12th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-12) 1-30 November 2008
http://www.usc.es/congresos/ecsoc/12/ECSOC12.htm & http://www.mdpi.org/ecsoc-12

[G003]

TOMOCOMD-CARDD Method in Early Drug Discovery-based Rational Drug Selection of Antifungal Agents. *A Comparative Study*

Ricardo Medina Marrero,[1,2] Yovani Marrero-Ponce,[1,3,4*] Yunaimy Echeverría Díaz,[1] Gerardo M. Casañola-Martín,[1] Milagros García Bernal,[2] Francisco Torrens,[3] Facundo Pérez Jimenez.[4]

[1]Unit of Computer-Aided Molecular "*Biosilico*" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy. Central University of Las Villas (UCLV), Santa Clara, 54830, Villa Clara, Cuba.

[2]Department of Microbiology, Chemical Bioactive Center. Central University of Las Villas (UCLV), 54830, Villa Clara, Cuba.

[3]Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, Poligon la Coma s/n (detrás de Canal Nou) P. O. Box 22085, E-46071 Valencia, Spain.

[4]Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain.

[*]*Corresponding author.*

Fax: 53-42-281130 [or 53-42-281455] (Cuba) and 963543156 (València)

Phone: 53-42-281192 [or 53-42-281473] (Cuba) and 963543156 (València)

Cell: 610028990

***e*-mail**: ymarrero77@yahoo.es; ymponce@gmail.com or yovanimp@qf.uclv.edu.cu

URL: http://www.uv.es/yoma/

ABSTRACT

The novel ***TOMOCOMD–CARDD*** approach has been introduced here for the classification and design of antifungal agents using computer-aided molecular design. For this purpose, no stochastic and stochastic atom-based quadratic fingerprinting were used to codify the antifungal-related chemical structure information from a comprehensive data set of 2478 organic compounds having a great structural variability, 1087 of them being antifungal agents covering the broadest antifungal mechanisms of action known so far. The two ligand-based antifungal-activity classification models obtained by using Linear Discriminant Analysis, including no stochastic and stochastic indices, classified correctly 90.73% and 92.47%, respectively, of 1772 chemicals in the training set. These models showed moderate-to-high Matthews correlation coefficients (MCC of 0.81 and 0.85) as well as a very good accuracy, sensitivity, specificity and false alarm rate. These models were able of classifying correctly 92.16% and 87.56% of 706 compounds in an external test set. In general, the ***TOMOCOMD–CARDD*** models were best in predicting antifungal activity when compared with six of the most recent models reported so far; indicating that this approach could be very useful to identify (design and/or select) new antifungal agents against life-threatening fungal infections.

Keywords: ***TOMOCOMD-CARDD*** *Software; non-stochastic and stochastic atom-based quadratic indices; LDA-based QSAR model; Learning Machine Tools, Computational Screening, Antifungal Agent.*

Running head: *Rational Drug Selection of Antifungal compounds…*

1. INTRODUCTION

Over the past two decades the incidence of life-threatening fungal infections have increased and is directly related to the increasing patient populations at risk for the development of serious fungal infections, which includes those having major surgery, HIV infection, chemotherapy-induced neutropenia, solid-organ and hematopoietic stem cell transplantation, hemodialysis, advanced age, premature birth, and from the use of broad-spectrum antibiotics and glucocorticosteroids.[1-5]

Serious infections are produced not only by the well known opportunists *Candida albicans*, *Cryptococcus neoformans*, and *Aspergillus fumigatus* but also by new and emerging fungal pathogens including species of *Candida* and *Aspergillus* other than *C. albicans* and *A. fumigatus*; opportunistic yeast like fungi such as *Trichosporon* spp., *Rhodotorula* spp., and *Blastoschizomyces capitatus*; hyaline molds such as *Fusarium*, *Acremonium*, *Scedosporium*, *Paecilomyces*, and *Trichoderma* species; zygomycetes such as *Rhizopus* spp*., Absidia* spp*.,* and *Rhizomucor* spp.; and a wide variety of dematiaceous fungi.[3, 6, 7]

The intrinsic resistance to even the very newest antifungal agents observed in some of these genera, along with the development of resistance during treatment in others, is becoming a major problem in the management of these diseases.[2, 8, 9] Furthermore, the clinical utility of the few classes of antifungal drugs on the market is limited by several shortcomings such as the lack of broad spectrum and fungicidal activity, unfavorable routes of administration, severe side effects and, undesirable drug-drug interactions.[10-12] To revert this situation, new effective antifungal agents need to be discovered in the next few years to come.

Computer-aided drug design has emerged in the pharmaceutical world as an important tool for the "*rational*" search of chemicals with desired properties. Different studies related to the *in silico* design have been reported in the literature during the last years.[13-19] In fact, many large pharmaceutical companies have reoriented their research strategies seeking to solve the problem of generation/selection of novel chemical entities (NCEs), one of the major bottlenecks in the drug discovery process. Currently most integration projects include efforts to integrate the data associated with NCE generation.[20] Alternatively, several approaches to the computer-aided molecular design and high-throughput *in silico* screening (or virtual high-throughput screening) have been introduced in the literature.[21] "*Nevertheless, novel computational methods and strategies are required to deliver a system that significantly reduces the time-to-market and research and development (R&D) spending, and increase the rate at which NCEs progress through the pipeline. Such studies if they are implemented successfully can deliver substantial benefits and act as the bedrock for NCE selection".*[20]

At present, there is an increasing interest on the development of *rational* approaches for antifungal drug discovery. In this sense, a very important role may be played by computer-aided drug design techniques based on quantitative-structure–activity-relationship (QSAR) studies. Unfortunately, almost all antifungal QSAR studies reported so far are based on very limited databases considering only structurally related compounds with *specific* action modes or acting against *a single* fungus species.[22-26] Therefore, most of the previous QSAR studies can be considered as ***local*** models having a small to medium spectrum of *chemical space* with limited power to predict the activity/inactivity of different ligands to specific molecular targets. For instance,

Gollapudy et al. [22] developed a 3D-model of the structure of *Aspergillus fumigatus* lanosterol 14-α demethylase (AF-CYP51A) using the crystal structure of *Mycobacterium tuberculosis* 14-α demethylase (PDB code:1EA1) as a template to investigate the interactions of azole antifungal with the enzyme(s) from fungi. Later, Gokhale and Kulkarni [23] performed a QSAR study on a series of 92 molecules using different physicochemical descriptors. In this report, inhibitors were divided into five classes depending on chemical structure and QSAR models were generated correlating the antifungal activity against *Candida albicans* by using the genetic function approximation (GFA) technique. Afterwards, the same research group examined a series of benzofuran antifungal in order to determine the structural requirements of N-myristoyltransferase (Nmt) enzyme inhibition by 3D-QSAR using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods as well as through docking studies where active and inactive molecules were docked into the active site of the *Candida albicans* Nmt (CaNmt) crystal structure to analyze enzyme–inhibitor interactions. The final results obtained from 3D-QSAR and docking studies were found complementary.

However, in the last few years, some *in silico* methods have been used to develop ***Global*** QSAR equations.[27-30] Firstly, two different ligand-based classification models on antifungal activity were developed by a group in Valencia University, which gave rise to a good discrimination of this activity.[28, 29] These models were used in *virtual screening* and four chemicals were *in silico* selected to be tested in the laboratory against *C. albicans*, *C. glabrata* and *S. cerevisiae*. All compounds, namely anethole, 2–methyl–4,5–diphenyloxazole, 2–mercaptobenzoxazole and β–naphtyl caproate, showed activity

against the three species with MIC_{50} ranging from 25 to 100 μg/mL. The main drawbacks of these equations are the *trim down* application-domain taken into account and the rather small diversity of chemical structural patterns contained in the data. Later, employed the same database, Li *et al*[30] obtained three new classification models by using a set of molecular descriptors (including electronic descriptors, topological descriptors, geometric descriptors and molecular shape Indices) and support vector machine (SVM) classification method. Comparison of the results by SVM method and those by other statistical classification methods, for instance k-nearest neighbor (k-NN) and C4.5 decision tree (that use the same pre-selected molecular descriptors) was also conducted. This study indicates the potential of SVM in facilitating the prediction of antifungal activity but the low spectrum of applicability of this model is the major advantage because it born form the same database that the previous counterpart analysis.[28, 29] More recently, one of our research groups proposed a global *unified* QSAR approach to predict the antifungal activity against different species of fungi.[27] The obtained model was trained using a data set composed by only 74 drugs tested in the literature against some of the fungi selected from a list of 87 species. The result was quite good representing the first unified model that allows predicting antifungal activity of organic compounds against a very *large diversity* of fungal pathogens; although the diversity of chemical structural patterns was low-to-medium considering the actual *chemical space*.

As a result, researchers interested on predicting the antifungal activity for a given series of compounds need to use/develop as many QSAR equations as combinations of structurally heterogeneous families of compounds are necessary to be predicted. Therefore, the development of a single equation explaining the antifungal activity of

structurally heterogeneous series of compounds covering as many as possible broad-range of mechanism of action is of major interest.

In this context, our research group has recently introduced a novel scheme known as ***TOMOCOMD*** (acronym of ***To***pological ***MO***lecular ***COM***puter ***D***esign) able of generating 2D (topologic), 2.5 (3D-chiral) and 3D (topographic and geometric) molecular descriptors based on the application of the discrete mathematics and linear algebra theory to chemistry. In this sense, atomic, group and atom-type as well as total linear, quadratic and bilinear molecular indices have been defined in analogy to the linear, quadratic and bilinear mathematical maps.[31-33] In first instance, this *in silico* method was successfully applied to the prediction of several physical, physicochemical and chemical properties of organic compounds[31, 32, 34, 35] and subsequently to perform rational –*in silico* molecular design (or selection/identification of lead drug-like chemicals) and QSA(P)R [Quantitative Structure-Activity(Property) Relationship] studies in its ***–CARDD*** extension (***TOMOCOMD-C***omputer ***A***ided " ***R***ational" ***D***rug ***D***esign) and to generate macromolecular fingerprintings in its ***–CANAR*** (***TOMOCOMD- C***omputer-***A***ided ***N***ucleic ***A***cid ***R***esearch)[36, 37] and ***–CAMPS*** extensions (***TOMOCOMD- C***omputer-***A***ided ***M***odeling in ***P***rotein ***S***cience),[38, 39] respectively.

The ***CARDD*** extension of ***TOMOCOMD*** approach has been successful used to estimate the intestinal–epithelial transport of drugs,[40, 41] to identify new tyrosinase inhibitors,[42] but it has been mainly validated and proved useful in the virtual screening of novel tyrosinase inhibitors as well as anthelmintic, trypanosomicidal and trichomonacidal compounds, which were then synthesized and *in vitro* evaluated on mushroom tyrosinase enzyme, *Fasciola hepatica, Trypanosoma cruzi* and *Trichomonas vaginalis,*[35, 43-47], as

well as in the fast-track discovery of novel paramphistomicides, antibacterial and antimalarial compounds.[15, 48-50] The predictive capacity of this ligand-based virtual screening methodology has remained high in all studies performed so far, which is an indication that ***TOMOCOMD-CARDD*** descriptors could become a powerful *in-silico* tool for the discovery of new drug or lead compounds[15, 35, 43-45, 48-50]

The main objectives of this paper are, first, to gather a large and structurally-diverse antifungal data base for modeling the so far broadest mechanisms of antifungal action, and second, to develop highly predictive lineal classification models using the *TOMOCOMD-CARDD* approach and linear discriminant analysis. Finally, the results of the current study are compared with those obtained in previous works showing the robustness of antifungal models developed herein.

2. MATERIALS AND METHODS

2.1. Computational Methods. The *CARDD* module of the *TOMOCOMD* approach was used to draw all structures and to generate molecular descriptors.[51] Briefly, the molecular pseudograph of each molecule was represented by the drawing mode of the *CARDD* module followed by the computation of the total and local (atom and atomtype), nonstochastic and stochastic quadratic indices of the k^{th} "nonstochastic and stochastic graph–theoretical electronic-density matrices" M^k and S^k, correspondingly, using the calculation mode[14, 31, 35, 40, 52] The k^{th} atomtype quadratic indices were calculated by adding the k^{th} atomic quadratic indices for all atoms of the same type in the molecule. In the atomtype quadratic indices formalism, each atom in the molecule is classified into an atomtype (fragment), such as heteroatoms, halogen atoms, aliphatic carbon chain,

aromatic atoms (aromatic rings), and so on. The mathematical basis and methodological explanation of this approach have been reported elsewhere[14, 31, 35, 40, 52] In this study, specifically we used the *k*th (k = 15) atomtype (heteroatoms: S, N, O) quadratic fingerprints not considering and considering H-atoms in the molecular pseudograph, correspondingly [$q_{kL}(\bar{x}_E)$ and $q_{kL}^{H}(\bar{x}_E)$].

In this report, Pauling electronegativities[53] were used as atomic weights (molecular vector's components). Finally, linear discriminant analysis (LDA) was performed to find quantitative relationship between the antifungal activity and the ***TOMOCOMD-CARDD***'s generated quadratic fingerprintings.

2.2. Data Set. Though antifungal compounds exhibit an enormous structural diversity and action modes, only a small proportion of that diversity has been seriously explored for its pharmacological potential so far, and there is therefore little reason to believe that this potential has now run dry. For this reason, a large database to facilitate the application of cheminformatics and molecular modelling to antifungal activity prediction has been constructed. Consequently, a data set of 2142 organic chemicals having a great structural variability - 1087 of them being antifungal agents[10, 23, 28, 54-112] covering the broadest antifungal mechanisms of actions known so far and the rest inactive ones (1055 compounds having other clinical uses, such as antivirals, sedative/hypnotics, diuretics, anticonvulsivants, hemostatics, oral hypoglycemics, antihypertensives, antihelminthics, anticancer compounds, and so on) - was selected.[113]

The data set of antifungal agents (active compounds) was chosen considering the largest representation of the so far known action modes; i.e., compounds interfering with cell wall synthesis (chitin synthesis inhibitors such as polyoxins and nikkomycins and, ß-

1,3 glucan synthesis inhibitors such as echinocandins), agents interfering with membrane sterols (polyenes, azoles, allylamines and morpholines), protein (sordarins) and DNA synthesis inhibitors (flucytosine and pentamidine analogs) as well as inhibitors of N-myristoyltransferase.[114] Several compounds reported as antifungals but having no known mechanism of action were also included.

2.3 Chemometric method. Linear discriminant analysis (LDA) was performed as implemented in the STATISTICA 6.0 for Windows package, using the forward stepwise procedure as a strategy for variable selection.[115] In this way, quantitative models with the following form were obtained:

$$\boldsymbol{P} = a_0\boldsymbol{q}_0(\bar{x}) + a_1\boldsymbol{q}_1(\bar{x}) + \ldots + a_n q_n(\bar{x}) + a_{n+1}\boldsymbol{q}_{0L}(\bar{x}) + a_{n+2}q_{1L}(\bar{x}) + \ldots + a_m q_{mL}(\bar{x}) \quad (\mathbf{1})$$

where $\boldsymbol{P}$ is the biological property (in this study $\boldsymbol{P}$ was designated as ***AFA***, acronym of **A**nti-**F**ugal **A**ctivity), $\boldsymbol{q_n}(\bar{x})$, the n^{th} total quadratic index, $\boldsymbol{q_{mL}}(\bar{x})$, the m^{th} local quadratic index and $\boldsymbol{a_n's}$ and $\boldsymbol{a_m's}$, the coefficients obtained by LDA (here, $k^{th} = n^{th}$ or m^{th}). The principle of parsimony (Occam's razor) was taken into consideration as a strategy for model selection. Accordingly, we selected models having the highest statistical significance but as few parameters as possible.

The quality of the models was determined by examining Wilks' λ parameter (U statistic), which can take values ranging from 0 (perfect discrimination) to 1 (no discrimination), the square Mahalanobis distance (D^2), which indicates the separation between active and inactive groups, the Fisher ratio (F) and its corresponding p level [$p(F)$]. Finally, the calculation of percentages of global good classification (accuracy), sensibility, specificity (also known as ''hit rate''), false positive rate (also known as

‘‘false alarm rate’’), and Matthews correlation coefficient (***C***) in the training and test sets were also used to assess the models.[116]

The Randić methodology for orthogonalization of descriptors was followed to avoid the exclusion of descriptors on the basis of its colinearity with other variables included in the model.[117-121] As a first step, an appropriate order of orthogonalization was considered following the order with which the variables were selected in the forward stepwise search procedure of the statistical analysis. The first variable (V_1) is taken as the first orthogonal descriptor $^1O(V_1)$, and the second one (V_2) is orthogonalized with respect to it [$^2O(V_2)$]. The residual of its correlation with $^1O(V_1)$ is that part of the descriptors V_2 not reproduced by $^1O(V_1)$. Similarly, from the regression of V_3 versus $^1O(V_1)$, the residual is the part of V_3 that is not reproduced by $^1O(V_1)$, and it is labeled $^1O(V_3)$. The orthogonal descriptor $^3O(V_3)$ is obtained by repeating this process in order to also make it orthogonal to $^2O(V_2)$. The process is continually repeated until all variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model.

3. RESULTS AND DISCUSSION

3.1 Development of lineal discriminant functions. When compared with other areas of pharmaceutical research, the screening of organic compounds in order to *in silico* identify new antifungal leads has suffered from a lack of data in an appropriate format. Particularly, the electronic information on chemical structure is found to be insufficient. While such information can serve a wide variety of purposes, it is perhaps in the field of virtual screening where it may have its greatest impact. For this reason, the first step of our study was to use a dataset having a molecular diversity as wide as possible in order to

search for good LDA-based QSAR models. To fulfill this requirement we have selected a data set of 2142 compounds, 1087 of them being antifungal agents covering all known as well as some unknown antifungal mechanisms of actions and, the rest 1055 chemicals having a series of other pharmacological uses. These compounds were randomly split into a training set containing 717 antifungal and 119 inactive compounds and a test set including 370 antifungal and 336 inactive compounds, respectively.

TOMOCOMD-generated data was used to derive discriminant functions able of classifying compounds as antifungal-like (positive) or no antifungal-like (negative) through LDA, using non-stochastic and stochastic atomtype quadratic indices as independent variables[14, 31, 35, 40, 52] For this purpose, the forward stepwise procedure of the statistic package STATISTICA[115] was fixed as a strategy for variable selection. The best discrimination functions obtained with nonstochastic and stochastic quadratic indices for the training set are given below, respectively:

$$\boldsymbol{AFA} = -5.01 + 5.94x10^{-4}\boldsymbol{q}_5(\bar{x}) - 1.24x10^{-4}\boldsymbol{q}_6(\bar{x}) - 0.02\boldsymbol{q}_{1L}{}^{H}(\bar{x}_E) - 1.59x10^{-7}\boldsymbol{q}_{12L}(\bar{x}_{E\text{-}H}) \quad (\mathbf{2})$$

N(Training) = 1436 $\lambda = 0.41$ $D^2 = 5.69$ $F(4,1431) = 509.72$ Rcan = 0.766 $\chi^2 = 1268.38$ $p<0.0001$

$$\boldsymbol{AFA} = -4.44 + 0.20^{s}\boldsymbol{q}_{11}(\bar{x}) + 0.50^{s}\boldsymbol{q}_{8L}{}^{H}(\bar{x}_E) - 0.08^{s}\boldsymbol{q}_0{}^{H}(\bar{x}) - 0.62^{s}\boldsymbol{q}_{6L}{}^{H}(\bar{x}_E) \quad (\mathbf{3})$$

N(Training) = 1436 $\lambda = 0.39$ $D^2 = 6.07$ $F(4,1431) = 544.51$ Rcan = 0.777 $\chi^2 = 1324.69$ $p<0.0001$

where, ***AFA*** refers to Antifungal Activity, N is the number of compounds, λ is Wilk's lambda, D^2 is the squared Mahalanobis distance, *F* is the Fisher ratio, *p*-value is the

significance level and Rcan and χ^2 are the correlation coefficient and chi-squared parameter of canonical LDA analysis, respectively.

While Eq. **2** classified correctly 90.73 % of the compounds in the training set, misclassifying only 164 chemicals out of a total of 1772, Eq. **3** classified correctly 92.47% of compounds, misclassifying only 133 chemicals. As it can be appreciated from Table 1, stochastic quadratic indices were best in predicting the antifungal activity than nonstochastic quadratic indices in the training set not only because of their better accuracy and Mathew's correlation coefficient but also due to their higher sensitivity, specificity and lower false positive rate. In general terms, however, both models were good to describe the antifungal activity of chemical compounds. The classification from Eqs. **2** and **3** of all active and inactive training compounds appears in Table SD1 and SD2, respectively, as Supplementary Data.

Table 1 comes about here (see end the document)

Results obtained from the training set provide some clues on the power of the developed models. However, their real power and final aim resides in the ability of predicting the biological properties of new compounds. Therefore, the use of a test set is essential to assess such a predictive power.[122, 123] For this purpose, a study aimed to test the predictive capacity of the two obtained discriminant functions was carried out with an external test set. In this case, (Eq**. 2**) correctly classified 92.16% (274/282) of the active compounds and 91.96% (141/160) of the inactives, whereas (Eq. **3**) correctly classified 87.56% (270/282) of the actives and 91.96% (143/160) of the inactive ones, for an overall accuracy of 92.06% (27/442) and 89.66% (29/442), respectively (for more details, see also Table 1). Contrary to what was observed in the training set, nonstochastic quadratic

indices were slightly superior in predicting the antifungal activity in the test set. Nonetheless, the predictive power of both models was really good with a relatively low number of misclassified compounds. This is a highly desirable property because the lower the number of misclassified inactive compounds, the less the waste of time and resources by sending inactive chemicals to biological tests;[18] similarly, the lower the number of misclassified active compounds, the less the chance of losing a potential drug candidate. The classification from Eqs. **2** and **3** of all active and inactive test compounds appears in Table SD3 and SD4, respectively, as Supplementary Data.

3.2 Orthogonalization of Descriptors. In the orthogonalization process, molecular descriptors are transformed in such a way that they do not mutually correlate to each other. In this philosophy, developed by Randić several years ago, the exclusion of descriptors based on their colinearity with other variables previously included in the model is avoided as a way to improve the statistical interpretation of the models by using interrelated indices.[117-121] Both, the nonorthogonal descriptors and the derived orthogonal descriptors contain the same information. Therefore, the same statistical parameters of the QSAR models are obtained.[117-121] It is known that the interrelation among different descriptors can result in highly unstable regression coefficients, which makes it almost impossible to know the relative importance of an index included in a model. In other cases, however, strongly interrelated descriptors can enhance the quality of a model because the small fraction of a descriptor that is not reproduced by its strongly interrelated pair can provide positive contributions to the model. Furthermore, the coefficient of the QSAR model based on orthogonal descriptors is stable to the inclusion

of novel descriptors, facilitating the interpretation of the regression coefficients and the evaluation of the role of individual fingerprints in the QSAR model.

The results of the orthogonalization of molecular descriptors included in both models are shown in Table 2. Eq.s **2a** and **3a** represents the final models with the orthogonalized molecular indices whereas in the symbolization ${}^{m}\mathrm{O}[\boldsymbol{q}_k(\bar{x})]$, the superscript *m* expresses the order of importance of the variable $[\boldsymbol{q}_k(\bar{x})]$ after a preliminary forward stepwise analysis and O means orthogonal (see Table 2). As it can be appreciated, there is a total coincidence in all statistical parameters between orthogonal descriptor-based models and linear descriptors-based models (i.e., the statistical coefficients of LDA-QSARs λ, D^2, F, $\boldsymbol{C}$, accuracy (Q_{total}) are the same whether a set of non-orthogonal descriptors or the corresponding set of orthogonal indices is used).

Table 2 comes about here (see end the document)

This fact facilitates the interpretation of the coefficients in the LDA-QSAR equations. In this sense, ${}^{m}\mathrm{O}(\boldsymbol{q}_k(\bar{x}))$ may be classified according to the distance *k* into short- (0-5), mid- (6-10), and long-range non-stochastic and stochastic quadratic indices. The information given in Table 2 clearly shows that all three short- middle- and long-range total and atom-type (heteroatoms and H-atoms bonding to heteroatoms) quadratic indices had a contribution to the antifungal activity, and local quadratic indices had the best variable combinations capable of describing such activity of compounds included in the training and test set. Nevertheless, total variables such as those of zero order were included in the models, indicating that the size and atom composition in a molecule are important for its activity. The high contribution of local variables can be explained by the fact that the mechanisms of action of antifungal drugs are direct and specific; therefore,

weak non-covalent interactions propitiated by heteroatoms' electronic distribution are very important for their interaction with receptors. However, the inclusion of local variables of superior order in both models demonstrates that an adequate molecular environment is also required for the interaction of antifungal drugs with their pharmacological target.

3.3 Comparison with other approaches for antifungal activity. In the last few years, some *in silico* methods have been used to develop ligand-based classification models on antifungal activity, which gave rise to a good discrimination of this activity.[28-30] However, an exhaustive comparison between these models and the models developed herein is not possible because of the differences in the experimental data used. Therefore, the comparison will be based on the number and diversity of chemical structural patterns contained in the data as well as on some classification and statistical parameters. Table 3 shows the comparison between antifungal models developed through ***TOMOCOMD–CARDD*** method and other reported approaches.

Firstly, the data set used to develop ***TOMOCOMD–CARDD*** based models have more than 26 and 22 times the number of antifungal agents with respect to models reported by Pastor *et al*[28] (Table 3, model 4 and 5) and Garcia-Domenech *et al*[29] (Table 3, model 6 as well as models 7-9 developed by Li *et al*[30]), respectively. Besides, the models developed by these authors cover a short range of mechanism of action compared with the broad range covered by our models.

Table 3 comes about here (see end the document)

Except for the training set of the model developed by Garcia-Domenech (Table 3, model 6) with a global good classification of 96.92%, ***TOMOCOMD–CARDD*** models

had a higher accuracy than all of the reported LDA equations both in the training and validation sets. This is remarkable taking into account the great structural diversity coded by ***TOMOCOMD–CARDD*** method.

4. CONCLUDING REMARKS AND FUTURE OUTLOOKS

In the last two decades the number of patients with severe fungal infections has dramatically increased and concern regarding the development of resistance to any of the few antifungal drugs available has developed.[124] On the other hand, the choice of suitable antifungal agents remains relatively limited due to their modest efficacy against life-threatening systemic fungal infections.[125] Despite aggressive management, the prognosis of invasive fungal disease, in particular those caused by filamentous fungi, continues to be dismal, with mortality rates exceeding 80% in selected categories of patients.[126]

Although the need for new drugs is clear, progress in that area is slow and unpredictable. It is stated that the ideal antifungal agent of the future should have a broad spectrum of fungicidal activity without mechanism-based host toxicity.[127] On the other hand, it takes around 13 years to bring a new antimicrobial to market[128] with an estimated cost of more than 800 million €.[129] This trend, which is similar to other therapeutic areas, has prompted different strategies within pharmaceutical companies and academic institutions to reduce the hit-to-drug timeline, increase the number of quality candidate drugs that make the transition from discovery to clinical development, and decrease the attrition rate (currently 90%) of candidate drugs in the clinical stages of the value chain.[130]

In this sense, the antifungal models developed in the present work are very valuable to design new agents able of fulfilling the above criteria. The fact that these models cover

a great structural diversity and known- and unknown-mechanism of actions coupled with their high accuracy, sensitivity and specificity to predict the antifungal activity is a step forward in the drug discovery process to any scientist wishing to develop new antifungals not only structurally related to known compounds but to generate new antifungal leads as well.

Supplementary Material Available: The complete list of compounds used in training and prediction sets, as well as their structures, posterior classification and scores according to LDA-based QSAR models is available free of charge via Internet at…

Acknowledgement: Yovani Marrero-Ponce (M-P. Y) acknowledges the Valencia University for kind hospitality during the first semester of 2008. M-P. Y thanks are given to the Generalitat Valenciana, (Spain) for partial financial support as well as the program 'Estades Temporals per an Investigadors Convidats' for a fellowship to work at Valencia University (2008). M-P. Y also thanks support from Spanish MEC (Project Reference: SAF2006-04698). Finally, but not less, M-P. Y thanks are also given to the project entitled "Strengthening postgraduate education and research in Pharmaceutical Sciences". This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium.

5. REFERENCES AND NOTES

1. Blumberg, H. M.; Jarvis, W. R.; Soucie, J. M.; Edwards, J. E.; Patterson, J. E.; Pfaller, M. A.; Rangel-Frausto, M. S.; Rinaldi, M. G.; Saiman, L.; Wiblin, R. T.; Wenzel, R. P. *Clin Infect Dis* **2001,** *33*, 177.
2. Maschmeyer, G. *Int J Antimicrob Agents* **2006,** *27 Suppl 1*, 3.
3. Nucci, M.; Marr, K. A. *Clin Infect Dis* **2005,** *41*, 521.
4. Rees, J. R.; Pinner, R. W.; Hajjeh, R. A.; Brandt, M. E.; Reingold, A. L. *Clin Infect Dis* **1998,** *27*, 1138.

5. Walsh, T. J.; Groll, A.; Hiemenz, J.; Fleming, R.; Roilides, E.; Anaissie, E. *Clin Microbiol Infect* **2004,** *10 Suppl 1*, 48.
6. Fridkin, S. K.; Jarvis, W. R. *Clin Microbiol Rev* **1996,** *9*, 499.
7. Groll, A. H.; Walsh, T. J. *Clin Microbiol Infect* **2001,** *7 Suppl 2*, 8.
8. Perea, S.; Patterson, T. F. *Clin Infect Dis* **2002,** *35*, 1073.
9. Pfaller, M. A.; Diekema, D. J. *J Clin Microbiol* **2004,** *42*, 4419.
10. Dickman, D. A.; Ding, H.; Li, Q.; Nilius, A. M.; Balli, D. J.; Ballaron, S. J.; Trevillyan, J. M.; Smith, M. L.; Seif, L. S.; Kim, K.; Sarthy, A.; Goldman, R. C.; Plattner, J. J.; Bennani, Y. L. *Bioorg Med Chem Lett* **2000,** *10*, 1405.
11. Spanakis, E. K.; Aperis, G.; Mylonakis, E. *Clin Infect Dis* **2006,** *43*, 1060.
12. Kauffman, C. A. *Proc Am Thorac Soc* **2006,** *3*, 35.
13. Gozalbes, R.; Galvez, J.; Moreno, A.; Garcia-Domenech, R. *J Pharm Pharmacol* **1999,** *51*, 111.
14. Marrero-Ponce, Y.; Montero-Torres, A.; Zaldivar, C. R.; Veitia, M. I.; Perez, M. M.; Sanchez, R. N. *Bioorg Med Chem* **2005,** *13*, 1293.
15. 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*, 1082.
16. McKie, J. H.; Douglas, K. T.; Chan, C.; Roser, S. A.; Yates, R.; Read, M.; Hyde, J. E.; Dascombe, M. J.; Yuthavong, Y.; Sirawaraporn, W. *J Med Chem* **1998,** *41*, 1367.
17. Estrada, E.; Pena, A.; Garcia-Domenech, R. *J Comput Aided Mol Des* **1998,** *12*, 583.
18. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J Med Chem* **2000,** *43*, 1975.
19. Gonzalez-Diaz, H.; Tenorio, E.; Castanedo, N.; Santana, L.; Uriarte, E. *Bioorg Med Chem* **2005,** *13*, 1523.
20. Scott, R. K. *Biosilico* **2003**, 14.
21. Seifert, M. H. J.; Wolf, K.; Vitt, D. *Biosilico* **2003**, 143.
22. Gollapudy, R.; Ajmani, S.; Kulkarni, S. A. *Bioorg Med Chem* **2004,** *12*, 2937.
23. Gokhale, V. M.; Kulkarni, V. M. *Bioorg Med Chem* **2000,** *8*, 2487.
24. Purushottamachar, P.; Kulkarni, V. M. *Bioorg Med Chem* **2003,** *11*, 3487.
25. Pranav Kumar, S. K.; Kulkarni, V. M. *Bioorg Med Chem* **2002,** *10*, 1153.
26. Fratev, F.; Benfenati, E. *J Chem Inf Model* **2005,** *45*, 634.
27. Gonzalez-Diaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. *Bioorg Med Chem* **2006,** *14*, 5973.
28. Pastor, L.; Garcia-Domenech, R.; Galvez, J.; Wolski, S.; Garcia, M. D. *Bioorg Med Chem Lett* **1998,** *8*, 2577.
29. García–Domenech, R.; Catalá–Gregori, A.; Calabuig, C.; Antón–Fos, G.; del Castillo, L.; Gálvez, J. *Internet Electron J Mol Des* **2002,** *7*, 339.
30. Chen, S.-W.; Li, Z.-R.; X-Y, L. *J. Mol. Struct. (THEOCHEM)* **2005,** *731*, 73.
31. Marrero-Ponce, Y. *Molecules* **2003,** *8*, 687.
32. Marrero-Ponce, Y. *J Chem Inf Comput Sci* **2004,** *44*, 2010.
33. 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.
34. Marrero-Ponce, Y. *Bioorg. Med. Chem.* **2004,** *12*, 6351.
35. 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*, 615.
36. Marrero-Ponce, Y.; Nodarse, D.; González, H. D.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. *Int. J. Mol. Sci.* **2004,** *5*, 276.
37. Marrero Ponce, Y.; Castillo Garit, J. A.; Nodarse, D. *Bioorg Med Chem* **2005,** *13*, 3397.
38. Marrero-Ponce, Y.; Medina-Marrero, R.; Castillo-Garit, J. A.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg Med Chem* **2005,** *13*, 3003.
39. Marrero-Ponce, Y.; Medina, R.; Castro, E. A.; de Armas, R.; González, H.; Romero, V.; Torrens, F. *Molecules* **2004,** *9* 1124.
40. Marrero Ponce, Y.; Cabrera Perez, M. A.; Romero Zaldivar, V.; Gonzalez Diaz, H.; Torrens, F. *J Pharm Pharm Sci* **2004,** *7*, 186.
41. Marrero-Ponce, Y.; Cabrera, M. A.; Romero-Zaldivar, V.; Bermejo, M.; Siverio, D.; Torrens, F. *Internet Electrón. J. Mol. Des.* **2005,** *4* 124.
42. Casanola-Martin, G. M.; Khan, M. T.; Marrero-Ponce, Y.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F. *Bioorg Med Chem Lett* **2006,** *16*, 324.
43. 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.
44. Vega, M. C.; Montero-Torres, A.; Marrero-Ponce, Y.; Rolon, M.; Gomez-Barrio, A.; Escario, J. A.; Aran, V. J.; Nogal, J. J.; Meneses-Marcel, A.; Torrens, F. *Bioorg Med Chem Lett* **2006,** *16*, 1898.
45. Marrero-Ponce, Y.; Machado-Tugores, Y.; Pereira, D. M.; Escario, J. A.; Barrio, A. G.; Nogal-Ruiz, J. J.; Ochoa, C.; Aran, V. J.; Martinez-Fernandez, A. R.; Sanchez, R. N.; Montero-Torres, A.; Torrens, F.; Meneses-Marcel, A. *Curr Drug Discov Technol* **2005,** *2*, 245.
46. Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. *Bioorg Med Chem* **2007,** *15*, 1483.
47. Marrero-Ponce, Y.; Khan, M. T.; Casanola Martin, G. M.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F.; Rotondo, R. *ChemMedChem* **2007,** *2*, 449.
48. Marrero-Ponce, Y.; Castillo-Garit, J. A. *J Comput Aided Mol Des* **2005,** *19*, 369.
49. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. *Bioorg Med Chem* **2005,** *13*, 2881.
50. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. *J Mol Model (Online)* **2006,** *12*, 255.
51. Marrero-Ponce, Y.; Romero, V.: Central University of Las Villas., 2002.
52. Marrero-Ponce, Y.; Díaz, H. G.; Romero, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004,** *12*, 5331–5342.
53. Pauling, L. *The Nature of Chemical Bond*; Ithaca, NY: Cornell University Press, 1939.
54. Bedi, P. M.; Mahajan, M. P.; Kapoor, V. K. *Bioorg Med Chem Lett* **2004,** *14*, 3821.
55. Brayman, T. G.; Wilks, J. W. *Antimicrob Agents Chemother* **2003,** *47*, 3305.

56. Bryskier, A. *Clin Infect Dis* **2000,** *31*, 1423.
57. Buchta, V.; Pour, M.; Kubanova, P.; Silva, L.; Votruba, I.; Voprsalova, M.; Schiller, R.; Fakova, H.; Spulak, M. *Antimicrob Agents Chemother* **2004,** *48*, 873.
58. Bueno, J. M.; Cuevas, J. C.; Fiandor, J. M.; Garcia-Ochoa, S.; Gomez de las Heras, F. *Bioorg Med Chem Lett* **2002,** *12*, 121.
59. Carta, A.; Paglietti, G.; Rahbar Nikookar, M. E.; Sanna, P.; Sechi, L.; Zanetti, S. *Eur J Med Chem* **2002,** *37*, 355.
60. Clemons, K. V.; Stevens, D. A. *Antimicrob Agents Chemother* **1997,** *41*, 200.
61. Debono, M.; Gordee, R. S. *Annu Rev Microbiol* **1994,** *48*, 471.
62. Del Poeta, M.; Schell, W. A.; Dykstra, C. C.; Jones, S.; Tidwell, R. R.; Czarny, A.; Bajic, M.; Kumar, A.; Boykin, D.; Perfect, J. R. *Antimicrob Agents Chemother* **1998,** *42*, 2495.
63. Del Poeta, M.; Schell, W. A.; Dykstra, C. C.; Jones, S. K.; Tidwell, R. R.; Kumar, A.; Boykin, D. W.; Perfect, J. R. *Antimicrob Agents Chemother* **1998,** *42*, 2503.
64. Denning, D. W. *J Antimicrob Chemother* **1997,** *40*, 611.
65. Dominguez, J. M.; Kelly, V. A.; Kinsman, O. S.; Marriott, M. S.; Gomez de las Heras, F.; Martin, J. J. *Antimicrob Agents Chemother* **1998,** *42*, 2274.
66. Ebiike, H.; Masubuchi, M.; Liu, P.; Kawasaki, K.; Morikami, K.; Sogabe, S.; Hayase, M.; Fujii, T.; Sakata, K.; Shindoh, H.; Shiratori, Y.; Aoki, Y.; Ohtsuka, T.; Shimma, N. *Bioorg Med Chem Lett* **2002,** *12*, 607.
67. Emami, S.; Falahati, M.; Banifatemi, A.; Shafiee, A. *Bioorg Med Chem* **2004,** *12*, 5881.
68. Favre, B.; Ryder, N. S. *Antimicrob Agents Chemother* **1996,** *40*, 443.
69. Fostel, J. M.; Lartey, P. A. *Drug Discov Today* **2000,** *5*, 25.
70. Goker, H.; Boykin, D. W.; Yildiz, S. *Bioorg Med Chem* **2005,** *13*, 1707.
71. Gokhale, V. M.; Kulkarni, V. M. *J Comput Aided Mol Des* **2000,** *14*, 495.
72. Hata, K.; Kimura, J.; Miki, H.; Toyosawa, T.; Moriyama, M.; Katsu, K. *Antimicrob Agents Chemother* **1996,** *40*, 2243.
73. Herreros, E.; Almela, M. J.; Lozano, S.; Gomez de las Heras, F.; Gargallo-Viola, D. *Antimicrob Agents Chemother* **2001,** *45*, 3132.
74. Herreros, E.; Martinez, C. M.; Almela, M. J.; Marriott, M. S.; De Las Heras, F. G.; Gargallo-Viola, D. *Antimicrob Agents Chemother* **1998,** *42*, 2863.
75. Kamai, Y.; Harasaki, T.; Fukuoka, T.; Ohya, S.; Uchida, K.; Yamaguchi, H.; Kuwahara, S. *Antimicrob Agents Chemother* **2002,** *46*, 367.
76. Khan, J. K.; Montaseri, H.; Poglod, M.; Bu, H. Z.; Zuo, Z.; Salama, S. M.; Daneshtalab, M.; Micetich, R. G. *Antimicrob Agents Chemother* **2000,** *44*, 910.
77. Kubo, I.; Xiao, P.; Fujita, K. *Bioorg Med Chem Lett* **2001,** *11*, 347.
78. Lal, B.; Gund, V. G.; Bhise, N. B.; Gangopadhyay, A. K. *Bioorg Med Chem* **2004,** *12*, 1751.
79. Lal, B.; Gund, V. G.; Gangopadhyay, A. K.; Nadkarni, S. R.; Dikshit, V.; Chatterjee, D. K.; Shirvaikar, R. *Bioorg Med Chem* **2003,** *11*, 5189.
80. Li, S.; Zhang, Z.; Cain, A.; Wang, B.; Long, M.; Taylor, J. *J Agric Food Chem* **2005,** *53*, 32.
81. Lopez, S. N.; Castelli, M. V.; Zacchino, S. A.; Dominguez, J. N.; Lobo, G.; Charris-Charris, J.; Cortes, J. C.; Ribas, J. C.; Devia, C.; Rodriguez, A. M.; Enriz, R. D. *Bioorg Med Chem* **2001,** *9*, 1999.

82. Mandala, S. M.; Thornton, R. A.; Milligan, J.; Rosenbach, M.; Garcia-Calvo, M.; Bull, H. G.; Harris, G.; Abruzzo, G. K.; Flattery, A. M.; Gill, C. J.; Bartizal, K.; Dreikorn, S.; Kurtz, M. B. *J Biol Chem* **1998,** *273*, 14942.
83. Mandala, S. M.; Thornton, R. A.; Rosenbach, M.; Milligan, J.; Garcia-Calvo, M.; Bull, H. G.; Kurtz, M. B. *J Biol Chem* **1997,** *272*, 32709.
84. Masubuchi, K.; Okada, T.; Kohchi, M.; Murata, T.; Tsukazaki, M.; Kondoh, O.; Yamazaki, T.; Satoh, Y.; Ono, Y.; Tsukaguchi, T.; Kobayashi, K.; Ono, N.; Inoue, T.; Horii, I.; Shimma, N. *Bioorg Med Chem Lett* **2001,** *11*, 1273.
85. Masubuchi, M.; Ebiike, H.; Kawasaki, K.; Sogabe, S.; Morikami, K.; Shiratori, Y.; Tsujii, S.; Fujii, T.; Sakata, K.; Hayase, M.; Shindoh, H.; Aoki, Y.; Ohtsuka, T.; Shimma, N. *Bioorg Med Chem* **2003,** *11*, 4463.
86. Masubuchi, M.; Kawasaki, K.; Ebiike, H.; Ikeda, Y.; Tsujii, S.; Sogabe, S.; Fujii, T.; Sakata, K.; Shiratori, Y.; Aoki, Y.; Ohtsuka, T.; Shimma, N. *Bioorg Med Chem Lett* **2001,** *11*, 1833.
87. Matsumoto, M.; Ishida, K.; Konagai, A.; Maebashi, K.; Asaoka, T. *Antimicrob Agents Chemother* **2002,** *46*, 308.
88. Menozzi, G.; Merello, L.; Fossa, P.; Schenone, S.; Ranise, A.; Mosti, L.; Bondavalli, F.; Loddo, R.; Murgioni, C.; Mascia, V.; La Colla, P.; Tamburini, E. *Bioorg Med Chem* **2004,** *12*, 5465.
89. Moreau, S.; Varache-Lembege, M.; Larrouture, S.; Fall, D.; Neveu, A.; Deffieux, G.; Vercauteren, J.; Nuhrich, A. *Eur J Med Chem* **2002,** *37*, 237.
90. Nam, N. H.; Sardari, S.; Selecky, M.; Parang, K. *Bioorg Med Chem* **2004,** *12*, 6255.
91. Nimura, K.; Niwano, Y.; Ishiduka, S.; Fukumoto, R. *Int J Antimicrob Agents* **2001,** *18*, 173.
92. Niwano, Y.; Kuzuhara, N.; Goto, Y.; Munechika, Y.; Kodama, H.; Kanai, K.; Yoshida, M.; Miyazaki, T.; Yamaguchi, H. *Int J Antimicrob Agents* **1999,** *12*, 221.
93. Panneerselvam, P.; Nair, R. R.; Vijayalakshmi, G.; Subramanian, E. H.; Sridhar, S. K. *Eur J Med Chem* **2005,** *40*, 225.
94. Quesnelle, C. A.; Gill, P.; Dodier, M.; St Laurent, D.; Serrano-Wu, M.; Marinier, A.; Martel, A.; Mazzucco, C. E.; Stickle, T. M.; Barrett, J. F.; Vyas, D. M.; Balasubramanian, B. N. *Bioorg Med Chem Lett* **2003,** *13*, 519.
95. Ryu, C. K.; Kang, H. Y.; Yi, Y. J.; Shin, K. H.; Lee, B. H. *Bioorg Med Chem Lett* **2000,** *10*, 1589.
96. Sheehan, D. J.; Hitchcock, C. A.; Sibley, C. M. *Clin Microbiol Rev* **1999,** *12*, 40.
97. Stephens, C. E.; Tanious, F.; Kim, S.; Wilson, W. D.; Schell, W. A.; Perfect, J. R.; Franzblau, S. G.; Boykin, D. W. *J Med Chem* **2001,** *44*, 1741.
98. Tkacz, J. S.; DiDomenico, B. *Curr Opin Microbiol* **2001,** *4*, 540.
99. Tsuchimori, N.; Hayashi, R.; Kitamoto, N.; Asai, K.; Kitazaki, T.; Iizawa, Y.; Itoh, K.; Okonogi, K. *Antimicrob Agents Chemother* **2002,** *46*, 1388.
100. Urbina, J. M.; Cortes, J. C.; Palma, A.; Lopez, S. N.; Zacchino, S. A.; Enriz, R. D.; Ribas, J. C.; Kouznetzov, V. V. *Bioorg Med Chem* **2000,** *8*, 691.
101. Vargas, M. L.; Castelli, M. V.; Kouznetsov, V. V.; Urbina, G. J.; Lopez, S. N.; Sortino, M.; Enriz, R. D.; Ribas, J. C.; Zacchino, S. *Bioorg Med Chem* **2003,** *11*, 1531.
102. Yalcin, I.; Oren, I.; Temiz, O.; Sener, E. A. *Acta Biochim Pol* **2000,** *47*, 481.

103. Yotsuji, A.; Shimizu, K.; Araki, H.; Fujimaki, K.; Nishida, N.; Hori, R.; Annen, N.; Yamamoto, S.; Hayakawa, H.; Imaizumi, H.; Watanbe, Y.; Narita, H. *Antimicrob Agents Chemother* **1997,** *41*, 30.
104. Zhang, Y. Z.; Sun, X.; Zeckner, D. J.; Sachs, R. K.; Current, W. L.; Chen, S. H. *Bioorg Med Chem Lett* **2001,** *11*, 123.
105. Ablordeppey, S. Y.; Fan, P.; Li, S.; Clark, A. M.; Hufford, C. D. *Bioorg Med Chem* **2002,** *10*, 1337.
106. Saha, A. K.; Liu, L.; Simoneaux, R. L.; Kukla, M. J.; Marichal, P.; Odds, F. *Bioorg Med Chem Lett* **2000,** *10*, 2175.
107. Barrett, D.; Tanaka, A.; Harada, K.; Ohki, H.; Watabe, E.; Maki, K.; Ikeda, F. *Bioorg Med Chem Lett* **2001,** *11*, 479.
108. Bryskier, A. *Clin Infect Dis* **1998,** *27*, 865.
109. Martinez, A.; Ferrer, S.; Santos, I.; Jimenez, E.; Sparrowe, J.; Regadera, J.; De Las Heras, F. G.; Gargallo-Viola, D. *Antimicrob Agents Chemother* **2001,** *45*, 3304.
110. Serrano-Wu, M. H.; St Laurent, D. R.; Mazzucco, C. E.; Stickle, T. M.; Barrett, J. F.; Vyas, D. M.; Balasubramanian, B. N. *Bioorg Med Chem Lett* **2002,** *12*, 943.
111. Turan-Zitouni, G.; Kaplancikli, Z. A.; Yildiz, M. T.; Chevallet, P.; Kaya, D. *Eur J Med Chem* **2005,** *40*, 607.
112. Na, Y. M.; Le Borgne, M.; Pagniez, F.; Le Baut, G.; Le Pape, P. *Eur J Med Chem* **2003,** *38*, 75.
113. ; Chapman & Hall
New York, 1996.
114. Ruiz-Herrera, J.; San-Blas, G. *Curr Drug Targets Infect Disord* **2003,** *3*, 77.
115. STATISTICA version. 6.0 Statsoft, I.
116. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. *Bioinformatics* **2000,** *16*, 412.
117. Randić, M. *J Chem Inf Comput Sci* **1991,** *31*, 311.
118. Randić, M. *New J Chem* **1991,** *15*, 517.
119. Randić, M. *J Mol Struct (Theochem)* **1991,** *233*, 45.
120. Estrada, E.; Uriarte, E. *Curr Med Chem* **2001,** *8*, 1573.
121. Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. *J Chem Inf Comput Sci* **2002,** *42*, 1194.
122. Wold, S.; Erikson, L. In *Chemometric Methods in Molecular Design;* van de Waterbeemd, H., Ed.; VCH Publishers: Weinheim, 1995, 309.
123. Golbraikh, A.; Tropsha, A. *Mol Divers* **2002,** *5*, 231.
124. Loeffler, J.; Stevens, D. A. *Clin Infect Dis* **2003,** *36*, S31.
125. Monk, B. C.; Cannon, R. D. *Curr Drug Targets Infect Disord* **2002,** *2*, 309.
126. Nosanchuk, J. D. *Recent Patents on Anti-Infective Drug Discovery* **2006,** *1*, 75.
127. Carrillo-Munoz, A. J.; Giusiano, G.; Ezkurra, P. A.; Quindos, G. *Rev Esp Quimioter* **2006,** *19*, 130.
128. Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. *Nat Rev Drug Discov* **2007,** *6*, 29.
129. Vicente, M.; Hodgson, J.; Massidda, O.; Tonjum, T.; Henriques-Normark, B.; Ron, E. Z. *FEMS Microbiol Rev* **2006,** *30*, 841.
130. Lyne, P. D. *Drug Discov Today* **2002,** *7*, 1047.

ANNEXES
(Tables, Figures and Schemes to be Inserted in the Main Text)

Table 1. Results of the Training and Prediction Series Performances using the Atom-based Quadratic Indices.

	Matthew´s corr. coefficient.	**Accuracy 'Q_{Total}' (%)**	**Sensitivity (%)**	**Specificity (%)**	**False positive rate ''false alarm rate'' (%)**
Non-stochastic descriptors (Eq. 2)					
Training Set	0.81	90.73	91.07	90.44	9.59
Test Set	0.84	92.06	92.16	92.66	8.03
Stochastic descriptors (Eq. 3)					
Training Set	0.85	92.47	91.21	93.56	6.25
Test Set	0.79	89.66	87.56	92.30	8.03

Table 2. Results of Randić's Orthogonalization Analysis.

		Non-orthogonal Quadratic Indices					
$q_5(\bar{x})$	$q_6(\bar{x})$	$q_{1L}{}^{H}(\bar{x}_E)$	$q_{12L}(\bar{x}_{E\text{-}H})$	${}^{s}q_{11}(\bar{x})$	${}^{s}q_{8L}{}^{H}(\bar{x}_E)$	${}^{s}q_{0}{}^{H}(\bar{x})$	${}^{s}q_{6L}{}^{H}(\bar{x}_E)$
1.00	0.99	0.61	0.54	1.00	0.90	0.99	0.90
	1.00	0.55	0.54		1.00	0.87	0.99
		1.00	0.56			1.00	0.87
			1.00				1.00
		Orthogonal Quadratic Indices					
$O(q_5(x))$	$O(q_6(x))$	$O(q_{1L}{}^{H}(x_E))$	$O(q_{12L}(x_{E\text{-}H}))$	${}^{1}O({}^{s}q_{11}(x))$	${}^{2}O({}^{s}q_{8L}{}^{H}(x_E))$	${}^{3}O({}^{s}q_{0}{}^{H}(x))$	${}^{4}O({}^{s}q_{6L}{}^{H}(x_E))$
1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	1.00	0.00	0.00		1.00	0.00	0.00
		1.00	0.00			1.00	0.00
			1.00				1.00

LDA based QSAR models derived from the orthogonal non-stochastic and stochastic quadratic indices

$$\boldsymbol{AFA} = -0.004 + 3.24\,{}^{1}O(q_5(\bar{x})) - 1.47\,{}^{2}O(q_{12L}(\bar{x}_{E\text{-}H})) - 10.81\,{}^{3}O(q_{1L}{}^{H}(\bar{x}_E)) - 1.001\,{}^{4}O(q_6(\bar{x}\,x)) \quad \textbf{(Eq. 2a)}$$

N = 1436 λ = 0.41 D^2 = 5.69 F(4, 1431) = 509.72
Rcan = 0.766 χ^2 = 1268.38 ***C*** = 0.81 Q_{Total} = 90.73

$$\boldsymbol{AFA} = -0.007 + 2.38\,{}^{1}O({}^{s}q_{11}(\bar{x})) - 4.67\,{}^{2}O({}^{s}q_{8L}{}^{H}(\bar{x}_E)) - 17.56\,{}^{3}O({}^{s}q_{0}{}^{H}(\bar{x})) - 37.0\,{}^{4}O({}^{s}q_{6L}{}^{H}(\bar{x}_E)) \quad \textbf{(Eq. 3a)}$$

N = 1436 λ = 0.39 D^2 = 6.07 F(4, 1431) = 544.51
Rcan = 0.777 χ^2 = 1324.69 ***C*** = 0.85 Q_{Total} = 92.47

Table 3. Comparison between the models developed in this study with other cheminformatic approaches.

Models' features to be compared[a]	***Classification models of the antifungal activity***							
	Eq. 2 (LDA)	Eq. 3 (LDA)	Eq. 4 (LDA)	Eq. 5 (LDA)	Eq. 6 (LDA)	**Eq. 7** (SVM)	**Eq. 8** (k-NN)	**Eq. 9** (C4.5)
N total	2142	2142	94	94	90	94	94	94
N antifungals	1087	1087	42	42	49	42	42	42
Wilks'λ	0.41	0.39	0.392	0.387	0.32	-	-	-
F	509.72	544.51	14.5	19.5	17.2	-	-	-
D^2	5.69	6.07	-	-	-	-	-	-
p-level	<0.0001	<0.0001	-	-	-	-	-	-
χ^2	1268.38	1324.69	-	-	-	-	-	-
Rcan	0.766	0.777	-	-	-	-	-	-
Variables in the model	4	4	9	9	8	60(30)*	30*	30*
Learning set								
N total	1436	1436	94	94	65	94**	94**	94**
N antifungals	717	717	42	42	36	42**	42**	42**
Matthews Corr. Coefficient (*C*)	0.81	0.85	0.74	0.77	0.94			
Accuracy 'Q_{Total}'	90.73	92.47	87.23	88.29	96.92	84.0(89.4)**	76.5**	75.6**
Specificity (%)	90.44	93.56	87.50	91.89	97.22	-	-	-
Sensitivity (%)	91.07	91.21	83.33	80.95	97.22	91.0(97.1)**	71.7	73.5
False + Rate (%)	9.56	6.25	9.62	5.76	3.44	-	-	-
Families of drugs[b]	Broad range	Broad range	Short range	Short range	Short range	Short range	Short range	Short range
Validation methods								
Validation method[c]	i	i	ii	ii	i	iii	iii	iii
N total	706	706	-	-	25	-	-	-
N antifungals	370	370	-	-	13	-	-	-
Matthews Corr. Coefficient (*C*)	0.84	0.79	0.71	0.75	0.60	-	-	-
Accuracy 'Q_{Total}'	92.06	89.66	85.56	87.80	80.00	77.8	75.0	85.7
Specificity (%)	92.66	92.30	85.71	88.57	83.33	-	-	-
Sensitivity (%)	92.16	87.56	85.71	83.78	76.92	-	-	-
False + Rate (%)	8.03	8.03	14.58	8.88	16.66	-	-	-
Families of drugs[b]	Broad range	Broad range	Short range	Short range	Short range	Short range	Short range	Short range

*Equation fited employed the whle set of 62 variables, between braket number of predictors selected by using Genetic algorithm after develop SVM.[30] **Result obtained form cross-validation experiment by using 5-fold out. [a]Equations 2 and 3 are reported in this work, the models 4 and 5 were reported by Pastor et al.,[28]

4 equation 6 was reported by Garcia-Domenech et al.[29] and models 7-9 were reported by Li *et al*.[30] [b]Only were taken into account compound families with a wide
5 representative's [c]Validation methods are: i) external prediction series, and ii) leave-20%-out, iii) an independent test set.
6