[g014]

QSAR & Network-based multi-species activity models for antifungals

FRANCISCO J. PRADO-PRADO, HUMBERTO GONZÁLEZ-DÍAZ,*
LOURDES SANTANA, AND EUGENIO URIARTE.

Unit of Bioinformatics & Connectivity Analysis (UBICA), Institute of Industrial Pharmacy and Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela 15782, Spain.

Abstract. There are many pathogen microbial species with very different antimicrobial drugs susceptibility. In this work, we selected pairs of antifungal drugs with similar/dissimilar species predicted-activity profile and represented it as a large network, which may be used to identify drugs with similar mechanism of action. Computational chemistry prediction of the biological activity based on quantitative structure-activity relationships (QSAR) susbtantialy increases the potentialities of this kind of networks avoiding time and resources consming 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, which outputs were the inputs of the above-mentioned network. Overall model classification accuracy was 87.0% (161/185 compounds) in training, 83.4% (50/61) in validation, and 83.7% for 288 additional antifungal compounds used to extent model validation for network construction. The network predicted has 59 nodes (compounds), 648 edges (pairs of compounds with similar activity), low coverage density d = 37.8%, and distribution more close to normal than to exponential. These results are more characteristic of a not-over-estimated random network, clustering different drug mechanisms of actions, than of a less useful power-law network with few mechanisms (network hubs).

Keywords: Molecular descriptor, Markov model, Networks, QSAR, co-expression network, Probability, Antimicrobials, Antifungals.

* ***Correspondence to***: GONZÁLEZ-DÍAZ, H. Faculty of Pharmacy, University of Santiago de Compostela 15782, Spain. Email: gonzalezdiazh@yahoo.es, Tel: +34-981-563100, Fax: +34-981 594912.

SHORT TITLE: UNIFY QSAR & NETWORK APPROACH TO ANTIFUNGALS

Introduction

There is a high interest on the search of rational approaches for antimicrobial drugs discovery. In particular, fungi-caused infections have increased dramatically during the past decades. Systemic mycoses mainly appear concomitant with other diseases or are caused by treatment with chemotherapeutics, for instance with cytostatics. At risk are patients after organ transplantation treated with immunosuppressives or those suffering with a weakened immune system, for example patients with AIDS. In this sense, quantitative structure-activity relationships (QSAR) studies may play an important role. Disappointingly, QSAR studies are generally based on databases considering only structurally parent compounds acting against one single microbial species.[1-3]

There are more than 1 600 molecular descriptors that may be in principle generalized and used to solve the former problem.[4] In any case, no one of these indices have been extended yet to encode information additional to chemical structure.[5-7] Our group has introduced elsewhere a Markov model (MM) method named MARCH-INSIDE, MARkovian CHemicals IN SIlico Design. AMRCH-INSIDE use matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties.[8-10] Stochastic spectral moments have been used for QSAR problems including the design of antimicrobial and anticancer drugs as well as for RNA and proteins QSAR problems.[11-15] Otherwise, entropy like parameters has demonstrated flexibility to treat many problems.[16-20] In recent studies the MARCH-INSIDE method has been extended to encompass molecular environment information in addition to molecular structure calculating thermodynamic free energy for many physicochemical and biological processes.[21,22] This approach take into consideration the molecular structure and the effect of the anitimicrobial over different species.[23,24]

In fact, there are many pathogen microbial species whith very different antimicrobial drugs susceptibility. This very high number of drug-species combinations may be investigated using networks to group or cluster drugs with similar multi-species activity profile and possibliy mechanism of action. We can use different classes of networks such as: artifical neural networks (ANN) [25-36] to mining datasets, depicting relationships between within the genetic code,[37-43] or representing relationships between proteins, genes, RNAs, organisms, or even non-living objects.[44-58] Specifically, co-expression networks can be constructed by measuring the expression of pairs or genes in different tissues.[59-63] Similarly, protein networks can be constructed from pair-wise experimentally or theoretically stablished protein-protein interactions.[44,64-66] In co-expression networks two RNAs are connected (supposed to be involved in common mechanism of regulation) if the levels of both RNAs for different tissues strongly correlate.[67] We propose to use the same network approach to study mulit-species antimicrobials drug action. In this case, the antimicrobial drug plays the role of the RNA molecule and the drug activity against different species activity play the role of RNA level of expression in different tissues. In the co-expression network we need to measure each RNA tissue profile if we do not have a computational approach to predict it.[59,68] For antimicrobials networks; we need to measure the activity of the drug against different species. Consequently, the QSAR method used for antimicrobials multi-species network construction must be able to make multi-species prediction of the antimicrobial activity.

In this work, we selected by the first time pairs of antifungal drugs with similar/dissimilar multi-species predicted-activity profile and represented it as a large network, which may be used to identify drugs with similar mechanism of action. First, we developed a multi-species QSAR classification model, which ouputs were the inputs of the above-mentioned network. Next, we used the outputs of this QSAR model to construct a network with low coverage density and normal-like distribution for antifungal compounds having similar multi-species activity. These results are more characteristic of a not-over-estimated random network, clustering different drug mechanisms of actions, than of a less useful power-law network with few mechanisms (network hubs).[69] We illustrate the use of the network for azole drugs such as voriconazole, miconazole, fluconazole and others. The present work reports by the first time the use of QSAR computational techniques to construct multi-species activity networks for antimicrobial drugs.

Materials and methods

Absolute probabilities for drug-target step-by-step interaction

By using, Chapman-Kolgomorov equations we can calculate the absolute probabilities ${}^{A}\pi_k(j,s)$ for the interaction in many step of different j-th atoms with the receptors in different microbial species (s). The ${}^{A}\pi_k(j,s)$ can be determined as the elements of the vectors $\mathbf{{}^{k}\pi(s)}$. These vectors are elemnts of a Markov chain based on the stochastic matrix $\mathbf{{}^{1}\Pi}$, which describes conditional probabilities of interaction of the j-th atom given that previously other i-th atom has interacted with the receptor. 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:[23,24]

$$ {}^{k}\pi_s = {}^{0}\pi \cdot \left[{}^{1}\Pi\right]^{k} = \left[{}^{A}\pi_0(1,s) \;\; {}^{A}\pi_0(2,s) \;\; {}^{A}\pi_0(3,s) \;\; \cdots \;\; {}^{A}\pi_0(n,s) \right] \cdot \begin{bmatrix} {}^{1}\pi_{11} & {}^{1}\pi_{12} & \cdots & \cdots & {}^{1}\pi_{1n} \\ {}^{1}\pi_{21} & {}^{1}\pi_{22} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ {}^{1}\pi_{n1} & \cdots & \cdots & \cdots & {}^{1}\pi_{nn} \end{bmatrix}^{k} \quad (2) $$

The ${}^{A}\pi_k(j,s)$ can be summed for specific sets of atoms (AS) 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 absolute probabilities for these AS are: ${}^{A}\pi_k(X,s)$, ${}^{A}\pi_k(C_{ins},s)$, ${}^{A}\pi_k(C_{sat},s)$, ${}^{A}\pi_k(Het,s)$, ${}^{A}\pi_k(H\text{-}Het,s)$. In this study, we calculated the first six classes of probabilities (k = 0 to 5) for the 5 AS in total 6·5 = 30 molecular descriptors.[23]

Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:[10]

$$ Actv = {}^{c}b_0 \cdot {}^{A}\pi_0(C,s) + {}^{c}b_1 \cdot {}^{A}\pi_1(C,s) + {}^{c}b_2 \cdot {}^{A}\pi_2(C,s) + {}^{c}b_3 \cdot {}^{A}\pi_3(C,s) + \ldots + {}^{c}b_k \cdot {}^{A}\pi_4(C,s) + b \quad (3) $$

Here, the absolute probabilities ${}^{A}\pi_k(C,s)$ play the role of molecule-target interaction descriptors for specific microbial species. We selected Linear Discriminant Analysis (LDA)[70] to fit the classification functions. The model deals with the classification of a set of compounds as active or not against different microbial species. A dummy variable (Actv) was used to codify the antimicrobial acitivity. This variable indicates either the presence (Actv = 1) or absence (Actv = –1) of antimicrobial activity of the drug against the microbe species in question. In equation (8), b_k represents the coefficients of the classification function, determined by the LDA module of the STATISTICA 6.0 software package[71] using forward stepwise strategy for variable selection.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 percentage of good classification. Validation of the model was corroborated with external prediction series.[22]

Data set

The data set was conformed by a set of marketed and/or very recently reported antifungal drugs which low reported $MIC_{50} < 10$ μM against different fungus. The three data sets used were as follows training series: 107 active compounds plus 78 non-active compounds (185 in total); predicting series: 36 + 74 = 110 in total; virtuals screening 288 active compounds. The literature reports experimental test of each drug against some but not all species of a list of 87. In consequence we were able to collect 583 cases (drug/species pairs). The names or codes for all compounds are depicted in **Table 1SM** (upon request to authors) of the supplementary material by reasons of space as well as the references consulted to compile the data, which appear bellow this table.

Network construction

In order to perform the antimicrobials multi-specie activity with a network approach we carried out the following steps:

1. First, we calculated the molecular descriptors include in the QSAR equation for 59 selected drugs using the MARCH-INSIDE software.[72]
2. We calculated the scores of biological activity of each one of the 59 drugs against all the fungus species studied here by substiting the molecular descriptors into the QSAR equation using the Microsoft Excel application.[73]
3. All the activity scores predicted were organized into a Table of drugs (rows) vs. species (columns), which was used as input for the software STATISTICA employed to calculate drug-drug multispecies correlations in the form of Pearson r coefficients. These correlations were represented actually as distances (1-Pearson r) between drugs pairs. The Pearson distance matrix was derived using the software package STATISTICA.[71]
4. Using Microsoft excel[73] again we transformed the drug pair distances matrix derived with Statistic into into a Boolean matrix. The elements of this matrix are equal to 1 if two drugs have a high correlation or the same are very close (short distance (1-Pearson r)). The threshold value used was a distance of 0.005. The line command used in Excel to transform the distance matrix into a Boolean matrix was f = if (A$1=$B2,0, if (B2>0.0051, 0, 1)). It allows transforming distance into Boolean values and equals the main diagonal elements to 0 avoiding loops in the future network. The Boolean matrix was saved as a txt format file.
5. After, renamed the .txt file as a .mat file we read it with the software CentiBin.[74,75] Using CentiBin we can not only represent the network but also highlight all drugs (nodes) connected to a specific drug and calculating many parameters including node degree.
6. The ChemOffice software[76] was used to draw and calculate topological indices (TIs): network radius (R_i), network diameter (D_i), sum of vertex degrees network radius (δ_i), and Wiener index (W) for small networks used in the example of related to **Table 1**. The loops (LP_i) or hub presence (HP_i) were easily calculated by visual inspection. Details on the definition of TIs and their uses for small molecules, macromolecules and complext networks can be found in the literature.[1,77-79]
7. CentiBin software was used to generate random networks by five different algorithms including: Barabasi-Albert random network, Kleinberg small wolrd network (SWN), 2D Lattice network, Erdos-Renyi network and Epsstein power law network (PLN).[75]

8. Last, all node degrees were used as input in STATISTICA in order to study the distribution of the network and compare it to other ideal network distributions including normal, exponential, gamma, and chi-square.[71]

Results and discussion

Training and validation of the model

This work introduces by the first time a single linear QSAR equation model to predict the antifungal activity of drugs against more than different 70 species based only on. The best model found was:

$$actv = -0.49 \cdot {}^{A}\pi_5 (s, C_{Sp\&Sp_2}) + 2.57 \cdot {}^{A}\pi_0 (s, X) + 1.43 \cdot {}^{A}\pi_0 (s, H - Het) + 0.90 \quad (1)$$

$$Rc = 0.75 \qquad \lambda = 0.44 \qquad p < 0.001$$

Where, Rc it is the canonical regression coefficient, λ it is the Wilk's statistics; and p the error level. In this equation, the absolute probabilities ${}^{A}\pi_k$ calculated refers to:

1. ${}^{A}\pi_5$(s, $C_{Sp\ \&\ Sp2}$) all unsaturated Carbon atoms (Sp and Sp_2 atoms) and all atoms placed at five or least atoms from them.
2. ${}^{A}\pi_0$(s,X) all halogens atoms.
3. ${}^{A}\pi_0$(s, H-Het) all Hydrogen atoms bound to a Heteroatom (N, O, or S).

The model, with only three variables, correctly classifies 90 out of 107 active compounds (84.1%) and 71 out of 78 non-active compounds (91.02%). Overall training accuracy was 87.02% (161 out of 185 compounds). Validation of the model was carried out by means of external predicting series. The model correctly classifies 30 out of 36 active compounds (83.3%) and 20 out of 25 non-active compounds (80.0%) in prediction series. Overall predictability was 83.38% (50 out of 61 compounds). Values in the range of 80 to 100 % are accepted as high accuracy for many authors that reported QSAR models based on LDA, including unique-specie antimicrobial QSAR models.[6,80-88] The present is the first model to predict the antifungal activity of any organic compound against a very large diversity of species based on molecular MM absolute probabilities, hence considering that the present is a multi-species QSAR the result is very good. The **Figure 1** illustrates this idea depicting overall prediction of the biological activity of broad spectrum antifungal drugs.

Two possible applications for the present model are the biomolecular screening of antifungals active against different species and the construction of multi-species activity profile networks for antifungals. In both cases, species susceptibility identification is imperative. For instance, the model recognizes 100% of the species studies that can be treated with ketoconazole. Detailed information on the names, predicted classification, and probability of action against different species of the drugs used to seek the model appear in **Table 1SM** of the supplementary material. The details of the forward-stepwise process for variable selection appear in the **Table 2SM** of the same supplementary material (upon request to authors).

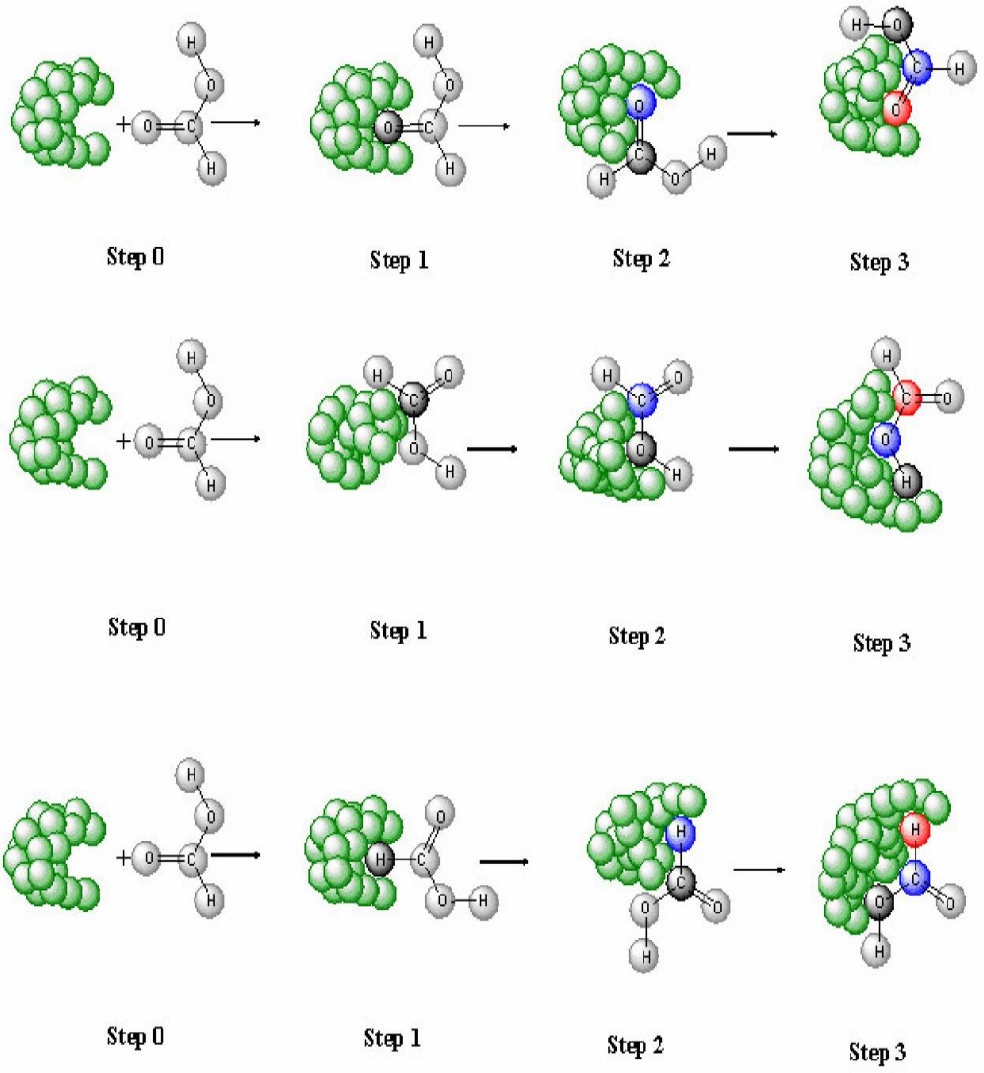


Figure 1. Overall prediction of the biological activity of broad spectrum antifungal drugs.

Computational chemistry based virtual screening experiment

A model for multi-species screening of antifungals and construction of multi-species activity profile networks necessarily have to be based on as diverse as possible series of chemical structural patterns. The compounds used to seek the model are structurally heterogeneous. However, in order to offer additional evidence on the validity of the model and also show how to use it in practice we carried out a virtual screening experiment. In this study, we try to predict the result of 288 positive activity tests for different compounds with diverse species. These results where never used in training or predicting series above. The model was able to correctly predict 241 out of 288 tests (83.68%). All these results were depicted in detail on **Table 2SM** of supplementary material (upon request to authors). Finally, the high potential of the present model to select

broad spectrum antifungal drugs can be illustrated also from the point of view of prediction of species multidrug susceptibility. The **Figure 2** depicts some selected values of the overall prediction of the antifungal drug susceptibitliy of selected species. For instance the model identifies 80.0% of the drugs that can be used to control *Candida spp*.

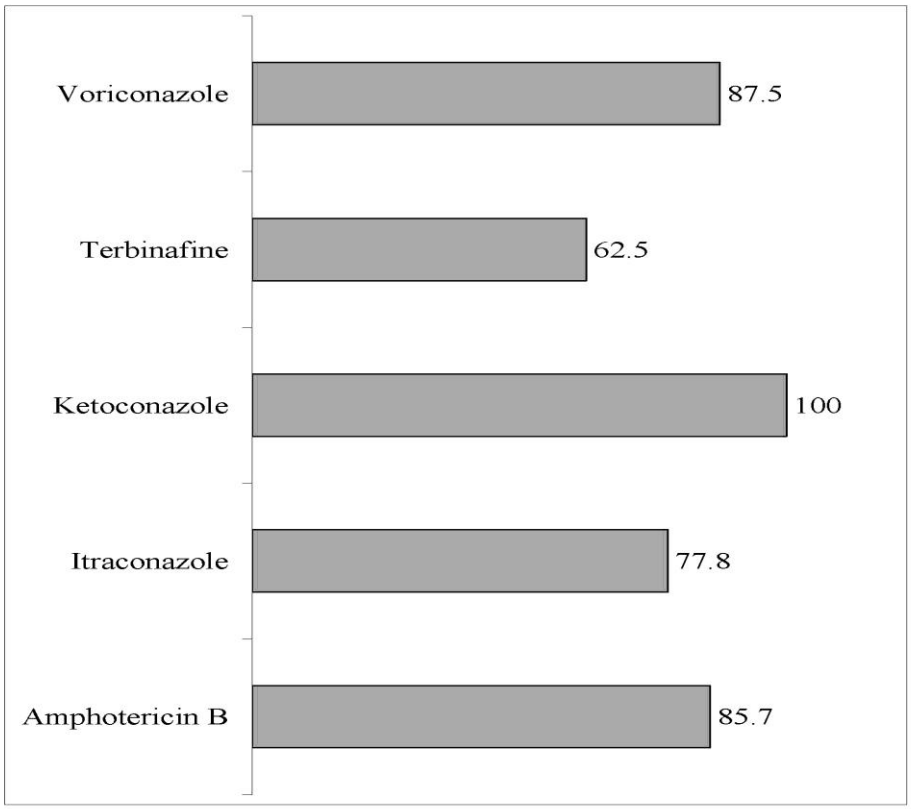


Figure 2. Overall prediction of the antifungal drug susceptibitliy of selected species.

Network approach to multi-species activity profile of antifungal drugs

First, used the multi-species QSAR predict the biological activity of 59 antifungal drugs against all the species studied. After correlation of activity score predicted with the QSAR equation of all possible pairs of drugs we decided which pairs of drugs have similar or dissimilar activity profile. The network has 59 nodes (compounds). We can determine a correlation threshold at which two genes are assessed to be co-expressed using a clustering coefficient.[89] We applied the same reckoning to pairs of antifungal drugs and decided the pairs of compounds connect to each other within the network after tree joining cluster analysis based on the 1-Pearson r. The **Figure 3** illustrates the tree joining formation of different clusters of compounds at different distances. The use of Tree joining clustering have been well documented in QSAR for clustering of antimicrobial compounds and in phylogenetics implications over organisms networks.[90,91]

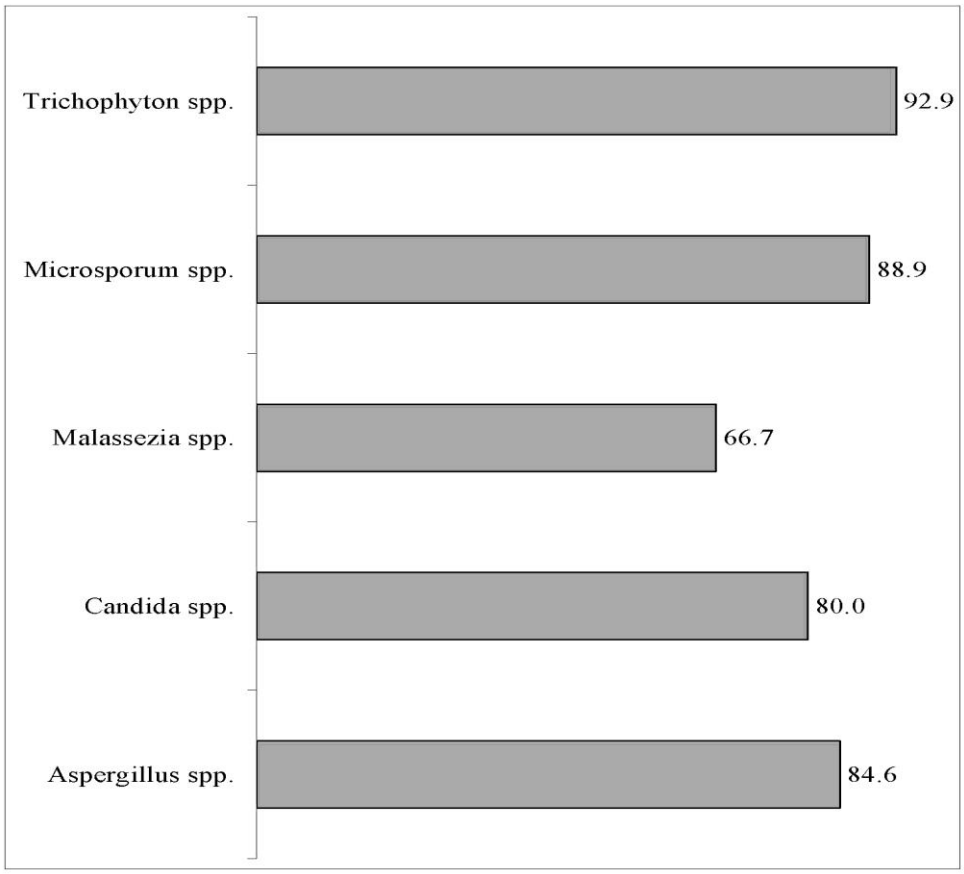


Figure 3. Clustering of antifungals drugs based on predicted species susceptibilities

We decided to use as threshold value for dissimilarity between the multi-species activities of two drugs the value distance-treshold = 0.0051. This threshold distance value was selected after inspection of the sinlge-linkage clustering of compounds in order to avoid network overcrowding. The **Figure 4** illustrates the distribution of drug-drug activity dissimilarity across linking steps for pairs of drugs. Using the combinatorial formula[88] we can calculate n!/[(n -2)!·2!] = 59!/(57! ·2!) = 59·58·57!/(57! ·2!) = (59·58)/2 = 1711 possible pairs of drugs by analogy to gene co-expression networks.[92]

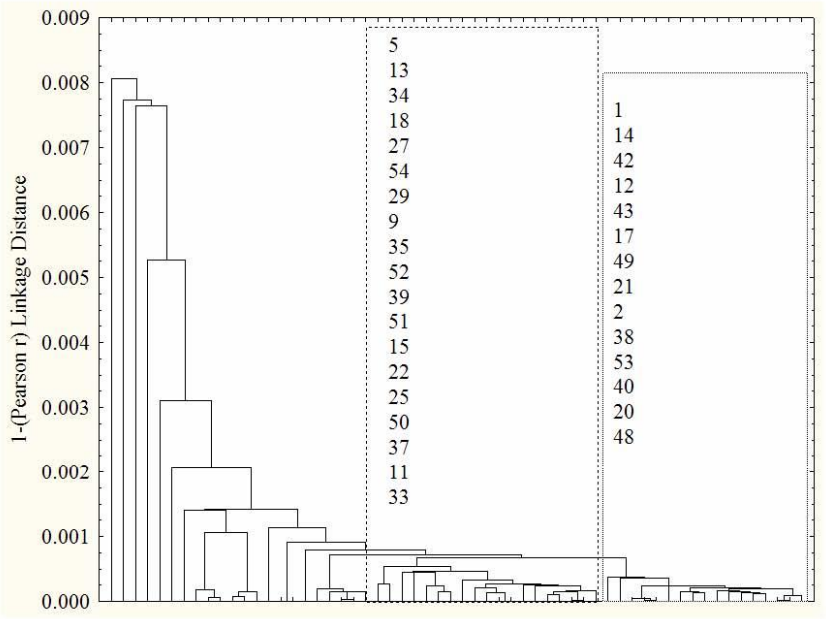


Figure 4. Plot of linkage distances across steps

Our multi-species QSAR predict 648 pairs of drugs with similar activity (dissimilarity lower than 0.0051) out of the 17711 possible pairs. So, we can predict low network edges coverage density d = 648/1711 = 37.9%.[92,93] Having a relatively low d is very important to avoid a netwrok that over-stimates thenumber of mechanism of actions for a drug or simply give so many possible mechanism to be investigated that becomes missuseful the prediction. The **Figure 5** depicts an overall representation of the present network in the CentiBin software interface.[94] We also give an example on the use of the network for the identification of similar mechanism of action for azole class of durgs such as: voriconazole, miconazole, fluconazole and others. Azole class is one of the more classic classes of anifungal drugs but the the synthesis, testing and QSAR study of novel azole derivatives constitutes a very promising field nowadays.[95]

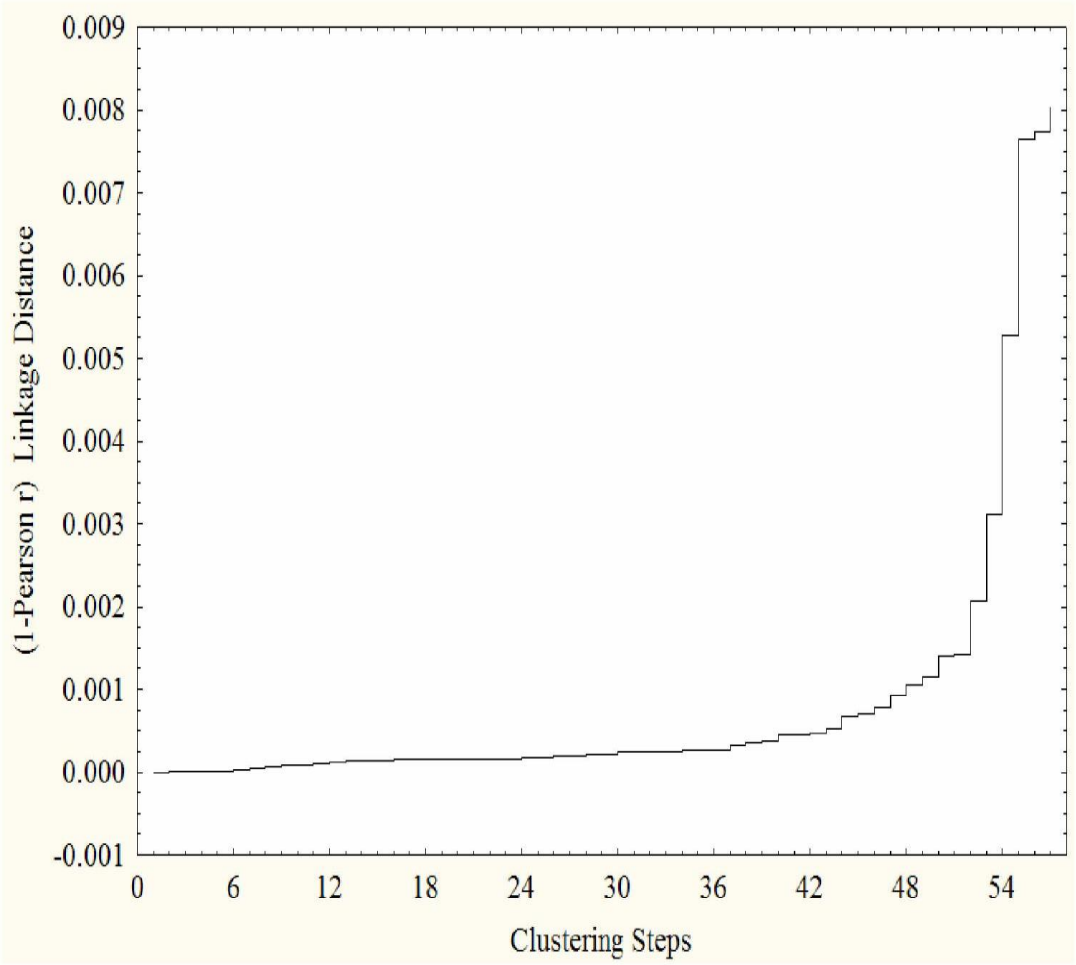
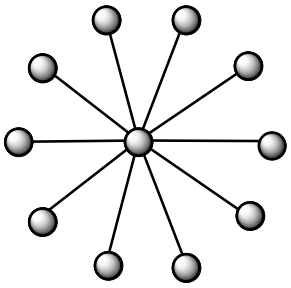
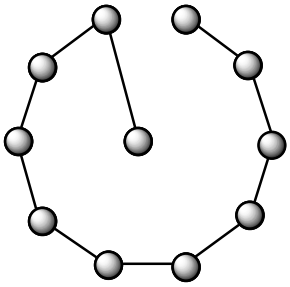
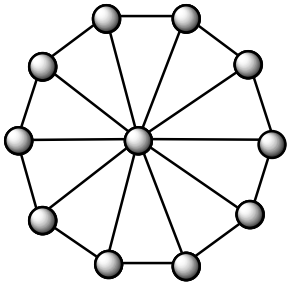
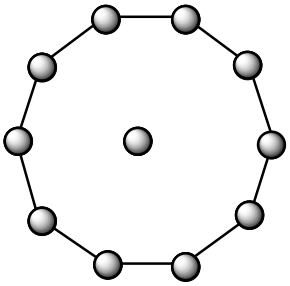
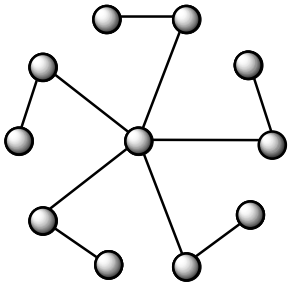


Figure 5. Antifungal drugs similar-mechanism-of-action network.

The accuracy of the model in terms of the percentages of good classification of active/non-active drugs is very important for network construction but is not the only aspect to be considered. However, the final topology of the network we pretend to construct is at less as important as model accuracy for inference of drugs multi-species activity similariy. For instance, we can find two models with the same overall accuracy but predicting networks with topological properties essentially different. In general, different methods for network reconstruction based on co-expresion not give as result the same network.[96] In **Table 1** we illustrate this fact with a hypothetic example. In this example we have a real network and four models derived to reconstruct it. The four models predict correctly the same number of drug similarities so they have the same accuracy. Nevertheless, the topologies predicted are in some cases very different each other and with respect to the real network too. For instance, the real network presents a central node (network hub),[69] which represent a drug with a possible mechanism of action similar to all other drugs.

Table 1. Comparing real network and four models

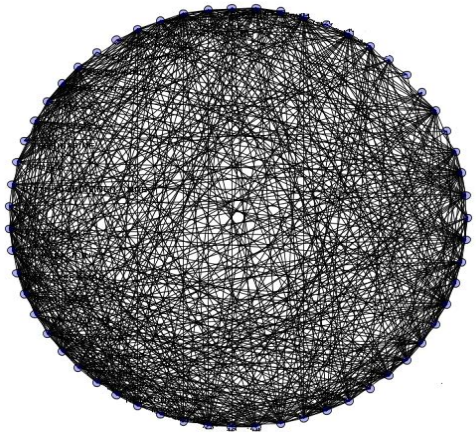
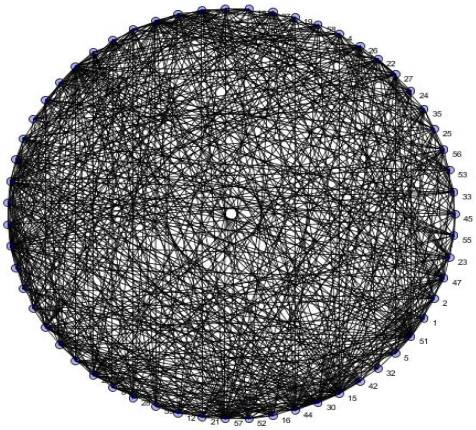
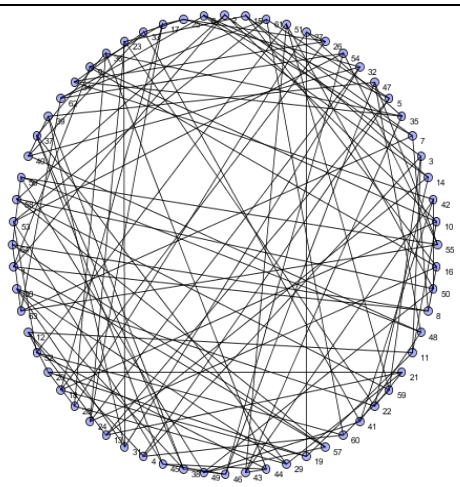
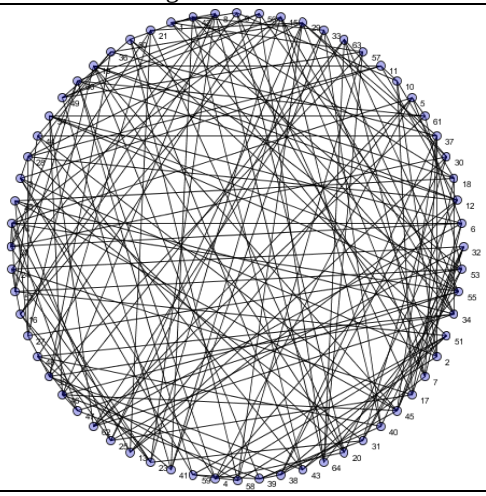
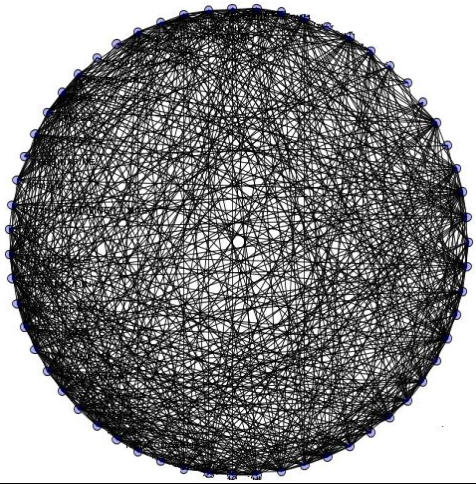
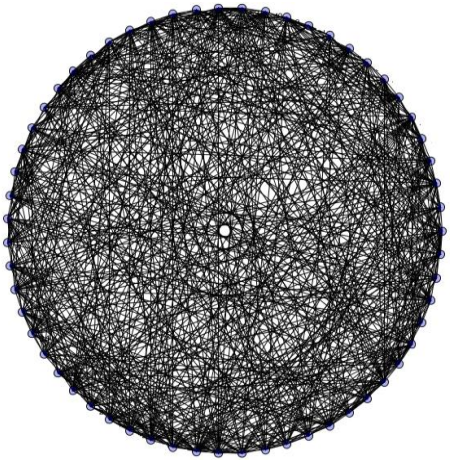
TIs[a]	Value[b]	Difference[c]		Value[b]	Difference[c]	TIs[a]
R_1	1	0	(1) 	5	4	R_3
D_1	2	0		5	3	D_3
δ_1	20	20		20	20	δ_3
W_1	100	10		125	35	W_3
LP_1	No	Yes		Yes	No	LP_3
HP_1	Yes	No		No	Yes	HP_3
(2) 				(3) 		
R_2	5	4		4	3	D_4
D_2	10	8	(4) 	2	0	R_4
δ_2	20	20		20	20	δ_4
W_2	220	130		140	50	W_4
LP_2	No	Yes		No	Yes	LP_4
HP_2	No	Yes		Yes	No	HP_4

[a] Topological indices (TIs) used to characterize the topology of the example-real network (center of the table) and networks predicted with hypothetic QSAR models: (1)-regular star with arms of length 1, (2)-linear, (3)-loop, (4)-regular star with arms of length 2; the TIs used were: network radius (R_i), network diameter (D_i), sum of vertex degrees network radius (δ_i), Wiener index (W) and loop (LP_i) or hub presence (HP_i). The values of the TIs for the example-real network are R = 1, D = 2, δ = 40, W = 90, LP = Yes and HP = Yes. [b] Value: is the value of the given TIs for the corresponding network. [c] Difference: is the difference between the value of the TIs for the predcted network and the value of the TIs for the example-real network.

Consequently, any other drug wihin the network present possibly the same mechanisn of action. The only one predicted network that reproduces this topology is the network (1) with a regular star topology having arms of length 1. By the contrary, the network (3) is a loop and predict that any drug in the network have the same mechanism as hub, which becomes in this case and isolated drug.[97] Consequently, in addition to QSAR model accuracy we should measure the topology of the network predicted and compare it with other known networks. It makes possible to derive general conclusions on the line of thinking above expressed. In the example of the **Table 1** we used different continuos and dummy measures of network topology such as the

Diameter (D, longest path), Radius (R, shortest path), the sum of node degrees (δ), the Wiener index (W), and presence of loops (PL) or hubs (HP).[92,98-100] We carried out a similar anlysis comparing the network predicted in this work with other recognized models of ideal networks. In the **Table 2** we illustrate the results of this analysis.

Table 2 Comparison with some ideal random network models

Antifungal drugs multi-species-activity QSAR network	Value	TIs	Value	Barabasi Albert Random Network
	59	n	59	
	648	m	631	
	6808	W	5582	
	6	D	2	
	5.13	R	1.42	
	22	δ	21	
	1.99	Dist	1.63	
	0.009	C	0.011	
	0.26	E	0.5	
	0.144	λ	0.123	
2D Lattice Network	**Value**	**TIs**	**Value**	**Kleinberg Small World Network**
	64	n	64	
	128	m	192	
	16384	W	10438	
	8	D	4	
	5.07	R	2.49	
	4	δ	6	
	4.06	Dist	2.58	
	0.004	C	0.006	
	0.12	E	0.25	
	0.125	λ	0.122	
Erdos Renyi Random Network	**Value**	**TIs**	**Value**	**Epsstein Power Law Network**
	59	n	59	
	640	m	648	
	5554	W	5548	
	2	D	2	
	1.43	R	1.43	
	22	δ	22	
	1.63	Dist	1.62	
	0.01	C	0.01	
	0.50	E	0.50	
	0.128	λ	0.128	

[a] The TIs used are: number of nodes (n), number of edges (m), Wiener index (W), diameter (D), and the network average values for radiality (R), node degree (δ), topological distance (Dist), node closeness (C), eccentricity (E) and node eigenvector value (λ).

We generated 5 ideal networks, one of each of the follwing classes: Barabasi-Albert random network, Kleinberg small wolrd network (SWN), 2D Lattice network, Erdos-Renyi network and Epsstein power law network (PLN).[75,101-103] The general topologic properties of these classes of networks have been studied in detail before. Consequently, if we pretend to study the features of our actual network, which characterize multi-species antifugal activity of drugs, could interesting to select between these networks the more similar to our actual network an study the deviations of the actual with respect to ideal behaviour. The networks were generated as similar a possible to actual. We measure 10 network features including: number of nodes (n), number of edges (m), Wiener index (W), diameter (D), and the network average values for radiality (R), node degree (δ), topological distance (Dist), node closeness (C), eccentricity (E) and node eigenvector value (λ). The description of these kind of parameters have been reported previously[94] and the applications for small molecules, macromolecules, and networks reviewed.[79] The deviation of the actual network with respect to ideal behaviour was measured in terms of Relative Difference Percentage (RD%) as follows RD% = ($TI_{actual} - TI_i$)·100/ TI_{actual} (see **Table 2** and **Table 3**).

Table 3. Summary of the comparative study of the actual vs. ideal networks

TIs[a]	Actual network	Barabasi-Albert	2D Lattice	Kleinberg SWN	Erdos-Renyi	Epsstein PLN
TIs values						
N	59	59	64	64	59	59
M	648	631	128	192	640	648
W	6808	5582	16384	10438	5554	5548
D	6	2	8	4	2	2
R	5.13	1.42	5.07	2.49	1.43	1.43
Δ	22	21	4	6	22	22
Dist	1.99	1.63	4.06	2.58	1.63	1.62
C	0.009	0.011	0.004	0.006	0.01	0.01
E	0.26	0.5	0.12	0.25	0.5	0.5
Λ	0.144	0.123	0.125	0.122	0.128	0.128
Relative Difference % = ($TI_{actual} - TI_i$)·100/ TI_{actual}						
N	-	**0**	-8.5	-8.5	**0**	**0**
m	-	2.6	80.2	70.4	1.2	**0**
W	-	18.0	-140.7	-53.3	18.4	18.5
D	-	66.7	-33.3	33.3	66.7	66.7
R	-	72.3	1.2	51.5	72.1	72.1
δ	-	4.5	81.8	72.7	**0**	**0**
Dist	-	18.1	-104.0	-29.6	18.1	18.6
C	-	-22.2	55.6	33.3	-11.1	-11.1
E	-	-92.3	53.8	3.8	-92.3	-92.3
λ	-	14.6	13.2	15.3	11.1	11.1
Summary statistics						
Mean		8.23	-0.07	18.89	8.42	8.36
Standard Deviation		45.83	75.34	41.61	45.18	45.22
Closer to 0		0	1.2	3.8	0	0
Max		72.3	81.8	70.4	66.7	72.1
Min		-92.3	-140.7	3.8	-92.3	-92.3
Absolute Values Mean		31.13	57.23	37.17	29.10	29.04
SD for AVM		33.15	45.12	24.34	34.34	34.43

[a] The TIs used are: number of nodes (n), number of edges (m), Wiener index (W), diameter (D), and the network average values for radiality (R), node degree (δ), topological distance (Dist), node closeness (C), eccentricity (E) and node eigenvector value (λ).

All networks present no more than 5 nodes than the actual with RD% lower than 10% in all cases. With respect to the number of edges all present the same or lower covering than the actual being any prediction in this sense non-over-estimated. The lower differences (more similar networks) for different features were: Epsstein PLN for m (RD% = 0%), Barabasi-Albert for W (18.0%), Erdos-Renyi and Epsstein PLN for δ (0%), 2D Lattice for R (1.2), Barabasi-Albert and Erdos-Renyi for Dist (18.1), Erdos-Renyi and Epsstein PLN for C (-11.1%), Kleinberg SWN for E (3.8%), and Erdos-Renyi and Epsstein PLN for (λ). Any network resembles significantly, in terms of D, to the actual (D = 6). However, we find a sort of upper and lower limits for actual network in the Kleinberg SWN and 2D Lattices which present D values of 4 and 8 having an RD% of 33.3 and -33.3% respectively. In closing, the actual network does not exactly match with any of ideal behaviours studied but, as usuall in the real world, have different properties of these ideal networks. This fact can be corroborated studinying the node degree distribution of the actual network. The **Figure 6** illustrates that the present network, even more close to normal distribution, does not significantly fit to normal, exponential, chi-square or gamma distributions. These results points to a not-over-covered normal random network, clustering different drug mechanisms of actions. The network is more far from a less useful (in this case) exponential network with few mechanisms (network hubs).[104]

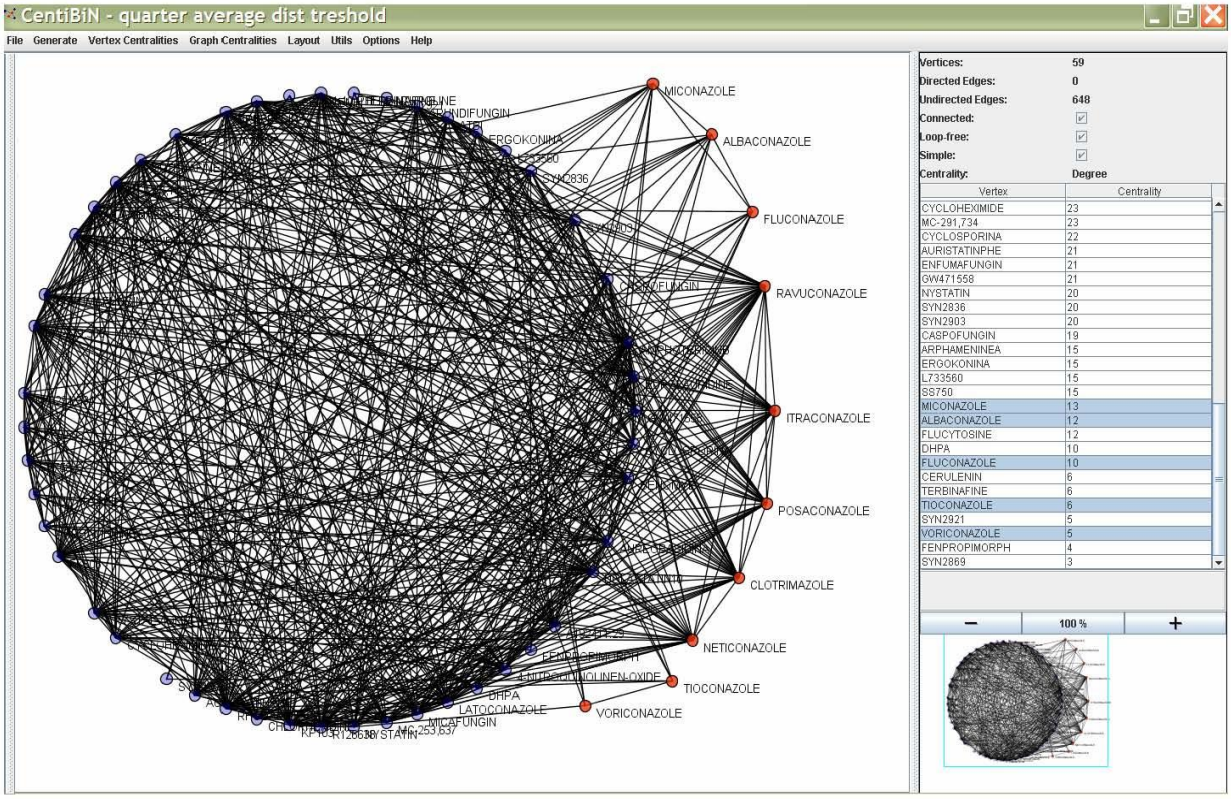


Figure 6. Distribution fitting study for antifungals similar-mechanism-of-action network, Kolmogorov-Smirnov difference d values are: d(Normal) = 0.097, d(Chis-Sqr) = 0.181 d(Gamma) = 0.201, d(Exponential) = 0.241 in all cases differences are significant with $p < 0.05$; however the lower d value is for normal distribution.

Conclusions

Using the MARCH-INSIDE approach is possible to seek a useful QSAR classifier for active/non-active drugs, which scores multi-species antifungal activity of chemicals. In analogy with gene, and transcripts co-expresion network we used the QSAR model outputs to derived a large network clustering antifungal drugs in terms of similar multi-species activity profile. Comparative studies reveal that the present network apparently has not an ideal behaviour but resemble some known network models in different aspects. In this sense, the network do not fit to some tested known distributions but is more close to normal than to exponential. These results are more characteristic of a not-over-estimated random network, clustering different drug mechanisms of actions. The present work reports by the first time the use of QSAR computational techniques in the construction of multi-species activity networks for antimicrobial drugs.

Acknowledgments

Gonzalez-Díaz H. acknowledges contract/grant sponsorship from the Program Isidro Parga Pondal of the "Dirección Xeral de Investigación y Desenvolvemento" of "Xunta de Galicia". This author also acknowledges two contracts as guest professor in the Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Spain in 2006. The authors thank the Xunta de Galicia (projects PXIB20304PR and BTF20302PR) and the Ministerio de Sanidad y Consumo (project PI061457) for partial financial support.

References

1. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH, 2002.
2. Otzen, T.; Wempe, E. G.; Kunz, B.; Bartels, R.; Lehwark-Yvetot, G.; Hansel, W.; Schaper, K. J.; Seydel, J. K. J Med Chem 2004, 47(1), 240-253.
3. Fratev, F.; Benfenati, E. J Chem Inf Model 2005, 45(3), 634-644.
4. Kubinyi, H. J Cancer Res Clin Oncol 1990, 116(6), 529-537.
5. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. Bioorg Med Chem 2005, 13(8), 2881-2899.
6. 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(4), 1005-1020.
7. Marrero-Ponce, Y.; Montero-Torres, A.; Zaldivar, C. R.; Veitia, M. I.; Perez, M. M.; Sanchez, R. N. Bioorg Med Chem 2005, 13(4), 1293-1304.
8. González-Díaz, H.; Olazabal, E.; Castanedo, N.; Sanchez, I. H.; Morales, A.; Serrano, H. S.; Gonzalez, J.; de Armas, R. R. J Mol Model (Online) 2002, 8(8), 237-245.
9. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernadez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. J Mol Model 2003, 9(6), 395-407.
10. González-Díaz, H.; Sanchez, I. H.; Uriarte, E.; Santana, L. Comput Biol Chem 2003, 27(3), 217-227.
11. González-Díaz, H.; de Armas, R. R.; Molina, R. Bull Math Biol 2003, 65(6), 991-1002.
12. González-Díaz, H.; Uriarte, E.; Ramos de Armas, R. Bioorg Med Chem 2005, 13(2), 323-331.
13. González-Díaz, H.; Molina, R.; Uriarte, E. Bioorg Med Chem Lett 2004, 14(18), 4691-4695.
14. González-Díaz, H.; de Armas, R. R.; Molina, R. Bioinformatics 2003, 19(16), 2079-2087.
15. González-Díaz, H.; Saiz-Urra, L.; Molina, R.; Gonzalez-Diaz, Y.; Sanchez-Gonzalez, A. J Comput Chem 2007, 28(6), 1042-1048.
16. Ramos de Armas, R.; González-Díaz, H.; Molina, R.; Perez Gonzalez, M.; Uriarte, E. Bioorg Med Chem 2004, 12(18), 4815-4822.
17. Ramos de Armas, R.; González-Díaz, H.; Molina, R.; Uriarte, E. Proteins 2004, 56(4), 715-723.
18. González-Díaz, H.; Bastida, I.; Castanedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; de Armas, R. R. Bull Math Biol 2004, 66(5), 1285-1311.
19. González-Díaz, H.; Marrero, Y.; Hernandez, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, E.; Castanedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Perez, M. Chem Res Toxicol 2003, 16(10), 1318-1327.
20. González-Díaz, H.; Aguero, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castanedo, N. Bioorg Med Chem Lett 2005, 15(3), 551-557.
21. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. Bioorg Med Chem 2005, 13(4), 1119-1129.
22. Cruz-Monteagudo, M.; González-Díaz, H.; Agüero-Chapin, G.; Santana, L.; Borges, F.; Domínguez, R. E.; Podda, G.; Uriarte, E. J Comput Chem 2007, doi:10.1002/jcc.20730.
23. González-Díaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. Bioorg Med Chem 2006, 14 5973–5980.
24. Prado-Prado, F.; González-Díaz, H.; Santana, L.; Uriarte, E. Bioorg Med Chem 2007, 15 897-902.
25. Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Karelson, M. Bioorg Med Chem 2005, 13(24), 6598-6608.
26. Basak, S. C.; Grunwald, G. D.; Gute, B. D.; Balasubramanian, K.; Opitz, D. J Chem Inf Comput Sci 2000, 40(4), 885-890.
27. Baskin, II; Ait, A. O.; Halberstam, N. M.; Palyulin, V. A.; Zefirov, N. S. SAR QSAR Environ Res 2002, 13(1), 35-41.
28. Benigni, R.; Giuliani, A. Mutat Res 1994, 306(2), 181-186.
29. Fernandez, M.; Caballero, J. J Mol Graph Model 2006.
30. Fernandez, M.; Caballero, J.; Tundidor-Camba, A. Bioorg Med Chem 2006, 14(12), 4137-4150.

31. Fernandez, M.; Tundidor-Camba, A.; Caballero, J. J Chem Inf Model 2005, 45(6), 1884-1895.
32. Fernandez, M.; Caballero, J.; Helguera, A. M.; Castro, E. A.; Gonzalez, M. P. Bioorg Med Chem 2005, 13(9), 3269-3277.
33. Caballero, J.; Garriga, M.; Fernandez, M. J Comput Aided Mol Des 2005, 19(11), 771-789.
34. Caballero, J.; Fernandez, M. J Mol Model (Online) 2006, 12(2), 168-181.
35. Caballero, J.; Fernandez, M. J Mol Model (Online) 2005, 1-14.
36. Caballero, J.; Fernandez, L.; Abreu, J. I.; Fernandez, M. J Chem Inf Model 2006, 46(3), 1255-1268.
37. Sanchez, R.; Grau, R. Bulletin of mathematical biology 2005, 67(5), 1017-1029.
38. Sanchez, R.; Grau, R. Acta biotheoretica 2006, 54(1), 27-42.
39. Sanchez, R.; Grau, R.; Morgado, E. Mathematical biosciences 2006, 202(1), 156-174.
40. Sanchez, R.; Morgado, E.; Grau, R. Bulletin of mathematical biology 2005, 67(1), 1-14.
41. Sanchez, R.; Morgado, E.; Grau, R. Journal of mathematical biology 2005, 51(4), 431-457.
42. Bashford, J. D.; Jarvis, P. D. Bio Systems 2000, 57(3), 147-161.
43. Beland, P.; Allen, T. F. Journal of theoretical biology 1994, 170(4), 359-365.
44. Zhang, Z.; Grigorov, M. G. Proteins 2006, 62(2), 470-478.
45. Voy, B. H.; Scharff, J. A.; Perkins, A. D.; Saxton, A. M.; Borate, B.; Chesler, E. J.; Branstetter, L. K.; Langston, M. A. PLoS Comput Biol 2006, 2(7), e89.
46. Tanaka, T.; Ikeo, K.; Gojobori, T. Gene 2006, 365, 88-94.
47. Sun, S.; Zhao, Y.; Jiao, Y.; Yin, Y.; Cai, L.; Zhang, Y.; Lu, H.; Chen, R.; Bu, D. FEBS Lett 2006, 580(7), 1891-1896.
48. Gelfand, M. S. Curr Opin Struct Biol 2006, 16(3), 420-429.
49. Barabasi, A. L. Science 2005, 308(5722), 639-641.
50. Barabasi, A. L.; Freeh, V. W.; Jeong, H.; Brockman, J. B. Nature 2001, 412(6850), 894-897.
51. Barabasi, A. L.; Oltvai, Z. N. Nature reviews 2004, 5(2), 101-113.
52. de Menezes, M. A.; Barabasi, A. L. Physical review letters 2004, 92(2), 028701.
53. Dezso, Z.; Barabasi, A. L. Physical review 2002, 65(5 Pt 2), 055103.
54. Dezso, Z.; Oltvai, Z. N.; Barabasi, A. L. Genome research 2003, 13(11), 2450-2454.
55. Dobrin, R.; Beg, Q. K.; Barabasi, A. L.; Oltvai, Z. N. BMC bioinformatics 2004, 5, 10.
56. Jeong, H.; Mason, S. P.; Barabasi, A. L.; Oltvai, Z. N. Nature 2001, 411(6833), 41-42.
57. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabasi, A. L. Nature 2000, 407(6804), 651-654.
58. Oliveira, J. G.; Barabasi, A. L. Nature 2005, 437(7063), 1251.
59. Yu, X.; Lin, J.; Masuda, T.; Esumi, N.; Zack, D. J.; Qian, J. Nucleic Acids Res 2006, 34(3), 917-927.
60. Carter, S. L.; Brechbuhler, C. M.; Griffin, M.; Bond, A. T. Bioinformatics 2004, 20(14), 2242-2250.
61. Carlson, M. R.; Zhang, B.; Fang, Z.; Mischel, P. S.; Horvath, S.; Nelson, S. F. BMC Genomics 2006, 7, 40.
62. Zhang, B.; Horvath, S. Statistical applications in genetics and molecular biology 2005, 4, Article17.
63. Reverter, A.; Barris, W.; McWilliam, S.; Byrne, K. A.; Wang, Y. H.; Tan, S. H.; Hudson, N.; Dalrymple, B. P. Bioinformatics 2005, 21(7), 1112-1120.
64. Estrada, E. Proteomics 2006, 6(1), 35-40.
65. Estrada, E.; Rodriguez-Velazquez, J. A. Phys Rev E Stat Nonlin Soft Matter Phys 2005, 71(5 Pt 2), 056103.
66. Jeong, H.; Mason, S. P.; Barabási, A.-L.; Oltvai, Z. N. Nature 2001, 411, 41–42.
67. Yu, X.; Lin, J.; Zack, D. J.; Qian, J. Nucleic Acids Res 2006, 34(17), 4925-4936.
68. Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R.; Califano, A. BMC Bioinformatics 2006, 7 Suppl 1, S7.
69. Jonsson, P. F.; Bates, P. A. Bioinformatics 2006.
70. Van Waterbeemd, H. Chemometric methods in molecular design; Wiley-VCH: New York, 1995.
71. StatSoft.Inc., 2002, p STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft.

72. González-Díaz, H.; Molina-Ruiz, R.; Hernandez, I., 2005, p MARCH-INSIDE version 2.0 (Markovian Chemicals In Silico Design). Main author information requesting contact email: gonzalezdiazh@yahoo.es.
73. Microsoft.Corp., 2002, p Microsoft Excel
74. Koschützki, D., 2006, p CentiBiN Version 1.4.2, Centralities in Biological Networks © 2004-2006 Dirk Koschützki Research Group Network Analysis, IPK Gatersleben, Germany.
75. Junker, B. H.; Koschutzki, D.; Schreiber, F. BMC bioinformatics 2006, 7, 219.
76. Cambridge.Soft., 2005, p ChemOffice, version 9.0, a CambridgeSoft Software Development Kit (SDK), integrating ChemDraw, ChemFinder, and Chem3D.
77. Estrada, E.; Uriarte, E. Curr Med Chem 2001, 8, 1573-1588.
78. Devillers, J.; Balaban, A. T., Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach The Netherlands, 1999.
79. González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Current Topics in Medicinal Chemistry 2007, 7, doi:1568-0266/1507.
80. Gozalbes, R.; Galvez, J.; Garcia-Domenech, R.; Derouin, F. SAR QSAR Environ Res 1999, 10(1), 47-60.
81. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Pereira, D. M.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Aran, V. J.; Martinez-Fernandez, A. R.; Garcia Sanchez, R. N. Bioorg Med Chem Lett 2005, 15(17), 3838-3843.
82. Murcia-Soler, M.; Perez-Gimenez, F.; Garcia-March, F. J.; Salabert-Salvador, M. T.; Diaz-Villanueva, W.; Medina-Casamayor, P. J Mol Graph Model 2003, 21(5), 375-390.
83. Pasqualoto, K. F.; Ferreira, E. I.; Santos-Filho, O. A.; Hopfinger, A. J. J Med Chem 2004, 47(15), 3755-3764.
84. Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. J Chem Inf Model 2005, 45(4), 1082-1100.
85. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. J Comput Aided Mol Des 2004, 18(10), 615-634.
86. Garcia-Garcia, A.; Galvez, J.; de Julian-Ortiz, J. V.; Garcia-Domenech, R.; Munoz, C.; Guna, R.; Borras, R. J Antimicrob Chemother 2004, 53(1), 65-73.
87. Cronin, M. T.; Aptula, A. O.; Dearden, J. C.; Duffy, J. C.; Netzeva, T. I.; Patel, H.; Rowe, P. H.; Schultz, T. W.; Worth, A. P.; Voutzoulidis, K.; Schuurmann, G. J Chem Inf Comput Sci 2002, 42(4), 869-878.
88. Graham, R. L.; Groetschel, M.; Lovász, L. Handbook of Combinatorics; Elsevier (North-Holland) and MIT Press: Amsterdam and Cambridge, 1996.
89. Gupta, A.; Maranas, C. D.; Albert, R. Bioinformatics 2006, 22(2), 209-214.
90. Saiz-Urra, L.; Gonzalez, M. P.; Fall, Y.; Gomez, G. Eur J Med Chem 2007, 42(1), 64-70.
91. Podani, J.; Oltvai, Z. N.; Jeong, H. G.; Tombor, B.; Barabási, A.-L.; Szathmáry, E. Nat Genet 2001, 29, 54-56.
92. Tsaparas, P.; Marino-Ramirez, L.; Bodenreider, O.; Koonin, E. V.; Jordan, I. K. BMC Evol Biol 2006, 6, 70.
93. Brun, C.; Herrmann, C.; Guenoche, A. BMC Bioinformatics 2004, 5, 95.
94. Junker, B. H.; Koschuetzki, D.; Schreiber, F. BMC Bioinformatics 2006, 7(1), 219.
95. Di Santo, R.; Tafi, A.; Costi, R.; Botta, M.; Artico, M.; Corelli, F.; Forte, M.; Caporuscio, F.; Angiolella, L.; Palamara, A. T. J Med Chem 2005, 48(16), 5140-5153.
96. Zhu, D.; Hero, A. O.; Cheng, H.; Khanna, R.; Swaroop, A. Bioinformatics 2005, 21(21), 4014-4020.
97. Kamp, C.; Christensen, K. Phys Rev E Stat Nonlin Soft Matter Phys 2005, 71(4 Pt 1), 041911.
98. Spirin, V.; Mirny, L. A. Proc Natl Acad Sci U S A 2003, 100(21), 12123-12128.
99. Koschützki D.; Lehmann, K. A.; Peeters, L.; Richter, S.; Tenfelde-Podehl, D.; Zlotowski, O. In Network Analysis: Methodological Foundations, LNCS Tutorial Brandes, U.; Erlebach, T., Eds.; Springer, 2005, p 16-61.

100. Valente, T. W.; Foreman, R. K. Social Networks 1998, 1, 89-105.
101. Barabasi, A. L.; Albert, R. Science 1999, 286(5439), 509-512.
102. Kleinberg, J. M. Nature 2000, 406(6798), 845.
103. Kleinberg, J.; Lawrence, S. Science 2001, 294(5548), 1849-1850.
104. Barabasi, A. L.; Bonabeau, E. Scientific American 2003, 288(5), 60-69.