

2D Chemometric Studies of a Series of Azole Derivatives Active against Fluconazole-Resistant *Cryptococcus gattii*

Humberto F. Freitas,^a Tania F. Barros^a and Marcelo S. Castilho^{*a,b}

^aFaculdade de Farmácia, Universidade Federal da Bahia, 40170-290 Salvador-BA, Brazil

^bInstituto Nacional de Ciência e Tecnologia em Biologia Estrutural e Bioimagem, Universidade Federal do Rio de Janeiro (CCS/UFRJ), 21944-970 Rio de Janeiro-RJ, Brazil

Apesar dos avanços no desenvolvimento de antifúngicos, tem ocorrido um aumento de casos de criptococose que não respondem de forma adequada a fluconazol (fármaco de primeira escolha). Portanto, é de suma importância investigar as propriedades químicas de derivados azólicos que sejam ativos contra cepas de *Cryptococcus neoformans* resistentes a fluconazol. Visando alcançar esse objetivo, o perfil de suscetibilidade de um isolado clínico de *C. neoformans* resistente contra 33 derivados azólicos comerciais foi avaliado junto com as suas respectivas concentrações inibitórias mínimas (MIC). Esses dados foram utilizados para construir modelos SIMCA (modelagem independente flexível por analogias de classes) que destacam a importância de propriedades eletrônicas (JGI10) para separar as moléculas ativas das inativas e para construir modelos de holograma-QSAR que apresentam bom ajuste, mas capacidade preditiva baixa (HQSAR, $r^2 = 0.85$, $q^2 = 0.35$ e $r^2_{pred} = 0.38$). Por outro lado, modelos de QSAR 2D desenvolvidos a partir de descritores topológicos apresentaram boa qualidade estatística ($r^2 = 0.95$, $q^2 = 0.86$, $r^2_{pred} = 0.72$) e destacam que a distribuição de cargas (GGI1) e a eletronegatividade topológica (GATS1e e MATS2e) devem ser modulados para contornar a resistência de *C. neoformans*.

Despite advances in the development of antifungal drugs, there has been an upsurge of cryptococosis infections that poorly respond to fluconazole (first choice drug). Hence, it is paramount to investigate the chemical properties of azole derivatives that are active against resistant *C. neoformans*. In order to achieve this goal, the susceptibility profile of a clinical isolate of resistant *C. neoformans* against 33 commercial azole derivatives was evaluated along with their potency (minimum inhibitory concentration, MIC). These data were employed to build SIMCA (soft independent modeling of class analogies) models that pinpoint the importance of electronic features (JGI10) to separate active from inactive compounds and hologram-QSAR models that have good fit but insufficient predictive power (HQSAR, $r^2 = 0.85$, $q^2 = 0.35$ and $r^2_{pred} = 0.38$). Conversely, 2D QSAR models built from topological descriptors improved the statistical quality ($r^2 = 0.95$, $q^2 = 0.86$, $r^2_{pred} = 0.72$) and highlight that charge distribution (GGI1) and topological electronegativity (GATS1e and MATS2e) should be modulated to overcome the *C. neoformans* resistance.

Keywords: *Cryptococosis*, azole resistance, 2D QSAR, SIMCA, HQSAR

Introduction

Fungal meningitis, caused either by *Cryptococcus neoformans* or *Cryptococcus gattii*,¹ distresses approximately 1 million people each year and causes more than 600 thousand deaths.² The first one is worldwide relevant, especially for HIV positive patients, while the second also can affect immunocompetent

persons in tropical and subtropical regions.³ Currently, azole derivatives are considered as the first choice drug for long-term therapy.⁴ However, the emergence of fluconazole-resistant *C. neoformans* isolates, along with the fact that *C. gattii* infections respond poorly to azole treatment,^{5,6} poses a tremendous hurdle to this approach.^{7,8} Taking this scenario into consideration, many research groups have focused their efforts towards the development of novel antifungal drugs that could circumvent resistance issues.⁹ Most antifungal drug design campaigns have relied on

*e-mail: castilho@ufba.br

old-fashioned trial and error paradigm, according to which lead compounds have their structures modified, guided by synthetic feasibility and intermediate biological assay results, until the desired potency/selectivity is achieved. Modern structure-based drug design strategies have been hampered by the lack of structural data on fungal lanosterol 14 α demethylase, the azole therapeutic target, despite the fact that homology modeling has been employed to partially solve this limitation.¹⁰⁻¹³

An alternative strategy would be to explore ligand-based strategies to extract information from the large amount of data available from phenotypic assays. This approach allows the screening of hundreds of compounds at low costs, within a reasonable amount of time, and takes into consideration pharmacokinetic issues, such as drug permeability through the cell membrane, that are not accounted for in target-based assays.¹⁴ This sort of biological data has already been used, for instance, to build descriptor-based classificatory models (k-nearest neighbor (KNN) and soft independent modeling of class analogy (SIMCA)) that hint at the importance of electronegativity and dipolar moment towards the biological activity of azole derivatives against *Moniliophthora perniciosa*, the causal agent of Witches Broom disease.¹⁵ Although all azole antifungals target the same macromolecule, subtle differences in the binding site might alter the structural requirements for potency. In fact, mutations in the binding site of *C. gattii* lanosterol 14- α demethylase have been associated with different susceptibility profile towards antifungal drugs.¹⁶⁻¹⁸ Thus, our group decided to investigate the structure-activity relationships of commercial azole against a fluconazole resistant *C. gattii* clinical isolate by means of 2D chemometric approaches.^{15,19} So, in order to accomplish the goal, the biological activities of 33 azole derivatives were measured under standardized conditions and used to build robust and predictive classificatory and quantitative 2D chemometric models that underscore crucial chemical features for antifungal activity towards a fluconazole resistant *C. gattii* strain.

Experimental

Reagents

The 33 imidazole and triazole derivatives employed in this study were purchased from the Sigma-Aldrich Fisher Company (St. Louis, MO, EUA) with purity equal or superior to 95% (Table S1 in the Supplementary Information (SI) section). All other reagents used in the culture medium preparation, buffer solutions and so on were acquired from well-known chemical companies

(example Sigma-Aldrich and Merck chemical (Darmstadt, Germany)) and used without further purification.

Disc diffusion susceptibility assay

The activity profile of 33 azole derivatives against a fluconazole resistant *C. gattii* strain (available in the Microbiology Research laboratory (LPMC) from Pharmacy School of Federal University of Bahia) was evaluated by disk diffusion susceptibility assay.²⁰

Considering the lack of standards concerning *Cryptococcus* spp assays, all compounds were initially evaluated at the standard concentrations employed for an American Type Culture Collection *Candida albicans* (ATCC 90028): 8.2 mmol L⁻¹ (25 μ g disc⁻¹) as suggested by M-44A2 guideline.²⁰ Twice the standard concentration (16.4 mmol L⁻¹) and half of it (4.1 mmol L⁻¹) were also evaluated to probe the susceptibility range profile (Table 1).

Table 1. Biological properties used for chemometric model development

Name	Diameter of inhibition halo / mm			MIC / (μ mol L ⁻¹)
	16.4 mmol L ⁻¹	8.2 mmol L ⁻¹	4.1 mmol L ⁻¹	
Azaconazole	23.3	0.0	0.0	106.62
Bifonazole	14.3	12.3	11.3	> 206.19
Bromuconazole	47.3	43.3	31.3	21.22
Ketoconazole	59.0	44.0	37.3	0.47
Climbazole	57.7	46.7	41.0	6.83
Clotrimazole	35.0	33.3	32.3	23.20
Cyproconazole	60.7	57.7	57.0	3.43
Diclobutrazole	29.3	24.7	14.0	24.37
Difenoconazole	36.0	33.3	31.7	2.46
Diniconazole	38.7	36.7	34.3	6.13
Econazole	30.3	28.0	27.0	10.48
Epoxiconazole	34.7	35.0	34.3	3.03
Etaconazole	38.0	31.0	23.3	24.38
Fembuconazole	28.0	29.0	26.0	23.75
Fluconazole	0.0	0.0	0.0	104.48
Fluotrimazole	0.0	0.0	0.0	168.70
Fluquinconazole	38.3	37.0	35.0	42.53
Flusilazole	26.0	26.3	19.7	25.37
Flutriafole	0.0	0.0	0.00	> 212.42
Hexaconazole	63.7	54.0	55.3	6.37
Imazalil	29.7	21.7	12.0	53.84
Itraconazole	22.3	21.0	19.3	0.18
Metconazole	54.3	51.7	44.0	3.13
Miconazole	39.3	33.3	26.7	2.40
Myclobutanil	25.7	13.3	0.0	55.41
Penconazole	38.7	33.0	22.0	14.08
Prochloraz	17.0	0.0	0.0	84.96
Propiconazole	36.0	35.7	35.7	11.69
Prothioconazole	42.7	42.3	40.3	92.95
Sulconazole	39.0	38.0	37.7	0.63
Tebuconazole	39.3	33.3	26.7	12.99
Triadimenol	18.0	11.0	0.00	10.11
Triticonazole	26.7	26.0	19.3	12.59
Average	33.0	28.3	24.1	

All assays were carried out in triplicate, along with positive growth, negative growth and dimethyl sulfoxide (DMSO) controls. Process control (manipulation) was probed by simultaneous evaluation of *Candida albicans* (ATCC 90028) and *Candida parapsilosis* (ATCC 22019) susceptibility profile against standard fluconazole concentration (8.2 mmol L^{-1}).

Minimum inhibitory concentration assays

The broth microdilution procedure was employed to determine the minimum inhibitory concentrations (MIC).²¹ Briefly, the azole derivatives were solubilized in DMSO, except for fluconazole that was solubilized in Roswell Park Memorial Institute medium (RPMI) 1640, to produce stock solutions ($1600\text{--}5120 \text{ }\mu\text{g mL}^{-1}$ for fluconazole) that were serially diluted and dispensed into microdilution plates. The final solutions (ranging from 128 to $0.062 \text{ }\mu\text{g mL}^{-1}$) were inoculated with a McFarland 0.5 standard saline suspension, containing *C. gattii* strain, diluted into saline:RPMI solution (1:100).

All MICs were defined as the lowest concentration of azole that produces more than 50% inhibition growth (Table 1). Every measurement was carried out in quadruplicate and positive control (fluconazole MIC against standard strains *Candida parapsilosis* (ATCC 22019) and *Candida krusei* (ATCC 6258)), sterility control (only RPMI) and DMSO control (20 μL of DMSO, 80 μL of RPMI and 100 μL of fungal suspension) were also carried out.²¹

Chemometric analysis

Data set

The data set used for the chemometric studies comprises 33 azole derivatives, whose biological

activity against *Cryptococcus gattii* was measured as described previously. The chemical structures were sketched in Sybyl-X 1.1 platform (Tripos Inc., St. Louis, USA) and energy optimized with Tripos molecular force-field (gradient < 0.01), using Gasteiger-Huckel charges and a dielectric constant equal to 80. Next, PM3 semi-empirical method (keywords: 1SCF XYZ ESP NOINTER SCALE=1.4 NSURF=2 SCINCR=0.4 NOMM) was used to assign partial charges to all molecules, as available in Sybyl-X 1.1 Mopac module.

Two biological properties were evaluated in this study: susceptibility assay distinguishes active from inactive azoles against fluconazole resistant *C. gattii*, whereas MIC values probe the potency profile. Hence, these values were separately used to build classification/pattern recognition models and multiple linear/partial least squares regression models. Thus, the diameter of inhibition halo, measured at 4.1 mmol L^{-1} (Table 1), was used to split the compounds into active (inhibition halo larger than 25 mm) or inactive (Figure 1 and Table 2) compounds. MIC values, expressed in molar concentration (Table 1), were transformed to pMIC ($-\log [\text{MIC}]$) and used as dependent variables in

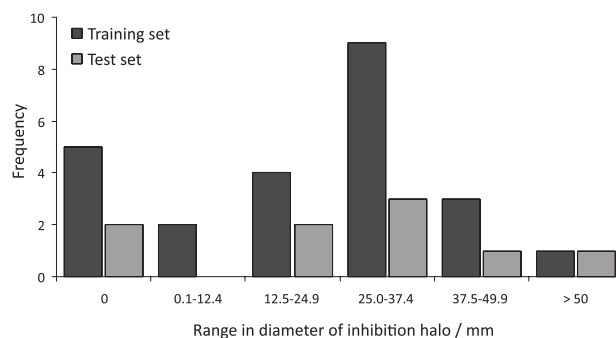


Figure 1. Frequency histogram of the compounds at the lower concentration (4.1 mmol L^{-1}) in relation to range in the diameter of inhibition halo against *C. gattii*.

Table 2. Division of the compounds into training and test sets

Active compound		Inactive compound	
Training set	Test set	Training set	Test set
Bromuconazole	Difenoconazole	Azaconazole	Fluconazole
Ketoconazole	Femuconazole	Bifonazole	Flusilazole
Climbazole	Hexaconazole	Diclobutrazole	Myclobutanil
Clotrimazole	Metconazole	Etaconazole	Triticonazole
Cyproconazole	Propiconazole	Fluotrimazole	
Diniconazole		Flutriafol	
Econazole		Imazalil	
Epoxiconazole		Itraconazole	
Fluquinconazole		Penconazole	
Miconazole		Prochloraz	
Prothioconazole		Triadimenol	
Sulconazole			
Tebuconazole			

QSAR model development (Figure S1 in the SI section and Table 4).

Descriptor calculation and selection

2D descriptors available in DRAGON 5.5 (Taletto SRL, Milan, Italy) were calculated for the whole dataset. Next, descriptors with high pairwise-correlation ($\geq 97\%$) and low variance ($< 10\%$) were excluded. The remaining descriptors were subjected to different selection protocols, as described below:

(i) Classification/pattern recognition models: Fisher's weight^{15,22} was employed to identify descriptors that individually differentiate the two classes of compounds. Descriptors with values above the mean plus two times the standard deviation (95% confidence interval) were selected for further chemometric analysis (KNN and SIMCA) available in the Pirouette 4.0 software (Infometrix Inc., Washington, USA).

(ii) MOBYDIGS 1.0 software (Taletto SRL, Milan, Italy) was employed to build preliminary multiple-linear regression QSAR (MLR-QSAR) models, with up to 4 descriptors *per* model through genetic algorithm.²³ MLR-QSAR model optimization was guided by the following rules: QUIK rule (0.05), asymptotic Q^2 rule (-0.005), redundancy RP rule (0.1) and overfitting RN rule (0.01).^{24,25} Due to the stochastic nature of genetic algorithm, the search was carried out in ten independent populations of 2000 models. Each population evolved for at least 1500 generations. Finally, the descriptors found in models with $q^2 > 0.64$ were pooled together, autoscaled and employed for further chemometric analysis (hierarchical cluster analysis (HCA), principal component analysis (PCA) and partial least square regression (PLS)) available in the Pirouette 4.0 software.

KNN and SIMCA studies

KNN models were built considering the Euclidean distance between the unknown sample (yet to be classified) and the *k*-th nearest neighbors, whose class is previously known.^{26,27} Too small or too large, a number of neighbors causes either instability or loss of precision to the model, so the optimum number of neighbors was determined by leave-one-out cross validation (LOO_{cv}).

LOO_{cv} approach was also employed to evaluate the number of PCs that should be used to describe each class in SIMCA models. The model improvement was carried out by the iterative elimination of descriptors with low discriminant or modeling power.

Hologram QSAR studies

Statistical HQSAR modeling was carried out as previously described.^{19,28-30} Briefly, all molecules were

broken down into unique fragments (linear, branched, cyclic, overlapping and so forth) according to different fragment distinction features (atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch) and donor/acceptor (DA)). Initially, only fragments with 4-7 atoms were considered. These fragments were hashed to a length-fixed array of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 or 401 bins, which controls the hologram length (HL). The molecular holograms created by the previous steps were employed along with the biological activity data (pMIC) to develop PLS HQSAR models. For the best HQSAR model obtained until this point, the influence of fragment size (3-6, 5-8) was also investigated.

Model evaluation and validation

As two biological properties were employed during chemometric model development, dissimilar criteria were necessary to evaluate their fit and predictive power. Nevertheless, in both cases, the complete data set was randomly split into training and test set compounds for external validation purposes (see Table 2 for KNN and SIMCA and Table 4 for descriptor-based and fragment-based QSAR models). Classification and pattern recognition models were evaluated according to the percentage of correctly classified training set compounds, whereas the equivalent measure for test set compounds was employed to account for the predictive ability of such models.

On the other hand, fragment-based (HQSAR) and descriptor-based (QSAR) models were evaluated by LOO cross-validation (q^2) and had their external predictive power (r^2_{pred}) calculated using the approach described in the literature.^{31,32}

Table 3. Descriptor selected through Fisher's weight

Descriptor ^a	Meaning
BEHe3	highest eigenvalue No. 3 of Burden matrix / weighted by atomic Sanderson electronegativities
BEHm2	highest eigenvalue No. 2 of Burden matrix / weighted by atomic masses
BEHm3	highest eigenvalue No. 3 of Burden matrix / weighted by atomic masses
C-008	atom-centered fragments - CHR2X ^b
EEig05d	Eigenvalue 05 from edge adj. matrix weighted by dipole moments
JGI10	mean topological charge index of order10
MATS8p	moran autocorrelation - lag 8 / weighted by atomic polarizabilities
nCl	number of chlorine atoms
nCs	number of total secondary C (sp3)
PJ12	2D Petitjean shape index

^aDescriptors found in final SIMCA model are highlighted in gray. ^bR stands for carbon; X stands for heteroatom (example, O, N, S).

Table 4. pMIC values of 33 azoles derivatives (training and test sets) against *C. gattii*

Training set		Test set	
Compound	pMIC	Compound	pMIC
Azaconazole	3.97	Hexaconazole	5.20
Bifonazole	3.69	Imazalil	4.27
Bromuconazole	4.67	Itraconazole	6.75
Clotrimazole	4.64	Metconazole	5.50
Difenoconazole	5.61	Miconazole	5.62
Diniconazole	5.21	Penconazole	4.85
Epoxiconazole	5.52	Prochloraz	4.07
Etaconazole	4.61	Propiconazole	4.93
Fembuconazole	4.62	Sulconazole	6.20
Fluquiconazole	4.37	Tebuconazole	4.89
Flusilazole	4.60	Triadimenol	5.00
Flutriafol	3.67	Triticonazole	4.90
		Ketoconazole	6.33
		Climbazole	5.17
		Cyproconazole	5.47
		Diclobutrazole	4.61
		Econazole	4.98
		Fluconazole	3.98
		Fluotrimazole	3.77
		Myclobutanil	4.26
		Prothioconazole	4.03

Results and Discussion

Although the classical SAR of azole antifungal derivatives has long been known,³³ evolution and mutation have produced subtle but crucial modifications in the active site of their molecular target, which results in lower susceptibility profile and appearance of resistant strains.^{5,8,18,34-36} It is noteworthy that although modern drug design methodologies have been successfully applied to fight *Candida spp.* azole-resistance and guide the development of antifungal compounds,^{19,37,38} similar strategies have been scarcely employed in the development of drugs against *Cryptococcus spp.*

One approach to circumvent this dilemma is to evaluate the structural and physicochemical features of azole derivatives that are currently available to fight *Cryptococcus spp.* resistant strains, such as the fluconazole resistant clinical isolate of *C. gattii* employed in this work. This new stand point of view over azole antifungals must be considered in two steps. First, it is essential to underscore which properties should be considered to properly select an azole derivative that is active against a fluconazole-resistant *C. gattii* infection. Secondly, it is important to highlight which features account for the different biological profile observed among azole derivatives against *C. gattii* since this information might shed some light on the chemical requirements to overcome fungal resistance. Most of the time, pharmacists and physicians measure fungal susceptibility profile by diffusion disc assay, so this technique was used to classify a dataset of 33 commercial azole antifungals as active or inactive against *C. gattii*.

Due to the lack of standardized microbiologic protocols to assay most azole derivatives included in this study, adaptations to M44-A2 protocol were required.²⁰ Preliminary assays were carried out with discs containing 8.2 mmol L⁻¹ of each compound, a concentration equivalent to that proposed for evaluating the susceptibility of *Candida albicans* (ATCC 90028) against fluconazole. Next, twice (16.4 mmol L⁻¹) and half (4.1 mmol L⁻¹) of the standard concentration were probed (Table 1). Although diffusion disk protocols do not allow us to evaluate the compound potency, the results clearly highlight that some azole derivatives are ineffective against *C. gattii* in all concentrations assessed (example, fluconazole, fluotrimazole and flutriafol). On the other hand, many compounds have inhibition halos greater than the average (example, ketoconazole, cyproconazole, difenoconazole, diniconazole, hexaconazole and so forth). Considering the average inhibition halo as criteria (Table 1), it is possible to split the azole derivatives into either inactive or active against *C. gattii*. Regardless the disc concentration, the compound classification remains the same for all but 3 compounds (highlighted in Table 1), so it was decided to use the results from the lower concentration disc to classify 18 compounds as active (54.5%) and 15 as inactive (45.5%) against *C. gattii* (Table 2).

In order to evaluate if the chemical information provided by this study would be restricted to the compounds analyzed or could be generalized to congeneric azole derivatives, not used for model calibration, the initial dataset was randomly split into training set (13 active and 11 inactive compounds) and test set (5 active and 4 inactive compounds) (Table 2).

Although DRAGON 5.1 software provides more than 2400 topological descriptors upon which chemometric tools could be applied, a limited number of those indeed is related to the biological activity of azole derivatives against *C. gattii*. In order to select those descriptors that have any correlation to fungal susceptibility assay, Fisher's weight was employed to select descriptors that individually differentiate active from inactive compounds. The 10 descriptors (Table 3) selected by this strategy were gathered, autoscaled and used for KNN and SIMCA model developments.

KNN approach ascribes to all descriptors similar importance to compound classification and is highly influenced by the number of neighbors. In order to avoid model instability as well the loss of precision caused by either smaller or greater than optimal number of neighbors, LOO_{cv} was used to select the number of neighbors that best fit the data. Accordingly, the KNN model built with 10 neighbors correctly classified 69% of active and 73% of inactive training set compounds. Similar accuracy degree was observed for the test set compounds (60% of the active and 75% of inactive compounds correctly classified). In order to improve the quality of the statistical model and avoid problems associated with highly correlated descriptors (example, BEHm3 *vs.* MATS8p ($r = 0.72$) or BEHm3 *vs.* BEHe3 ($r = 0.79$)), it was resorted to chemometric approaches that rely on PCs, such as SIMCA. This approach not only reduces the dimensionality of the models and avoids correlation issues,³⁹⁻⁴¹ but also allows a compound to be classified as "unknown" in case it lies in a chemical space outside the applicability domain of the model.^{42,43} In addition, SIMCA models can be optimized by the iterative exclusion of descriptors that have minor contribution to class separation (discriminating power) or that display little influence on the model (modeling power).⁴⁴ Despite the fact that the initial model presents good statistical results (83% of active and 81.3% of inactive compounds correctly classified with 4 PCs), the exclusion of 3 descriptors lead to a less complex SIMCA model, which correctly classifies 85% to active and 100% of inactive training set compounds, using 3 PCs for each classes. Although, a somewhat lower classification power was observed for the test set compounds of best model (60% of the active and 75% of inactive compounds correctly predicted), this result is still better than the initial model (33% of active and 45% of inactive compounds) (Figure 2).

A model is as good as the data it is built from, so it is not surprising that the disc diffusion assays provided models with limited classification/predictive power. In fact, it has been already pointed out that biological results from microdilution in broth assays are more reliable than disc

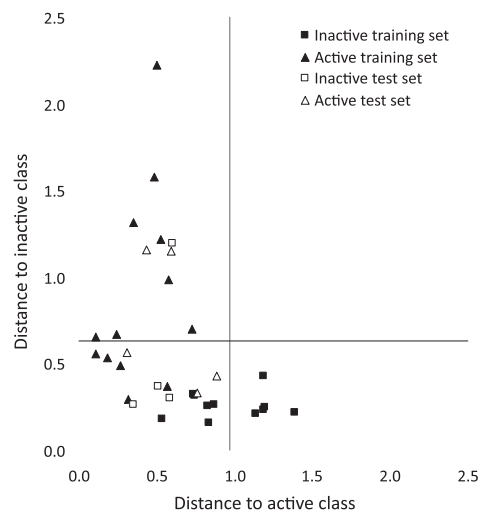


Figure 2. Interclass distance for active and inactive compounds of the training and test sets according to final SIMCA model.

diffusion assays.⁴⁵⁻⁴⁷ The limitation of disc diffusion assays can be partially a consequence of compound diffusion rate in agar. In fact, a potent compound with low diffusion rate would produce a small inhibition halo and thus would be considered "inactive". This subject deserves some thought if triticonazole is taken into consideration. According to disc diffusion assay, this compound is inactive (19.3 mm inhibition halo), but microdilution in broth assay reveals that its MIC ($12.59 \mu\text{mol L}^{-1}$) is comparable to the most potent compounds. Interestingly, the best SIMCA model "erroneously" classifies this compound as active.

Despite moderate predictive capability, the best SIMCA model can shed some light on the structure-activity relationship of azole derivatives that are active against fluconazole-resistant *C. gattii*. A similar strategy was successfully employed by Mota *et al.*¹⁵ which uncovered the importance of electronegativity (BEHe3) and dipolar moment (JGI4), as well as H-bonding towards the activity of azole derivatives against *M. perniciosa*. As expected, evolutionary pressure, both natural and caused by antifungal-therapy, leads *C. gattii* to have a different susceptibility profile from that observed for *M. perniciosa*. Accordingly, azole derivative SAR must also have been altered. In order to investigate this matter, the descriptors with highest modeling power (BEHe3) and discriminant power (JGI10) (Figure S2) were further analyzed, as described below.

JGI10 is an average index of topological charge which evaluates the charge transference between pairs of atoms that are 10 bonds apart.^{48,49} Most of the inactive molecules have zero value for this descriptor (73%), whereas 50% of the active molecules have greater than zero values. This result suggests that charge dispersion has a large effect towards the antifungal activity.

The descriptor BEHe3 is a Burden eigenvalue descriptor that accounts for the Sanderson atomic electronegativity of atoms separated by 3 bonds.^{50,51} The analysis of BEHe3 values suggests that extreme values are found only in inactive compounds (Figure S3 in the SI section).

This analysis is in good agreement with the information provided by JGI10, once both descriptors highlight that electronic features are essential to the biological property.

Similarly, Mota *et al.*¹⁵ showed that these types of descriptor (JGI4 and BEHe3) are useful for describing the activity profile of azole derivatives against the filamentous fungus *Moniliophthora perniciosa*.

Although promising, the classification models developed so far are not suited to predict the potency of azole derivatives and, therefore, have limited usefulness in the design of novel azole derivatives that would be more active against fluconazole-resistant *C. gattii* strains. In order to overcome this limitation, MIC values, obtained from microdilution in broth assay, were used to build QSAR models.

The pMIC values of training and test set compounds not only are normally distributed across the potency range (Table 4 and Figure S3 in the SI section) and were obtained under standardized conditions, but also account for pharmacokinetic effects that would not be captured in kinetic assays with the purified macromolecular target.

Hologram QSAR modeling

HQSAR technique was used as a first resource because it is not biased by subjective alignment rules required by 3D QSAR approaches, such as CoMFA and CoMSIA, but shows comparable statistical quality to those methods.^{29,50}

Our hypothesis was that fragment-like descriptors (molecular holograms) would be useful to build robust 2D QSAR models. Initially, the influence of fragment distinction over the statistical parameters was assessed using default fragment size (FS) (Table 5). If only atoms, bonds and connections (A/B/C) parameters are taken into consideration, marginal fit ($r^2 = 0.55$) and poor internal consistency ($q^2 = 0.11$) are achieved. Despite the fact that addition of hydrogen (H), chirality (Ch) and donor and acceptor (DA) improves the model fit, no significant increase was observed in the internal consistency (compare model 2, 3 and 4 vs. model 1). Similarly, other combinations of fragment distinction parameters produced no further statistical improvements in q^2 values.

Sometimes, variation of fragment size leads to models with higher statistical values, but in this case, no further improvement was achieved for the q^2 values (Table 6).

These initial results indicate that biological activity of azole derivatives cannot be properly captured by molecular holograms only. This might be a consequence of the fact that HQSAR models do not consider charge properties *per se* (not accounted for in fragment distinction options), whereas the best SIMCA model suggests that electronic features are highly influential for the antifungal activity. Aiming at circumvent this problem, topological descriptors weighted by steric or electrostatic features were employed to build 2D QSAR models.

Classical 2D QSAR

As an initial approach, MLR was used to build models with up to 4 variables, from a set of topological descriptors calculated with DRAGON software. Despite good

Table 5. Influence of fragment distinction over the statistical parameters of HQSAR models, using default fragment size (4-7)

Model	Fragment distinction	q^2	r^2	Hologram length	N
1	ABC	0.11	0.55	61	2
2	ABCH	0.25	0.46	61	2
3	ABCHCh	0.25	0.88	307	4
4	ABCHChDA	0.28	0.81	59	4
5	ABH	0.20	0.41	71	2
6	ACH	0.23	0.45	59	2
7	ABCCh	0.21	0.60	61	2
8	ACHCh	0.23	0.48	61	2
9	AHCh	0.22	0.44	71	2
10	ABHCh	0.29	0.61	59	3
11	AHChDA	0.19	0.30	71	1

q^2 : cross-validated correlation coefficient; r^2 : noncross-validated correlation coefficient; N : optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

Table 6. Influence of fragment size over the statistical parameters of the best HQSAR models

Model	Fragment distinction	q^2	r^2	Hologram length	N	Fragment size
12	ABHCh	0.14	0.47	59	2	2-5
13	ABHCh	0.35	0.85	59	4	3-6
14	ABHCh	0.15	0.59	59	3	5-8

q^2 : cross-validated correlation coefficient; r^2 : noncross-validated correlation coefficient; N : optimal number of components. Fragment distinction (FD): A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

statistical parameters were achieved ($r^2 = 0.78$, $q^2 = 0.74$), these models have poor predictive ability ($r^2_{\text{pred}} = 0.1$). Among the reasons that might explain this result is: (i) collinearity among descriptors, which results in unstable regression models; (ii) inadequate description of chemical-biological space by 2D descriptors. As pointed out before, no structural data are available for the macromolecular target of azole antifungals, rendering the development of 3D QSAR model subjective. On the other hand, the collinearity issue can be simply solved by using PLS.⁵¹ This statistical tool has the additional benefit of reducing the risk of building chance correlation QSAR models, once just a few descriptors (PCs) are employed to build the models, instead of the 428 descriptors available at first.

Accordingly, 36 descriptors found in the best 10 MLR models ($q^2 > 0.64$) were gathered, autoscaled and used for further PLS analysis, as available in the Pirouette 4.0 software. (Table S2 in the SI section).

The initial QSAR model shows statistical results ($r^2 = 0.88$ and $q^2 = 0.69$, 3PCs) similar to those of preliminary MLR models, with improved but still low predictive ability ($r^2_{\text{pred}} = 0.45$). Aiming at producing statistically sound QSAR models, an iterative exclusion of descriptors with low leverage towards the regression vector was carried out. This strategy afforded a significant improvement of the statistical quality ($r^2 = 0.95$, $q^2 = 0.86$, 3 PCs) (Table 7).

It is important to note though that high q^2 values do not guarantee that the QSAR model has any predictive

Table 7. Descriptors contained in the final 2D QSAR model

Variable	Description
Me	mean atomic Sanderson electronegativity (scaled on carbon atom)
Ms	mean electrotopological state
nDB	number of double bonds
MAXDN	maximal electrotopological negative variation
PW3	path/walk 3 - Randic shape index
T(F.F)	sum of topological distances between F.F
X2v	valence connectivity index chi-2
X5v	valence connectivity index chi-5
X2Av	average valence connectivity index chi-2
IC3	information content index (neighborhood symmetry of 3-order)
ATS6p	broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities
MATS7m	Moran autocorrelation - lag 7 / weighted by atomic masses
MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses
MATS2e	Moran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities
GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses
GATS1e	Geary autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities
GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities
ESpm03u	spectral moment 03 from edge adj. Matrix
BELe8	lowest eigenvalue No. 8 of burden matrix / weighted by atomic sanderson electronegativities
GGI1	Topological charge index of order 1
SEigZ	eigenvalue sum from Z weighted distance matrix (Barysz matrix)
nArX	number of halogen linked to aromatic ring
C-003	atom-centred fragments - CHR3 ^a
C-040	atom-centred fragments - R-C(=X)-X / R-C#X / X=C=X ^{a,b}

^aR stands for any groups linked to the central carbon atom; ^b=: stands for double bond; # stands for the triple bond; X stands for heteroatom (i.e., O, N, S).

ability.⁵² Thus, the predictive power was assessed by means of external validation protocol using test set compounds (Table 4). The good agreement between experimental and predicted values (residues below 0.65 log units) indicates the high robustness and reliability of descriptors used to build this QSAR model (Figure 3 and Table 8), thus suggesting that this model might be useful to understand the structural and physicochemical requirements for inhibition of fluconazole-resistant *C. gattii* growth and, in the future, guide the identification of novel azole derivatives that would be more active.

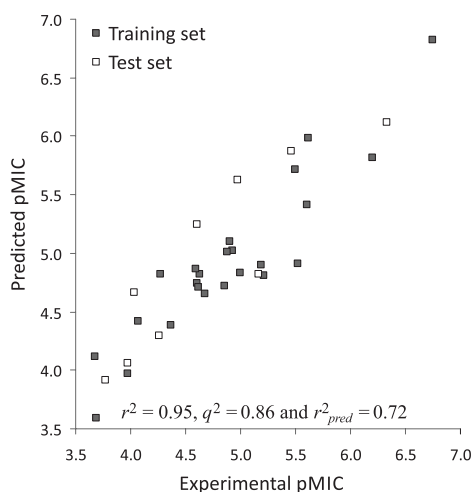


Figure 3. Plot of experimental and predicted pMIC values according to the classical 2D QSAR model.

Table 8. Experimental and predicted pMIC values of the test set compounds for 2D QSAR based on topological descriptors models

Name	Experimental pMIC	Descriptor-based QSAR	
		Predicted pMIC	Residue
Climbazole	5.17	4.81	-0.35
Cyproconazole	5.47	5.87	0.40
Diclobutrazole	4.61	5.25	0.63
Econazole	4.98	5.62	0.64
Fluconazole	3.98	4.05	0.07
Fluotrimazole	3.77	3.91	0.14
Ketoconazole	6.33	6.11	-0.21
Myclobutanil	4.26	4.29	0.04
Prothioconazole	4.03	4.66	0.63

In order to achieve this goal, descriptors with high leverage towards the regression vectors were selected for further analysis (Figure 4).

It is interesting to note that most of the descriptors with high vector regression values describe electrostatic features (example, GGI1, ESPm03u, X2v, BELe8, ATS6p, X2Av, GATS4p, GATS1e and MATS2e), thus emphasizing the influence of electronic features to the antifungal activity.

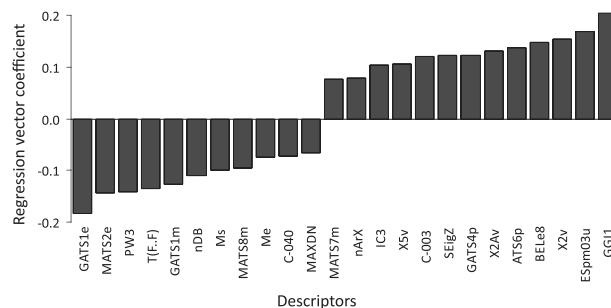


Figure 4. Coefficient of vector regression for descriptors of final 2D QSAR model based on topological descriptors.

GGI1, which is inversely correlated to global charge transfer in the molecule, suggests that reducing charge spread might improve the potency of azole derivatives.^{53,54} In fact, high GGI1 values are found in potent azole derivatives (difenoconazole (GGI1 = 7.0; MIC = 2.46 $\mu\text{mol L}^{-1}$), ketoconazole (8.5; MIC = 0.47 $\mu\text{mol L}^{-1}$) and itraconazole (10.5; MIC = 0.18 $\mu\text{mol L}^{-1}$)) whereas weak derivatives follow the opposite trend (e.g., imazalil (GGI1 = 3.5; MIC = 53.84 $\mu\text{mol L}^{-1}$) and bifonazole (GGI1 = 2.5; MIC = 206.19 $\mu\text{mol L}^{-1}$)).

GATS1e and MATS2e account for the difference on the atomic electronegativity at the topological distance of 1 and 2 bonds, respectively. Although these descriptors point out to the same characteristic, at similar topological distance, they are mathematically uncorrelated ($r = 0.05$). The MATS descriptors provide global information whereas GATS descriptors are sensitive to differences in the atomic neighborhood (example, branching).^{55,56} Nevertheless, both descriptors show that low electronegativity difference between atoms one or two bonds apart increases potency. Taken together, the most influential descriptors underscore that modulating electronic features of azole derivatives is the most promising strategy to circumvent resistance issues in fluconazole-resistant *C. gattii* strains.

Conclusion

The chemometric models presented in this work take advantage of well-established ligand-based tools to highlight chemical features that should be modulated if *Cryptococcus* spp resistance to azole derivatives is to be overcome. First, physicochemical and structural changes that play a major role to azole derivatives antifungal activity were investigated by SIMCA models. This preliminary study suffers from the limitations imposed by disc diffusion assays. Nevertheless, it highlights that electronic features (accounted by BEH3 and JGI10) could be useful to explain azole derivative activity profile against fluconazole resistant *C. gattii*.

Furthermore, fragment-based QSAR (HQSAR) studies proved inappropriate to describe the structure-activity relationship of azole derivatives ($r^2 = 0.85$, $q^2 = 0.35$, $r^2_{pred} = 0.38$), whereas topological descriptors weighted by electronic and steric properties proved suitable to develop high quality 2D QSAR models ($r^2 = 0.95$, $q^2 = 0.86$, $r^2_{pred} = 0.72$). The analysis of the best QSAR model supports the hypothesis that electronic features, described by GGI1, MATS2 and GATS1, should be fine-tuned in order to develop congeneric molecules with improved potency against fluconazole-resistant *C. gattii* strains.

Supplementary Information

Supplementary material includes the complete set of descriptor employed for PLS model development (Figures S1-S3, Tables S1 and S2) and is available free of charge at <http://jbcs.sbq.org.br> as a PDF file.

Acknowledgments

We gratefully acknowledge financial support from the Bahia State Research Foundation (FAPESB-BOL0160/2009) and the Brazilian National Council for Scientific and Technological Development (CNPq 57.3767/2008-4).

References

- Kidd, S. E.; Hagen, F.; Tschärke, R. L.; Huynh, M.; Bartlett, K. H.; Fyfe, M.; Macdougall, L.; Boekhout, T.; Kwon-Chung, K. J.; Meyer, W.; *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17258.
- Park, B. J.; Wannemuehler, K. A.; Marston, B. J.; Govender, N.; Pappas, P. G.; Chiller, T. M.; *AIDS* **2009**, *23*, 525.
- <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5928a1.htm> accessed in April 2013.
- Friese, G.; Discher, T.; Fussle, R.; Schmalreck, A.; Lohmeyer, J.; *AIDS* **2001**, *15*, 2344.
- Khan, Z. U.; Randhawa, H. S.; Kowshik, T.; Chowdhary, A.; Chandu, R.; *J. Antimicrob. Chemother.* **2007**, *60*, 312.
- Chong, H. S.; Dagg, R.; Malik, R.; Chen, S.; Carter, D.; *J. Clin. Microbiol.* **2010**, *48*, 4115.
- Yamazumi, T.; Pfaller, M. A.; Messer, S. A.; Houston, A. K.; Boyken, L.; Hollis, R. J.; Furuta, I.; Jones, R. N.; *J. Clin. Microbiol.* **2003**, *41*, 267.
- Varma, A.; Kwon-Chung, K. J.; *Antimicrob. Agents Chemother.* **2010**, *54*, 2303.
- Sheng, C.; Zhang, W.; Ji, H.; Zhang, M.; Song, Y.; Xu, H.; Zhu, J.; Miao, Z.; Jiang, Q.; Yao, J.; Zhou, Y.; Lu, J.; *J. Med. Chem.* **2006**, *49*, 2512.
- Sheng, C. Q.; Miao, Z. Y.; Ji, H. T.; Yao, J. Z.; Wang, W. Y.; Che, X. Y.; Dong, G. Q.; Lu, J. G.; Guo, W.; Zhang, W. N. A.; *Antimicrob. Agents Chemother.* **2009**, *53*, 3487.
- Sheng, C.; Che, X.; Wang, W.; Wang, S.; Cao, Y.; Yao, J.; Miao, Z.; Zhang, W.; *Chem. Biol. Drug Des.* **2011**, *78*, 309.
- Wang, S.; Jin, G.; Wang, W.; Zhu, L.; Zhang, Y.; Dong, G.; Liu, Y.; Zhuang, C.; Miao, Z.; Yao, J.; Zhang, W.; Sheng, C.; *Eur. J. Med. Chem.* **2012**, *53*, 292.
- Xu, Y.; Sheng, C.; Wang, W.; Che, X.; Cao, Y.; Dong, G.; Wang, S.; Ji, H.; Miao, Z.; Yao, J.; Zhang, W.; *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2942.
- Wang, J.; Urban, L.; *Drug Discovery* **2004**, 73.
- Mota, S. G. R.; Barros, T. F.; Castilho, M. S.; *J. Braz. Chem. Soc.* **2010**, *21*, 510.
- Leroux, P.; Walker, A. S.; *Pestic. Manag. Sci.* **2011**, *67*, 44.
- Diaz-Guerra, T. M.; Mellado, E.; Cuenca-Estrella, M.; Rodriguez-Tudela, J. L.; *Antimicrob. Agents Chemother.* **2003**, *47*, 1120.
- Morio, F.; Loge, C.; Besse, B.; Hennequin, C.; Le Pape, P.; *Diagn. Microbiol. Infect. Dis.* **2010**, *66*, 373.
- Mota, S. G. R.; Barros, T. F.; Castilho, M. S.; *J. Braz. Chem. Soc.* **2009**, *20*, 451.
- Clinical and Laboratory Standards Institute (CLSI); *Method for Antifungal Disk Diffusion Susceptibility Testing of Yeasts*; CLSI: USA, 2009, p. 29.
- Clinical and Laboratory Standards Institute. *Reference method for broth dilution antifungal susceptibility testing of yeasts*. M27-A2, CLSI: USA, 2005, p. 22.
- Cavalcanti, A. R. O.; Soares Leite, E.; Neto, B. B.; Ferreira, R.; *Origins Life Evol. Biosphere* **2004**, *34*, 407.
- Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; *Data Handling Sci. Tech.* **2003**, *23*, 141.
- Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; *Anal. Chim. Acta* **2004**, *515*, 199.
- Todeschini, R.; Consonni, V.; Maiocchi, A.; *Chemom. Intell. Lab. Syst.* **1999**, *46*, 13.
- Molfetta, F. A.; Bruni, A. T.; Honório, K. M.; da Silva, A. B. F.; *Eur. J. Med. Chem.* **2005**, *40*, 329.
- De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. L.; *Chemom. Intell. Lab. Syst.* **2000**, *50*, 1.
- Castilho, M. S.; Guido, R. V.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2007**, *15*, 6242.
- Castilho, M. S.; Postigo, M. P.; de Paula, C. B.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2006**, *14*, 516.
- Guido, R.; Castilho, M.; Mota, S.; Oliva, G.; Andricopulo, A.; *QSAR Comb. Sci.* **2008**, *27*, 768.
- Schuurmann, G.; Ebert, R. U.; Chen, J.; Wang, B.; Kuhne, R.; *J. Chem. Inf. Model.* **2008**, *48*, 2140.
- Doddareddy, M. R.; Cho, Y. S.; Koh, H. Y.; Pae, A. N.; *Bioorg. Med. Chem.* **2004**, *12*, 3977.

33. Bell, A. S.; *Comprehensive Medicinal Chemistry II*; Elsevier Science: Oxford, UK, 2007.
34. Marichal, P.; Koymans, L.; Willemsens, S.; Bellens, D.; Verhasselt, P.; Luyten, W.; Borgers, M.; Ramaekers, F. C. S.; Odds, F. C.; Vanden Bossche, H.; *Microbiol.* **1999**, *145*, 2701.
35. Sanglard, D.; Ischer, F.; Koymans, L.; Bille, J.; *Antimicrob. Agents Chemother.* **1998**, *42*, 241.
36. Kontoyiannis, D. P.; Lewis, R. E.; *Med. Micol.* **2002**, *359*, 1135.
37. Duchowicz, P. R.; Vitale, M. G.; Castro, E. A.; Fernandez, M.; Caballero, J.; *Bioorg. Med. Chem.* **2007**, *15*, 2680.
38. Katritzky, A. R.; Slavov, S. H.; Dobchev, D. A.; Karelson, M.; *Bioorg. Med. Chem.* **2008**, *16*, 7055.
39. Jolliffe, I.; *Principal Component Analysis*, 2nd ed.; Springer: New York, USA, 2002.
40. Tan, C.; Qin, X.; Li, M.; *Vib. Spectrosc.* **2009**, *51*, 276.
41. Flåten, G. R.; Grung, B.; Kvalheim, O. M.; *Chemom. Intell. Lab. Syst.* **2004**, *72*, 101.
42. Wold, S.; *Pattern Recognit.* **1976**, *8*, 127.
43. Daszykowski, M.; Kaczmarek, K.; Stanimirova, I.; Heyden, Y. V.; Walczak, B.; *Chemom. Intell. Lab. Syst.* **2007**, *87*, 95.
44. Pirouette, Infometrix; *Multivariate Data Analysis*; Woodinville, USA, 2008.
45. Scorzoni, L.; Benaducci, T.; Almeida, A. M. F.; Silva, D. H. S.; Bolzani, V. S.; Mendes-Giannini, M. J. S.; *J. Basic Applied Sci.* **2007**, *28*, 25.
46. Mendez, C. C.; Serrano, M. C.; Valverde, A.; Peman, J.; Almeida, C.; Martin-Mazuelos, E.; *Med. Mycol.* **2008**, *46*, 119.
47. Kiraz, N.; Dag, I.; Oz, Y.; Yamac, M.; Kiremitci, A.; Kasifoglu, N.; *J. Microbiol. Methods* **2010**, *82*, 136.
48. Lewis, D. R.; *Tripes Technical Notes*; St. Louis, USA, 1997.
49. Weber, K. C.; da Silva, A. B. F.; *Eur. J. Med. Chem.* **2008**, *43*, 364.
50. Burden, F. R.; *Quant. Struct.-Act. Rel.* **1997**, *16*, 309.
51. Gonzalez, M. P.; Teran, C.; Teijeira, M.; Besada, P.; Gonzalez-Moa, M.; *J. Bioorg. Med. Chem. Lett.* **2005**, *15*, 3491.
52. Golbraikh, A.; Tropsha, A.; *J. Mol. Graphics Modell.* **2002**, *20*, 269.
53. Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R.; *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520.
54. Galvez, J.; Garcia-Domenech, R.; de Julian-Ortiz, J. V.; Soler, R.; *J. Chem. Inf. Comp. Sci.* **1995**, *35*, 272.
55. Geary, R. C.; *The Incorporated Statistician* **1954**, *5*, 115.
56. Moran, P. A.; *Biometrika* **1950**, *37*, 17.

Submitted: February 8, 2013

Published online: May 21, 2013

Supplementary Information

2D Chemometric Studies of a Series of Azole Derivatives Active against Fluconazole-Resistant *Cryptococcus gattii*

Humberto F. Freitas,^a Tania F. Barros^a and Marcelo S. Castilho^{*,a,b}

^aFaculdade de Farmácia, Universidade Federal da Bahia, 40170-290 Salvador-BA, Brazil

^bInstituto Nacional de Ciência e Tecnologia em Biologia Estrutural e Bioimagem, Universidade Federal do Rio de Janeiro (CCS/UFRJ), 21944-970 Rio de Janeiro-RJ, Brazil

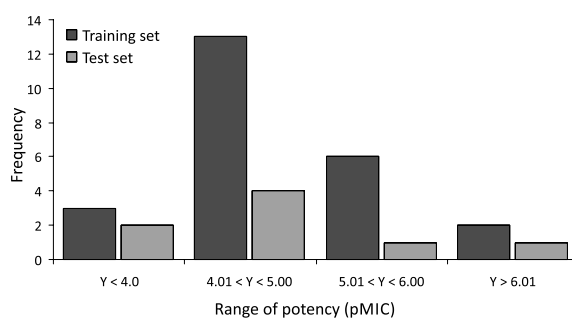


Figure S1. Frequency histogram of compounds of the training and test sets in relation to range of potency (pMIC).

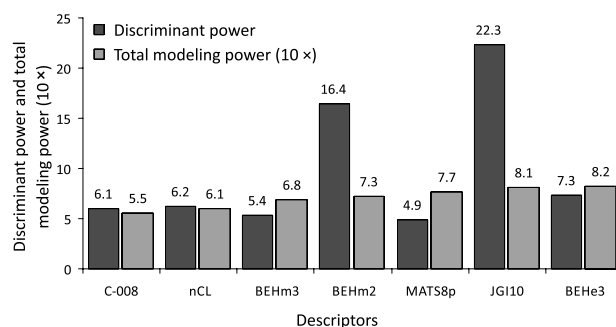


Figure S2. Coefficients of discriminant power (black bar) and total modeling power (10^x) (gray bar) to the descriptors present in the final SIMCA model.

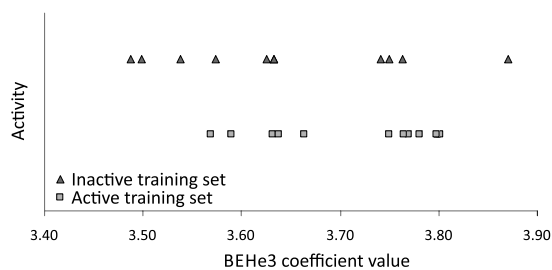


Figure S3. Profile of activity coefficients vs. BEHe3 for compounds of the training and test sets.

Table S1. Chemical structure of the azole derivatives used in chemometric studies

Name	Structure	Name	Structure
Azaconazole		Bifonazole	
Bromuconazole		Ketoconazole	
Climbazole		Clotrimazole	
Ciproconazole		Diclobutrazole	
Difenoconazole		Diniconazole	
Econazole		Epoiconazole	
Etaconazole		Fenbuconazol	
Fluconazole		Fluotrimazol	

Table S1. continuation

Name	Structure	Name	Structure
Fluquinconazole		Flusilazol	
Flutriafol		Hexaconazol	
Imazalil		Itraconazole	
Metconazole		Miconazole	
Myclobutanil		Penconazole	
Prochloraz		Propiconazole	
Prothioconazole		Sulconazole	

Table S1. continuation

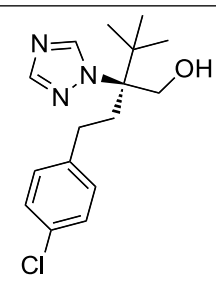
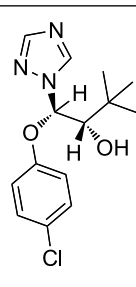
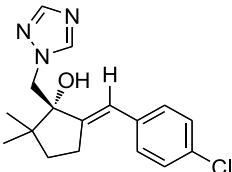
Name	Structure	Name	Structure
Tebuconazole		Triadimenol	
Triticonazole			

Table S2. All the descriptors pre-selected by RLM model using MOBYDIGS 1.0. The descriptors highlighted in gray were used in final 2D QSAR model

Descriptor	Meaning
Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)
Ms	mean electrotopological state
nDB	number of double bonds
J	Balaban distance connectivity index
MAXDN	maximal electrotopological negative variation
PW3	path/walk 3 - Randic shape index
T(F.F)	sum of topological distances between F.F
X2v	valence connectivity index chi-2
X5v	valence connectivity index chi-5
X2Av	average valence connectivity index chi-2
IC3	information content index (neighborhood symmetry of 3-order)
SIC2	structural information content (neighborhood symmetry of 2-order)
ATS5e	Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by Sanderson electronegativities
ATS6p	Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities
MATS7m	Moran autocorrelation - lag 7 / weighted by atomic masses
MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses
MATS2e	Moran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities
GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses
GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses
GATS1e	Geary autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities
GATS5e	Geary autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities
GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities
GATS6p	Geary autocorrelation - lag 6 / weighted by atomic polarizabilities
GATS3v	Geary autocorrelation - lag 6 / weighted by
ESpm03u	spectral moment 03 from edge adj. Matrix
BELe8	lowest eigenvalue n. 8 of burden matrix / weighted by atomic sanderson electronegativities

Table S2. continuation

Descriptor	Meaning
GGI1	topological charge index of order 1
SEigZ	eigenvalue sum from Z weighted distance matrix (Barysz matrix)
nArX	number of halogen linked to aromatic ring
C-003	atom-centred fragments - CHR3 ^a
C-040	atom-centred fragments - R-C(=X)–X / R-C#X / X=C=X ^{b,c}
H-052	atom-centred fragments - H attached to C ⁰ (sp ³) with 1X attached to next C ^d
N-072	atom-centred fragments - RCO–N< / >N–X=X
nCONN	number of urea (–thio) derivatives
nN=C–N<	number of number of amidine derivatives
nCXr	number of halogens on ring C(sp ³)

^aR stands for any group linked through carbon; ^bX stands for any electronegative atom (O, N, S, P, Se, halogens); ^c# stands for triple bond; ^dthe superscript number represents the formal oxidation number of the atoms (example, C⁰).