.
37. Shen, H.; Chou, K. C. *Biochem Biophys Res Commun* **2005,** *334*, 288.
38. Chou, K. C.; Shen, H. B. *J Cell Biochem* **2006**.
39. Wang, S. Q.; Yang, J.; Chou, K. C. *J Theor Biol* **2006**.
40. Feng, K. Y.; Cai, Y. D.; Chou, K. C. *Biochem Biophys Res Commun* **2005,** *334*, 213.
41. Cai, Y. D.; Feng, K. Y.; Lu, W. C.; Chou, K. C. *J Theor Biol* **2006,** *238*, 172.
42. Shen, H. B.; Yang, J.; Chou, K. C. *Amino Acids* **2007,** *33*, 57.
43. Chen, J.; Liu, H.; Yang, J.; Chou, K. C. *Amino Acids* **2007,** *33*, 423.
44. Shen, H. B.; Yang, J.; Chou, K. C. *Expert Rev Proteomics* **2007,** *4*, 453.
45. Zhang, T. L.; Ding, Y. S.; Chou, K. C. *J Theor Biol* **2008,** *250*, 186.
46. Shen, H. B.; Chou, K. C. *Protein Eng Des Sel* **2007,** *20*, 561.
47. Estrada, E.; Uriarte, E.; Molina, E.; Simon-Manso, Y.; Milne, G. W. *J Chem Inf Model* **2006,** *46*, 2709.
48. Estrada, E.; Molina, E. *J Mol Graph Model* **2006,** *25*, 275.
49. Patil, V.; Guerrant, W.; Chen, P. C.; Gryder, B.; Benicewicz, D. B.; Khan, S. I.; Tekwani, B. L.; Oyelere, A. K. *Bioorg Med Chem* **2009**.
50. Fernandez, M.; Caballero, J.; Tundidor-Camba, A. *Bioorg Med Chem* **2006,** *14*, 4137.