

## 2D QSAR Studies on a Series of Bifonazole Derivatives with Antifungal Activity

Sabrina G. R. Mota,<sup>a</sup> Tânia F. Barros<sup>b</sup> and Marcelo S. Castilho<sup>\*,a</sup>

<sup>a</sup>Laboratório de Bioinformática e Modelagem Molecular and <sup>b</sup>Laboratório de Microbiologia Clínica, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290 Salvador-BA, Brazil

*Candida albicans* (CA) é considerado como o principal patógeno oportunista em pacientes imunossuprimidos. A maior parte dos fármacos disponíveis para o tratamento de cepas resistentes são altamente tóxicos ou ineficazes. Uma forma de amenizar esse cenário seria através de modificações na estrutura de derivados de azóis que resultassem no aumento da potência e seletividade. Visando esclarecer quais propriedades químicas e estruturais são importantes para atividade antifúngica de derivados de azol, estudos de QSAR 2D clássico e holograma QSAR (HQSAR) foram realizados para um conjunto diverso de 52 derivados de bifonazol com atividade antifúngica. Os descritores topológicos utilizados nos estudos de QSAR 2D clássico originaram modelos com baixa consistência interna ( $r^2 = 0,38$ ,  $q^2 = 0,27$ ) e poder preditivo nulo ( $r^2_{\text{pred}} = -0,6$ ). Por outro lado, a utilização de hologramas moleculares possibilitou a criação de modelos de HQSAR robustos ( $r^2 = 0,92$ ,  $q^2 = 0,65$ ) e com bom poder preditivo ( $r^2_{\text{pred}} = 0,79$ ).

*Candida albicans* (CA) has been identified as the major opportunistic pathogen in immunosuppressed patients. Most of currently available drugs are either highly toxic or becoming ineffective against resistant strains. An approach to overcome this burden relies on azole derivatives with increased potency and selectivity. Aiming at shedding some light on structural and chemical features that are important for the antifungal activity of azole derivatives, classical 2D QSAR and hologram QSAR (HQSAR) studies were performed for a diverse set of 52 bifonazole derivatives with antifungal activity. Topological descriptors, employed in Classical QSAR studies, resulted in models with low correlation ( $r^2 = 0.38$ ,  $q^2 = 0.27$ ) and lack of predictive power ( $r^2_{\text{pred}} = -0.6$ ). On the other hand molecular holograms afforded HQSAR models with good correlation coefficients ( $r^2 = 0.92$ ,  $q^2 = 0.65$ ) and good predictive ability ( $r^2_{\text{pred}} = 0.79$ ).

**Keywords:** 2D QSAR, *Candida albicans*, azole, ergosterol

### Introduction

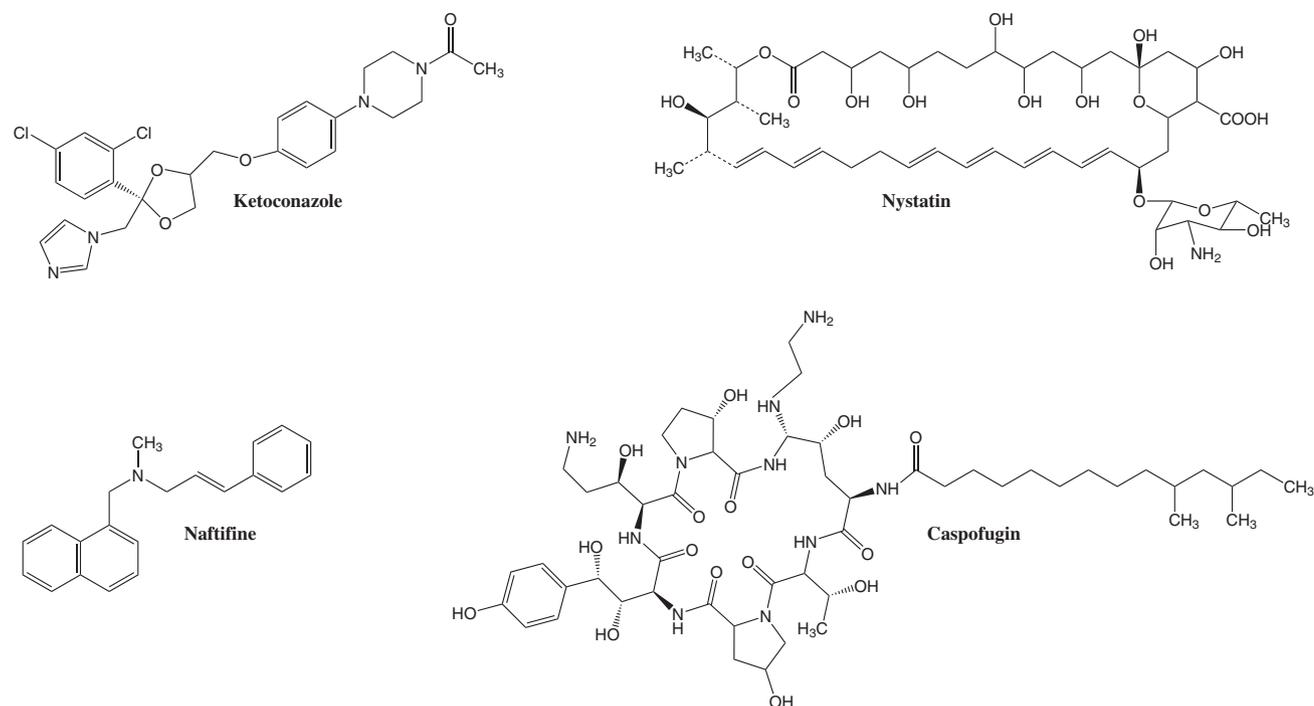
During the past two decades, the incidence of invasive and systemic fungal infections has increased dramatically. This fact can be ascribed to acquired immunodeficiency syndrome (AIDS), greater use of immunosuppressive drugs and chemotherapeutic agents, long term use of corticosteroids, and even the indiscriminate use of antibiotics.<sup>1-5</sup>

Despite *Candida albicans* (CA) has been identified as the major opportunistic pathogen in many fungal infections, the number of infections with other *Candida* species has been increasing.<sup>1</sup> Currently available drugs to treat these infections include azoles (such as fluconazole, ketoconazole, and itraconazole), polyenes (such as amphotericin B and nystatin), echinocandins (such as caspofungin and micafungin), and allylamines (such as naftifine and

terbinafine) (Figure 1). Among them, azole derivatives are the most common drugs used against fungal infections.<sup>2</sup> The azoles are fungistatic compounds that inhibit the C-14 lanosterol demethylase, a key enzyme in sterol biosynthesis pathway leading to the accumulation of C-14 methyl sterols that alter normal membrane function.<sup>6,7</sup>

Nevertheless, some antifungal drugs are either highly toxic (e.g., amphotericin B, AMB) or becoming ineffective against resistant strains that affect mostly hospitalized patients (e.g., flucytosine and azoles).<sup>3</sup> In fact, azole resistance is a major concern in long-course treatment of AIDS patients. The causes of resistance are generally associated with mutations in lanosterol 14 $\alpha$ -demethylase that reduce azole binding and decreased intracellular drug accumulation due to increased expression of efflux pump genes.<sup>4</sup> Moreover, long-term treatments may also cause hepatotoxicity, as azole derivatives can also interact to P450 enzymes from mammalian cytochromes.<sup>8</sup>

\*e-mail: castilho@ufba.br



**Figure 1.** Representative examples of currently available drugs employed in fungal infections treatment.

In order to improve antifungal potency and selectivity, efforts have been made to synthesize new classes of antifungal agents or modify the structures of so far effective azole drugs.<sup>5</sup> Indeed, several three-dimensional quantitative structure-activity relationship studies (3D QSAR) have been reported for different datasets of azole derivatives<sup>2,6,9-12</sup> however, this strategy afforded models with moderate robustness and local predictive ability only. For instance, Di Santo and co-workers<sup>1</sup> report that their CoMFA model was contradicted by the synthesis and evaluation of novel compounds. This result highlighted that the predictive ability of their model was restrict and should not be used to guide the development of novel azole derivatives. One may argue that alignment rules, based on energy minimized or homology model driven conformations contribute significantly to this outcome.

Pharmacophore models have also been developed by means of docking several antifungal agents to 3D homology models of CYP51.<sup>2</sup> Although important information regarding selectivity towards candidosis or aspergillosis was made available, this study did not focus on molecular modifications that might increase potency or overcome resistance. One of the major limitations faced in the previous studies is the lack of accurate 3D structural information from *C. albicans* CYP51. An initial step to overcome this dilemma was the resolution of the crystallographic structure of *Mycobacterium tuberculosis* sterol 14 $\alpha$ -demethylase (MTCYP51) in complex with two azole inhibitors (4-phenylimidazole and fluconazole).<sup>9</sup> Yet, the bacterial source of this enzyme

still poses as a problem for the development of 3D QSAR (ex.: comparative molecular field analysis –CoMFA) and 3D pharmacophore models, as these methods are highly dependent on molecular alignment.<sup>13</sup>

Alternatively, we resorted to 2D QSAR approaches that require no explicit 3D information for the ligands (e.g. putative binding conformations and molecular alignment), employing both classical and fragment-based hologram QSAR (HQSAR) methods.<sup>14,15</sup> HQSAR is an important drug design tool that encodes useful fragment-based information of the molecular structures. Nevertheless, more robust QSAR analysis can be carried out when molecular properties (e.g., physicochemical parameters) are also accounted for.<sup>16</sup> Such approach highlights the complementary nature of classical and HQSAR methods. Besides, to the best of our knowledge, no HQSAR investigation for this class of antifungal compounds has been reported to date. The results of modeling this data set are reported herein.

## Experimental

### Data set

The data set used for the QSAR studies contains 52 derivatives of bifonazole with antifungal activity that were selected from the literature.<sup>17,18</sup> The biological property of this dataset is reported as MIC values, which is the antifungal concentration required to substantially

inhibit organism (*C. albicans*) growth. As most azoles are fungistatics, the breakpoints cannot be clearly defined as residual growth persists for all concentrations above the MIC (National Committee for Clinical Laboratory Standards- CLSI M27-A2 document). For this reason the 90% growth inhibition ( $MIC_{90}$ ) was considered as an accurate measurement of antifungal activity. In order to make comparison with the reference compound and other results from the literature<sup>2,6,9-12</sup> easier, the same relationship employed in previous papers ( $MIC_{90}/MIC_{90bifonazole}$ ), was used to derive 2D QSAR models.

The structures and corresponding  $MIC_{90}/MIC_{90bifonazole}$  values for the whole set of inhibitors are included in Table 1. The  $MIC_{90}/MIC_{90bifonazole}$  values were converted to  $pMIC_{90}/MIC_{90bifonazole}$  ( $-\log MIC_{90}/MIC_{90bifonazole}$ ) values and used as dependent variables in the QSAR analyses.

The chemical structures were drawn in the 2D format and converted to 3D, using Sybyl 7.3 platform (Tripos Inc., St. Louis, USA). All structures were single point optimized using the AM1 semi-empiric method. A hierarchical cluster analysis, carried out with Pirouette 3.11 software (Infometrix, Washington, USA), using the complete linkage clustering method (Euclidean distances) and data autoscaling, guided the division of the complete dataset into training (compounds **1–43**, Table 1) and test (compounds **44–52**, Table 1) sets.

#### Classical QSAR studies

Classical QSAR studies require the calculation and selection of suitable descriptors. The following software was employed for this task: DRAGON 5.4 (Taletto SRL, Milan, Italy) and BUILDQSAR.<sup>19</sup> Briefly, 2D molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation descriptors, Burden eigenvalues indices, among others, were computed using the software DRAGON 5.4 and used as independent variables in the QSAR analyses. A total of 929 molecular descriptors were calculated. Descriptors with zero variance or with poor correlation to biological activity ( $r^2 < 0.10$ ) were discarded. Then, BUILDQSAR software was employed to systematically search for models of up to 4 variables that give rise to multiple linear regression (MLR) models with  $r^2 > 0.81$ . All descriptors present in the MLR models were pooled together, autoscaled and used for the partial least squares (PLS) analysis performed with the PIROUETTE 3.11.

#### HQSAR analysis

The HQSAR modeling analyses, calculations and visualizations were performed using the SYBYL 7.3

package (Tripos Inc., St. Louis, USA) running on Red Hat Enterprise IV workstations. HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size and fragment distinction. The generation of the molecular holograms was carried out using several combinations of the following fragment distinction: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor/acceptor (DA). The influence of fragment size, which controls the minimum and maximum length of fragments to be included in the hologram, was further investigated for the model with better  $q^2$ , by using 5 distinct fragment sizes over the 12 default series of hologram lengths values ranging from 53 to 401 bins. The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using the PLS analysis.

#### QSAR model validation

All QSAR models were investigated using full cross-validated  $r^2$  ( $q^2$ ) PLS. Leave one-out (LOO) cross-validation has been applied to determine the number of principal components that yield optimally predictive models. External validation was performed with a test set of 10 compounds, which were not considered for QSAR model development. The predictive ability of the models is expressed by predictive  $r^2$  values ( $r^2_{pred}$ ), calculated as follows (equation 1):

$$r^2_{pred} = \frac{SD - PRESS}{SD} \quad (1)$$

SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules.<sup>20</sup>

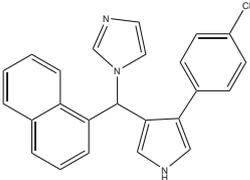
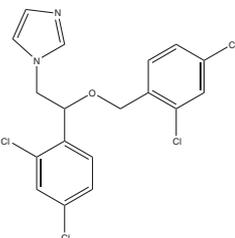
#### Fisher's weight

The Fisher weight is a measure of the distance between two categories, it is given by the difference of the mean values of each category, divided by the sum of the categories variances and can be interpreted as a normalized distance between the Classes.<sup>21</sup> The Fisher's weight is defined as:

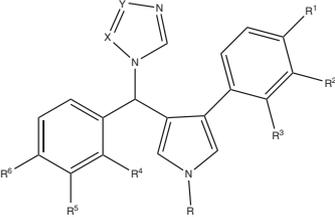
$$F = \frac{(\bar{x}_{p,1} - \bar{x}_{p,2})^2}{S_{p,1} + S_{p,2}} \quad (2)$$

Where  $\bar{x}_{p,1}$ ,  $\bar{x}_{p,2}$ , denote the average values of descriptor  $p$  in class 1 and class 2 respectively, and  $S_{p,1}$ ,  $S_{p,2}$  denote the standard deviation of descriptor  $p$  in class 1 and class 2, respectively.<sup>22</sup> This value was calculated for every

**Table 1.** Chemical structures and corresponding MIC<sub>90</sub>/MIC<sub>90bifonazole</sub> values for a series of bifonazole derivatives with antifungal activity

| Training Set Compounds |   |                 |                |  |                |  |                 |    |  |  |
|------------------------|---|-----------------|----------------|--|----------------|--|-----------------|----|--|--|
| Compd.                 | R   | R <sup>1</sup>  | R <sup>2</sup> | R <sup>3</sup>                                 | R <sup>4</sup> | R <sup>5</sup>   | R <sup>6</sup>  | X  | Y  | MIC <sub>90</sub> /MIC <sub>90bifonazole</sub> |
| 1                      | H   | H               | H              | H  | H              | H  | H               | CH | CH   | 4  |
| 2                      | H   | H               | H              | H  | H              | H  | CH <sub>3</sub> | CH | CH   | 4  |
| 3                      | H   | H               | H              | H  | H              | H  | NO <sub>2</sub> | CH | CH   | 4  |
| 4                      | H   | H               | H              | H  | H              | H  | NH <sub>2</sub> | CH | CH   | 16   |
| 5                      | H   | H               | H              | H  | H              | Cl   | Cl              | CH | CH   | 4  |
| 6                      | H   | H               | H              | Cl   | H              | H  | F               | CH | CH   | 4  |
| 7                      | H   | H               | H              | Cl   | H              | H  | CH <sub>3</sub> | CH | CH   | 4  |
| 8                      | H   | H               | H              | Cl   | H              | Cl   | NH <sub>2</sub> | CH | CH   | 16   |
| 9                      | H   | H               | H              | Cl   | Cl             | H  | Cl              | CH | CH   | 1  |
| 10                     | H   | Cl              | H              | H  | H              | H  | H               | CH | CH   | 1  |
| 11                     | H   | Cl              | H              | H  | H              | H  | Cl              | CH | CH   | 2  |
| 12                     | H   | Cl              | H              | H  | H              | H  | F               | CH | CH   | 2  |
| 13                     | H   | Cl              | H              | H  | H              | H  | CH <sub>3</sub> | CH | CH   | 2  |
| 14                     | H   | Cl              | H              | H  | H              | H  | NH <sub>2</sub> | CH | CH   | 8  |
| 15                     | H   | Cl              | H              | H  | H              | Cl   | Cl              | CH | CH   | 4  |
| 16                     | H   | Cl              | H              | Cl   | H              | H  | H               | CH | CH   | 1  |
| 17                     | H   | Cl              | H              | Cl   | H              | H  | F               | CH | CH   | 2  |
| 18                     | H   | Cl              | H              | Cl   | H              | H  | CH <sub>3</sub> | CH | CH   | 1  |
| 19                     | H   | Cl              | H              | Cl   | H              | H  | NO <sub>2</sub> | CH | CH   | 4  |
| 20                     | H   | Cl              | H              | Cl   | Cl             | H  | Cl              | CH | CH   | 1  |
| 21                     | H   | Cl              | Cl             | H  | H              | H  | H               | CH | CH   | 1  |
| 22                     | H   | Cl              | Cl             | H  | H              | H  | F               | CH | CH   | 1  |
| 23                     | H   | Cl              | Cl             | H  | H              | H  | CH <sub>3</sub> | CH | CH   | 1  |
| 24                     | H   | H               | H              | Cl   | H              | H  | NO <sub>2</sub> | CH | CH   | 16   |
| 25                     | H   | Cl              | Cl             | H  | Cl             | H  | Cl              | CH | CH   | 1  |
| 26                     | H   | Cl              | Cl             | H  | H              | Cl   | Cl              | CH | CH   | 2  |
| 27                     | H   | H               | H              | Cl   | H              | H  | 1-pyrrolyl      | CH | CH   | 4  |
| 28                     | H   | Cl              | H              | H  | H              | H  | 1-pyrrolyl      | CH | CH   | 16   |
| 29                     | H   | Cl              | H              | Cl   | H              | H  | 1-pyrrolyl      | CH | CH   | 16   |
| 30                     | H   | Cl              | Cl             | H  | H              | H  | 1-pyrrolyl      | CH | CH   | 16   |
| 31                     | H   | H               | H              | H  | H              | H  | <i>p</i> -Ph    | CH | CH   | 16   |
| 32                     | H   | NH <sub>2</sub> | H              | H  | Cl             | H  | Cl              | CH | CH   | 64   |
| 33                     | H   | Ph              | H              | H  | Cl             | H  | Cl              | CH | CH   | 2  |
| 34                     | H   | CF <sub>3</sub> | H              | H  | Cl             | H  | Cl              | CH | CH   | 2  |
| 35                     | H   | CN              | H              | H  | Cl             | H  | Cl              | CH | CH   | 8  |
| 36                     | H   | 1-pyrrolyl      | H              | H  | Cl             | H  | Cl              | CH | CH   | 1  |
| 37                     | H   | SMe             | H              | H  | Cl             | H  | Cl              | CH | CH   | 2  |
| 38                     | CH <sub>3</sub>   | Cl              | H              | H  | Cl             | H  | Cl              | CH | CH   | 4  |
| 39                     | H   | Cl              | H              | H  | Cl             | H  | Cl              | N  | CH   | 64   |
| 40                     | H   | Cl              | H              | H  | Cl             | H  | Cl              | CH | N  | 64   |
| Compd.                 | Structure   |                 |                | MIC <sub>90</sub> /MIC <sub>90bifonazole</sub> | Compd.         | Structure  |                 |    | MIC <sub>90</sub> /MIC <sub>90bifonazole</sub> |  |
| 41                     |  |                 |                | 1  | 42             |  |                 |    | 2  |  |

**Table 1.** continuation

| Test Set Compounds  |   |                    |                |                |                |                |                 |    |    |  |
|---|---|--------------------|----------------|----------------|----------------|----------------|-----------------|----|----|--|
|  |   |                    |                |                |                |                |                 |    |    |  |
| Compd.  | R | R <sup>1</sup>     | R <sup>2</sup> | R <sup>3</sup> | R <sup>4</sup> | R <sup>5</sup> | R <sup>6</sup>  | X  | Y  | MIC <sub>90</sub> /MIC <sub>90bifonazole</sub> |
| 43  | H | H                  | H              | H              | H              | H              | Cl              | CH | CH | 8  |
| 44  | H | H                  | H              | H              | H              | H              | F               | CH | CH | 4  |
| 45  | H | H                  | H              | Cl             | H              | H              | H               | CH | CH | 2  |
| 46  | H | H                  | H              | Cl             | H              | H              | Cl              | CH | CH | 1  |
| 47  | H | Cl                 | Cl             | H              | H              | H              | NO <sub>2</sub> | CH | CH | 16   |
| 48  | H | H                  | H              | Cl             | H              | Cl             | Cl              | CH | CH | 2  |
| 49  | H | Cl                 | H              | H              | H              | H              | NO <sub>2</sub> | CH | CH | 4  |
| 50  | H | Cl                 | H              | Cl             | H              | Cl             | Cl              | CH | CH | 2  |
| 51  | H | Cl                 | Cl             | H              | H              | H              | Cl              | CH | CH | 1  |
| 52  | H | <i>tert-t</i> -But | H              | H              | Cl             | H              | Cl              | CH | CH | 64   |

descriptor of amino and nitro substituted compounds, in order to identify those descriptors that best differentiate between them.

## Results and Discussion

### *Chemical and biological data*

Classical QSAR and HQSAR models were derived for a series of bifonazole derivatives with antifungal activity (Table 1). An initial exploratory analysis was carried out by hierarchical cluster analysis (HCA), available in PIROUETTE 3.11 using the complete linkage clustering method (Euclidean distances) and data autoscaling. The cluster analysis shows 8 distinct clusters at 50% similarity, suggesting a reasonable structural diversity of the data set. The generation of consistent statistical models is dependent on the adequacy of the training and test sets. Therefore, molecules from each cluster were randomly assigned to either training set (compounds **1-43**, Table 1) or test set (compounds **44-52**, Table 1) so that structurally diverse molecules, possessing activities of wide range were used for model generation, whereas the 10 inhibitors from test set were employed for external validation.

The *in vitro* MIC<sub>90</sub>/MIC<sub>90bifonazole</sub> values employed in this work were measured under the same experimental conditions,<sup>17,18</sup> a fundamental requirement for QSAR

studies.<sup>15,23</sup> Taken together, these two aspects indicate that this data set is suitable for QSAR modeling.

### *Classical QSAR analysis*

Classical 2D QSAR studies require the calculation of a variety of molecular descriptors (e.g., connectivity indices, 2D autocorrelation descriptors, Burden eigenvalues) that are used as independent variables in QSAR modeling. The DRAGON 5.4 software was used to generate the descriptors for model development. The selection of the descriptors was carried out according to the following criteria. In order to reduce the number of descriptors, BUILDQSAR software was employed to systematically search for MLR models of up to 4 variables with correlation coefficients  $r^2 > 0.81$ . All descriptors present in MLR models (9) were pooled together, autoscaled and then explored using more robust statistical methods such as PCA and PLS, as implemented in the PIROUETTE 3.11 software (Table 2).

The PCA results shows that the first principal component accounts for 46.7% of total variance, while PC2 and PC3 accounts for 25.09% and 5.52% respectively. Additional components have insignificant contribution and were not considered further. The first PC broadly accounts for potency: the less potent inhibitors have positive PC1 values whereas the most potent ones display negative values. These preliminary results prompted us to use the

**Table 2.** Descriptors selected for QSAR model development

| Symbol    | Definition and Description  |
|-----------|---|
| nH        | Number of hydrogen atoms  |
| nN        | Number of nitrogen atoms  |
| SIC5      | Structural information content (neighborhood symmetry of 5-order)     |
| CIC5      | Complementary information content (neighborhood symmetry of 5-order)  |
| BEHm4     | Highest eigen value No. 4 of Burden matrix/ Weighted by atomic masses |
| TPSA(NO)  | Topological polar surface area using N, O polar contributions         |
| TPSA(Tot) | Topological polar surface area using N, O, S, P polar contributions   |
| ALOGP     | Ghose-Crippen octanol-water partition coeff. (logP)                   |
| MLOGP     | Moriguchi octanol-water partition coeff. (logP)                       |

selected descriptors for QSAR modeling studies using PLS. Unfortunately, all models show poor statistical parameters (best model  $r^2 = 0.38$ ,  $q^2 = 0.27$ ) and no predictive power at all ( $r^2_{\text{pred}} = -0.6$ ) (Table 3). This scenario suggests that selected descriptors could not capture chemical features that are important for azole derivatives biological activity. As  $\text{pMIC}_{90}/\text{MIC}_{90\text{bifonazole}}$  values are influenced by both pharmacokinetics and pharmacodynamics events, and 2D descriptors account for whole molecule properties, it is difficult to identify which phenomena description should be improved.

As an alternative to gain further insight into the fragment-based structure-activity relationships for this

**Table 3.** Experimental and predicted activities ( $\text{pMIC}_{90}/\text{MIC}_{90\text{bifonazole}}$ ) with residual values for the test set compounds

| Compd. | Test Set       |           |           |           |           |
|--------|----------------|-----------|-----------|-----------|-----------|
|        | Classical QSAR |           |           | HQSAR     |           |
|        | Experimental   | Predicted | Residual* | Predicted | Residual* |
| 43     | 4.62           | 4.91      | -0.29     | 4.78      | -0.18     |
| 44     | 4.9            | 4.87      | 0.32      | 4.87      | 0.02      |
| 45     | 5.22           | 5.10      | 0.12      | 5.04      | 0.19      |
| 46     | 5.57           | 5.18      | 0.39      | 5.01      | 0.56      |
| 47     | 5.3            | 5.42      | -0.12     | 4.95      | 0.35      |
| 48     | 4.98           | 4.34      | 0.64      | 4.86      | 0.12      |
| 49     | 5.34           | 5.52      | -0.18     | 5.57      | -0.23     |
| 50     | 5.6            | 5.30      | 0.30      | 5.37      | 0.23      |
| 51     | 4.41           | 4.57      | -0.16     | 4.92      | -0.51     |
| 52     | 3.82           | 4.83      | -1.00     | 4.10      | -0.28     |

\* Residual values are calculated from experimental minus predicted values.

series of bifonazole derivatives with antifungal activity, we resorted to another QSAR approach, Hologram QSAR.

### HQSAR analysis

HQSAR relates biological activity to structural fragments. Basically, this analysis involves three main steps: 1. generation of structural fragments for each azole derivative of the training set; 2. the encoding of these fragments into a molecular hologram; 3. the statistical generation of PLS QSAR models.<sup>15</sup>

In our studies, the influence of the three parameters: fragment distinction, fragment size, and hologram length (HL), on the statistical values of our models were investigated. Thus, several combinations of fragment distinction were considered during the QSAR modeling runs using the fragment size default (4-7), as follows: ABC, ABCH, ABCChH, ABCChDAH, ABH, ABDA, ABCDA, ABHDA, ABCHDA, ABHChDA, ABCChDA, and ABCCh. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The statistical results from the PLS analyses for the 42 training set inhibitors are summarized in Table 4.

**Table 4.** HQSAR analyses for various fragment distinctions on the key statistical parameters using fragment size default (4-7)

| Model | Fragment Distinction | $q^2$ | $r^2$ | HL* | PC** |
|-------|----------------------|-------|-------|-----|------|
| 1     | A/B/C                | 0.27  | 0.75  | 59  | 4    |
| 2     | A/B/C/H              | 0.22  | 0.82  | 151 | 6    |
| 3     | A/B/C/Ch/H           | 0.41  | 0.79  | 59  | 5    |
| 4     | A/B/C/Ch/DA/H        | 0.65  | 0.92  | 59  | 6    |
| 5     | A/B/H                | 0.22  | 0.63  | 61  | 3    |
| 6     | A/B/C/Ch             | 0.36  | 0.68  | 59  | 2    |
| 7     | A/B/DA               | 0.22  | 0.65  | 199 | 2    |
| 8     | A/B/C/DA             | 0.11  | 0.66  | 59  | 3    |
| 9     | A/B/H/DA             | 0.22  | 0.72  | 199 | 4    |
| 10    | A/B/C/H/DA           | 0.15  | 0.75  | 151 | 4    |
| 11    | A/B/H/Ch/DA          | 0.29  | 0.75  | 199 | 4    |
| 12    | A/B/C/Ch/DA          | 0.54  | 0.92  | 71  | 6    |

\*HL stands for Hologram length. \*\*PC stands for number of principal components.

According to HQSAR analysis, donor/acceptor atoms (model 8) or hydrogen atoms (model 2) add no information to default fragment distinction (model 1). On the other hand, chirality seems to play an important role for antifungal activity (compare model 6 and 1). Furthermore, a small improvement was achieved by including donor/acceptor

atoms in fragment distinction (model 12). Interestingly, when hydrogen atoms were considered, in addition to chirality and donor/acceptor atoms, a further improvement was observed (model 4). This result is unexpected as previous studies have shown that stereochemical features do not significantly influence the antifungal activity of azole derivatives.<sup>18,24,25</sup>

Nevertheless, lack of stereo selectivity does not necessarily indicate that inhibitors' stereochemistry is irrelevant to ligand binding in the active site of the putative fungal target, otherwise it might be related to other phenomena such as cell uptake or different cellular localization. Hence, the importance of chirality in the best HQSAR model might be a consequence of structural features that are important for azole derivatives interaction into lanosterol 14 $\alpha$ -demethylase binding site. The influence of different fragment sizes, which control the minimum and maximum length of fragments to be included in the hologram, was further investigated as shown in Table 5, but no improvement was achieved.

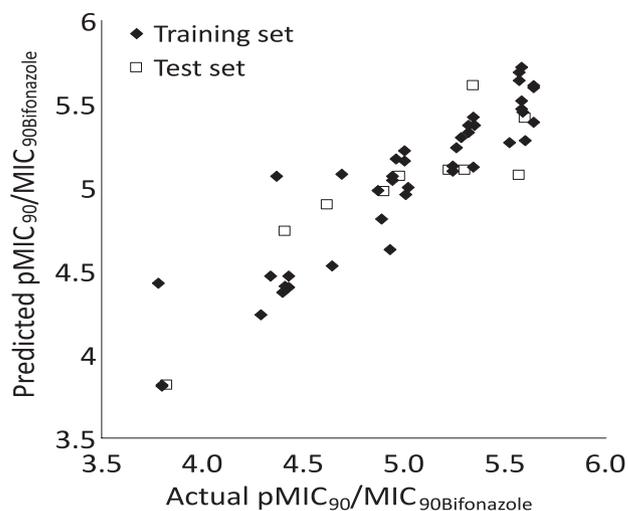
**Table 5.** HQSAR analysis for the influence of different fragment sizes on the statistical parameters (HL stands for Hologram length and PC stands for number of principal components)

| Fragment Distinction | Model | Fragment size | $q^2$ | $r^2$ | HL* | PC** |
|----------------------|-------|---------------|-------|-------|-----|------|
| A/B/C/Ch/DA/H        | 15    | 2-5           | 0.43  | 0.87  | 71  | 6    |
| A/B/C/Ch/DA/H        | 16    | 3-6           | 0.47  | 0.88  | 59  | 5    |
| A/B/C/Ch/DA/H        | 4     | 4-7           | 0.65  | 0.92  | 59  | 6    |
| A/B/C/Ch/DA/H        | 17    | 5-8           | 0.57  | 0.88  | 53  | 5    |
| A/B/C/Ch/DA/H        | 18    | 6-9           | 0.41  | 0.78  | 71  | 3    |

\*\*HL stands for Hologram length. \*\*PC stands for number of principal components.

As the molecular structure encoded within the molecular hologram is directly related to anti-*Candida* activity of training set compounds, the HQSAR model should be able to predict the activity of new related molecules. Thus, the predictive ability of the model was assessed using the same test set compounds employed in the classical 2D QSAR studies (compounds **43-52**, Table 1). The results of the external validation for the best predictive model (4,  $r^2_{\text{pred}} = 0.79$ ) are displayed in Table 3 and the graphic results for the experimental versus predicted for both training and test sets are shown in Figure 2. As can be seen, the predicted values are in good agreement with experimental  $\text{pMIC}_{90}/\text{MIC}_{90\text{Bifonazole}}$  values, deviating by no more than 0.56 log units. Therefore, the model presents good correlative and predictive abilities.

Besides predicting the property value of interest of untested molecules, HQSAR models should also provide



**Figure 2.** Predicted versus actual  $\text{pMIC}_{90}/\text{MIC}_{90\text{Bifonazole}}$  values for the training and test set.

hints about the relationships of different molecular fragments to biological activity.<sup>14,15</sup> HQSAR models can be graphically represented in the form of contribution maps where the color of each molecular fragment reflects the contribution of an atom or a small number of atoms to the activity of the molecule under study. The colors at the red end of the spectrum (e.g., red, red-orange, and orange) reflect poor (or negative) contributions, while colors at the green end (e.g., yellow, green-blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are colored white.

The HQSAR contribution map, Figure 3, shows that compounds *ortho*- (compare **45** and **1**), or *para*-substituted (compare **10** and **1**) at the benzyl ring, next to the pyrrol ring, have good activity against *C. albicans*. Another interesting feature, highlighted in contribution maps, is the opposite effect of nitro and amine moieties to potency (ex.: **3** and **4**). Unfortunately, HQSAR affords no plain explanation for the different contribution of the fragments towards potency. The synergic use classical and HQSAR could circumvent this sort of limitation,<sup>16</sup> but the poor correlation of classical QSAR models prevents this approach. Instead, topological descriptors computed in the early steps of this work were employed to shed some light in this subject. A subset of compounds bearing either amine (**4**, **8** and **14**) or nitro (**3**, **24** and **49**) moieties, at equivalent positions, was investigated in the search for descriptors that could discriminate between them. The assumption being that such descriptor would somehow explain the different inhibitory profile of amine and nitro substituted compounds. In order to accomplish this task, a low dimensional classification rule (Fisher's weight) was employed. As seen in Table 6, two descriptors are important for class discrimination,

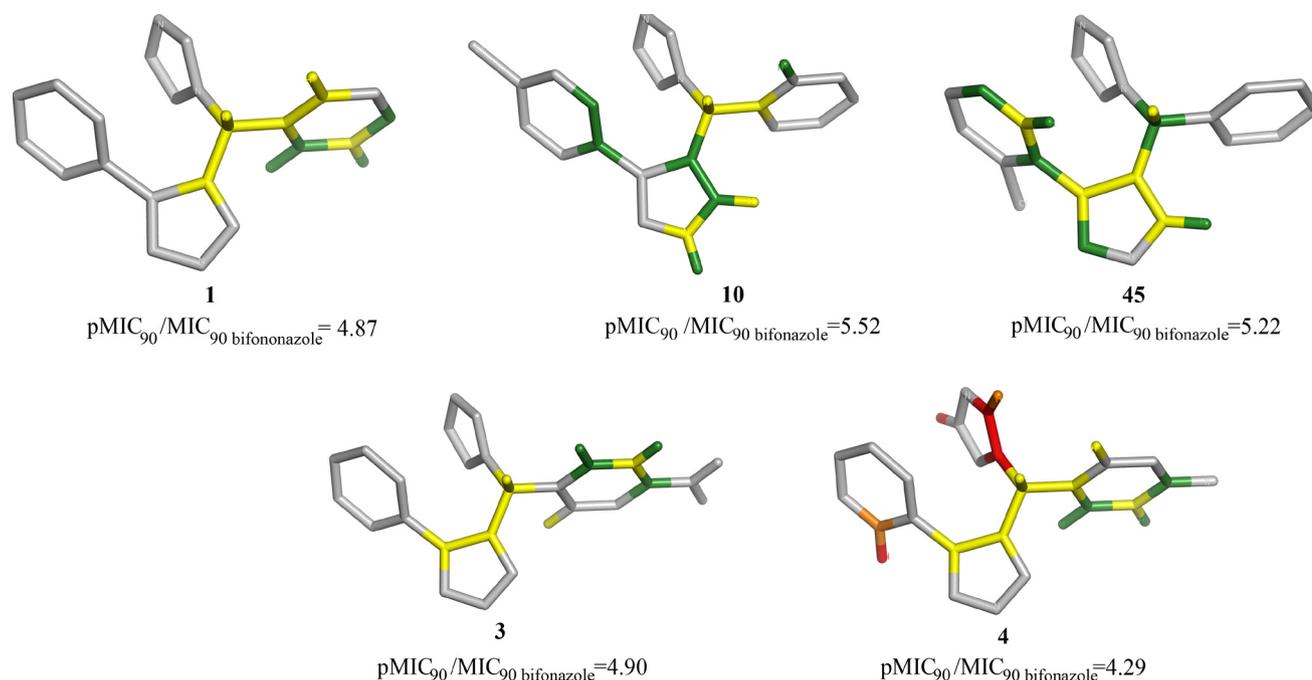


Figure 3. Contribution map according do HQSAR model 4.

Table 6. Comparison of nitro and amine substituted compounds by Fisher's weight

| Compd.        |                 | pMIC <sub>90</sub> | MAXDN  | BEHe1 | MLOGP  | ALOGP  | TPSA(NO) | ZM1V   |
|---------------|-----------------|--------------------|--------|-------|--------|--------|----------|--------|
| 3             | NO <sub>2</sub> | 4.93               | 2.39   | 4.07  | 3.24   | 3.63   | 79.43    | 378.00 |
| 24            | NO <sub>2</sub> | 4.37               | 2.40   | 4.07  | 3.73   | 4.30   | 79.43    | 385.61 |
| 49            | NO <sub>2</sub> | 4.98               | 2.40   | 4.07  | 3.73   | 4.30   | 79.43    | 385.61 |
| 4             | NH <sub>2</sub> | 4.29               | 1.30   | 4.06  | 3.70   | 2.98   | 4.06     | 290.00 |
| 8             | NH <sub>2</sub> | 4.34               | 1.36   | 4.06  | 3.19   | 3.65   | 59.63    | 297.61 |
| 14            | NH <sub>2</sub> | 4.64               | 1.33   | 4.06  | 3.19   | 3.65   | 59.63    | 297.61 |
| Fisher Weight |                 |                    | 698.64 | 0     | 0.1384 | 0.1672 | 0        | 133.89 |

the Zagreb index by valence vertex degrees (ZM1V) that accounts for molecular branching in hydrogen-depleted structures<sup>26</sup> and the maximum negative intrinsic state difference in the molecules (MAXDN), which can be related to the nucleophilicity of the molecules.<sup>27</sup> The greater value of MAXDN suggests that it is the most appropriate descriptor to separate amine substituted compounds from nitro substituted ones.

Therefore, it is tentative to assume that better anti-*Candida* activity of nitro substituted compounds is correlated to selective binding to electrophilic residues in the active site of lanosterol 14 $\alpha$ -demethylase. This hypothesis shed new light on the previous work of Artico and co-workers,<sup>17</sup> which shows that lipophilicity (logP) is responsible for the different potency of nitro and amine substituted compounds. This apparent contradiction can be reasoned if one considers that azole derivatives

potency depends on pharmacokinetic (logP) as well as pharmacodynamic (MAXDN) factors.

## Conclusions

Though classical QSAR models were unable to describe azole derivatives' anti-*Candida* activity, fragment based hologram QSAR succeed in this task. The good correlation between experimental and predicted anti-candida activity for 10 test set compounds and the new insight into nucleophilicity importance for azole potency further highlights the worth of constructed HQSAR models. Moreover, HQSAR contribution maps provide information about the importance of benzyl substitution pattern towards potency. The combined use of these results should be useful to develop selective and more potent azole derivatives.

## Acknowledgements

We gratefully acknowledge financial support from the State of Bahia Research Foundation (FAPESB), the National Council for Scientific and Technological Development (CNPq), Brazil. We also acknowledge Dr. Adriano Defini Andricopulo for kindly providing access to SYBYL 7.3 package of programs.

## References

1. Di Santo, R.; Tafi, A.; Costi, R.; Botta, M.; Artico, M.; Corelli, F.; Forte, M.; Caporuscio, F.; Angiolella, L.; Palamara, A.T.; *J. Med. Chem.* **2005**, *48*, 5140.
2. Sheng, C.; Zhang, W.; Ji, H.; Zhang, M.; Song, Y.; Xu, H.; Zhu, J.; Miao, Z.; Jiang, Q.; Yao, J.; Zhou, Y.; Zhu, J.; Lu, J.; *J. Med. Chem.* **2006**, *49*, 2512.
3. Nam, N.; Sardari, S.; Selecky, M.; Parang, K.; *Bioorg. Med. Chem.* **2004**, *12*, 6255.
4. Xiao, L.; Madison, V.; Chau, A.S.; Loebenberg, D.; Palermo, R.E.; McNicholas, P.M.; *Antimicrob. Agents Chemother.* **2004**, *48*, 568.
5. Matysiak, J.; Niewiadomy, A.; Krajewaska-Kulak, E.; Macik-Niewiadomy, G.; *Il Farmaco* **2003**, *58*, 455.
6. Talele, T.T.; Kulkarni, V.M.; *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 204.
7. Podust, L.M.; Kries, J.P.; Eddine, A.N.; Kim, Y.; Yermalitskaya, L.V.; Kuehne, R.; Oullet, H.; Warriar, T.; Altekko, M.; Lee, J.; Rademann, J.; Oschkinat, H.; Kaufman, S.H.E.; Waterman M.R.; *Antimicrob. Agents Chemother.* **2007**, *51*, 3915.
8. Rupp, B.; Raub, S.; Marian, C.; Holtje, H. D.; *J. Comput. Aided. Mol. Des.* **2005**, *19*, 149.
9. Ji, H.; Zhang, W.; Zhang, M.; Zhou, Y.; Zhang, M.; Zhu, J.; Song, Y.; Lu, J.; Lu, J.; Zhu, J.; *J. Med. Chem.* **2000**, *43*, 2493.
10. Fratev, F.; Benfenati, E.; *J. Chem. Inf. Model.* **2005**, *45*, 634.
11. Talele, T.T.; Kulkarni, S.S.; Kulkarni, V.M.; *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 958.
12. Tafi, A.; Anastassopoulou, J.; Theophanides, T.; Botta, M.; Corelli, F.; Massa, S.; Artico, M.; Costi, R.; Di Santo, R.; Ragno, R.; *J. Med. Chem.* **1996**, *39*, 1227.
13. Lemmen, C.; Lengauer, T.; *J. Comput. Aided Mol. Des.* **2000**, *14*, 215.
14. Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D.; *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119.
15. Castilho, M.S.; Guido, R.V.C.; Andricopulo, A.D.; *Bioorg. Med. Chem.* **2007**, *15*, 6242.
16. Guido, R.V. C.; Castilho, M. S.; Mota, S. R. G.; Oliva, G.; Andricopulo, A. D.; *QSAR Comb. Sci.* **2008**, *27*, 768
17. Artico, M.; Santo, R.; Costi, R.; Massa, S.; Rético, A.; Artico, M.; Apuzzo, G.; Simonetti, G.; Strippoli, J.; *J. Med. Chem.* **1995**, *38*, 4223.
18. Tafi, A.; Costi, R.; Botta, M.; Santo, R.; Corelli, F.; Massa, S.; Ciacci, A.; Manetti, F.; Artico, M.; *J. Med. Chem.* **2002**, *45*, 2720.
19. Oliveira, D.B.; Glaudio, A.C.; *QSAR Comb. Sci.* **2003**, *19*, 599.
20. Doddareddy, M.R.; Cho, Y.S.; Koh, H.Y.; Pae, A.N.; *Bioorg. Med. Chem.* **2004**, *12*, 3977.
21. Cavalcanti, A.R.O.; Leite, E.S.; Neto, B.B.; Ferreira, R.; *Origins Life Evol. B* **2004**, *34*, 407.
22. Wang, K.L.; Wen, Z.N.; Nie, F.S.; Li, M.L.; *Chin. Chem. Lett.* **2005**, *16*, 1133.
23. Farutin, V.; Masterson, L.; Andricopulo, A.D.; Cheng, J.; Riley, R.; Hakimi, R.; Frazer, J.; Cordes, E.H.; *J. Med. Chem.* **1999**, 2422.
24. Botta, M.; Corelli, F.; Gasparrini, F.; Messina, F.; Mugnaini, C.; *J. Org. Chem.* **2000**, *65*, 4736.
25. Cirilli, R.; Costi, R.; Santo, R.; Ferretti, R.; Torre, F.; Angiolella, L.; Micocci, M.; *J. Chromatogr. A* **2002**, 942, 107.
26. Gutman, I.; Ruscic, B.; Trinajstic, N.; Wicox, C. F.; *J. Chem. Phys.* **1975**, *62*, 3399.
27. Gramatica, P.; Corradi, M.; Consonni, V.; *Chemosphere* **2000**, *41*, 763.

Received: July 15, 2008

Web Release Date: February 13, 2009