Article

Hologram- and Descriptor-Based QSAR Studies for a Series of Non-Azoles Derivatives Active Against *C. neoformans*

Lara B. Cunha,[a] Humberto F. Freitas[a] and Marcelo S. Castilho[*,a,b]*

[a] Faculdade de Farmácia, Universidade Federal da Bahia, 40170-290 Salvador-BA, Brazil

[b] Instituto 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

Durante as últimas décadas as infecções fúngicas têm se tornado um crescente problema de saúde, especialmente para pacientes imunocomprometidos. Infelizmente, o padrão-ouro de terapia profilática para tal doença é baseada em derivados azólicos, que são mais fungistáticos do que fungicidas contra *C. neoformans* e causam hepatotoxicidade. Objetivando contornar estes problemas, inibidores não–azólicos de CYP51 foram planejados. Aqui, um abrangente estudo de relação estrutura-atividade foi executado para uma série de 110 moléculas através de um estudo de QSAR baseado em hologramas e descritores moleculares. O melhor modelo de QSAR baseado em descritores ($r^2 = 0,92$, $q^2 = 0,90$, 6 LVs e $r^2_{pred} = 0,86$) sugere que o efeito de ressonância (ESpm08r) desempenha um papel principal para a atividade antifúngica. O modelo de QSAR baseado em hologramas ($r^2 = 0,87$, $q^2 = 0,81$, 6 LVs e $r^2_{pred} = 0,84$) sustenta esta hipótese. Estas percepções obtidas pelas análises integradas dos modelos de QSAR, juntamente com o bom poder preditivo comprovam sua utilidade para futuros esforços em planejamento de fármacos.

Over the last decades fungal infections have become an increasing health problem, especially for immunocompromised patients. Unfortunately, the gold standard prophylactic therapy for such ailment is based on azole derivatives, which are fungistatic rather than fungicidal against *C. neoformans* and cause hepatotoxicity. Aiming at circumvent these problems, non-azole CYP51 inhibitors were designed. Herein a comprehensive structure-activity relationships study was carried out for a dataset of 110 molecules by means of hologram– and descriptor–based QSAR studies. The best descriptor-based QSAR model ($r^2 = 0.92$, $q^2 = 0.90$, 6 LVs and $r^2_{pred} = 0.86$) suggests that resonance effects (ESpm08r) play a major role for antifungal activity. The hologram-based QSAR ($r^2 = 0.87$, $q^2 = 0.81$, 6 LVs and $r^2_{pred} = 0.84$) supports this hypothesis and hints at steric properties that should also contribute to non-azole inhibitors potency. The insights provided by the integrated analysis of QSAR models, along with their good predictive power prove their usefulness to future drug design efforts.

Keywords: antifungal, non-azole, chemometrics, 2D QSAR, HQSAR

Introduction

Over the last decades fungal infections have become an increasing health problem, especially for immunocompromised patients, such as those receiving chemotherapy, organ transplant recipients and HIV positive patients,[1-3] for whom even opportunistic fungi, such as *Cryptococcus* sp present a life-threatens issue. In fact *Cryptococcal meningoencephalitis*, common among HIV patients, has a high mortality rate even when treated with first-line antifungal drugs.[4,5] Currently, the standard prophylactic therapy and long term management of systemic deep seated fungal infections rely on azole derivatives, for instance fluconazole, or combinations thereof.[1-6]

Unfortunately, the extensive use and prolonged therapy with azole antifungal agents have led to severe resistance, which significantly limits their use.[7] Infections caused by *C. neoformans* present an additional complexity as it has inherent resistance to fluconazole.[8,9] Furthermore, fungistatic rather than fungicidal activities and hepatotoxicity pose as serious drawbacks for azole use. Apparently, similarity between human and fungal therapeutic targets makes the pharmacophore requirements

*e-mail: castilho@ufba.br

of azole drugs profoundly correlated to their toxicophore, as the binding of heterocyclic nitrogen atom (N-3 of imidazole and N-4 of triazole) to the heme iron atom in the active site of lanosterol 14α-demethylase (CYP51) enzyme is responsible both for the pharmacological outcome and the hepatic side-effects. In fact, cases of death due to liver failure are ascribed to the coordination of azole drugs to the heme of host cytochrome P450 enzymes, such as CYP3A4,[10,11] whereas their binding towards fungal CYP51 leads to fungal growth arrest following the accumulation of precursor 14α-methylated sterols in the fungi membrane.

Aiming at separating these effects, extensive medicinal chemistry efforts have been carried out.[12-14] Despite novel azole derivatives, for instance posaconazole and voriconazole, present lower hepatotoxicity than first generation azoles, they still cause liver enzyme elevation.[15] On the other hand, inhibitors that explore specific interactions (H-bonding, hydrophobic interaction and so forth) in the active site of fungi CYP51, but do not interact with heme prosthetic group are expected to have antifungal effect only.[16] Taking this hypothesis into consideration homology modeling strategies guided the design and synthesis of diverse non-azole CYP51 inhibitors (Figure 1).[10,16,17]
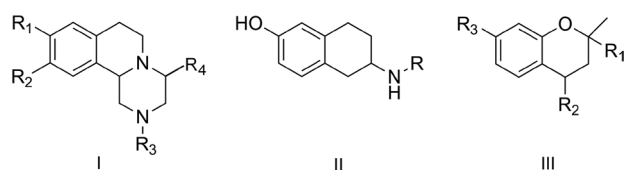

Figure 1. Non-azole antifungals designed as lanosterol 14α-demethylase inhibitors.

Although this strategy relied on subjective structural data and docking studies, compounds that are more potent than fluconazole against *C. neoformans*, have been designed. As these compounds are structurally different from azole drugs, they probably have dissimilar physicochemical requirements for CYP51 inhibition. Nevertheless, as far as we are aware, no effort was made to investigate their structure-activity relationships from a quantitative point of view. The results described herein try to shorten this knowledge-gap by means of hologram-based and descriptor-based QSAR models that do not require structural information from the molecular target and thus are less prone to bias induced by other modeling strategies.

Materials and Methods

Data set

The data set used for the QSAR studies contains 110 derivatives of isoquinolines, chromenes and 2-aminetetralines along with their antifungal activity (Table 1).[2,10,18] The biological property of this dataset is reported as MIC_{80} values, which is the antifungal concentration required to inhibit 80% of *C. neoformans* growth. In order to overcome small data variability among different experiments, an internal standard was employed ($MIC_{80fluconazole}$) so that the biological data used for QSAR models development is based on the ratio ($MIC_{80cpd}/MIC_{80fluconazole}$). These values were converted to $pMIC_{80}$ ($-\log(MIC_{80cpd}/MIC_{80fluconazole})$) and used as dependent variables in the QSAR model development.

The chemical structures were drawn in the two-dimensional (2D) format and converted to 3D using CONCORD standard parameters, as available in "translate molecular file" tool, from Sybyl-X 1.1 plataform (Tripos Inc., St. Louis, USA) and then energy minimized by conjugated gradient using Tripos force field (convergence criteria 0.001 kcal mol^{-1}). Next, MOPAC charges were added to each molecule (AM1 semi-empirical method with the following keywords: 1SCF XYZ ESP NOINTER NSURF = 2 SCALE = 1.4 SCINCR = 0.4 NOMM). This protocol is necessary so that charge-related descriptors can be properly calculated in DRAGON. A unsupervised method HCA (Hierarchical Cluster Analysis), carried out with Pirouette 4.0 software (Infometrix, Washington, USA), using the complete linkage clustering method (Euclidean distances) and data autoscaling, was used to guide to split the complete dataset into training (compounds **1**-**75**) and test (compounds **76**-**110**) sets. (Table 1) Accordingly, at least 1 compound from each cluster (similarity degree = 90%) was assigned to the test set (Figure 1S).
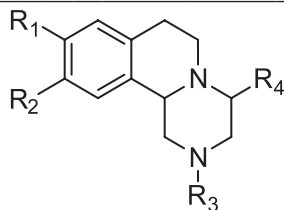
Descriptor-based QSAR approach

About 2,500 2D molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation and physicochemical descriptors and so forth, were computed using the DRAGON 5.5 software (Talette SRL, Milan, Italy) and then pre-selected as follows: descriptors with high inter-correlation ($\geq 97\%$) or those poorly related to the biological property ($r^2 < 0.10$) were discarded. This strategy yielded 215 physicochemical descriptors that were employed to build multiple linear regression models (MLR) with up to 6 descriptors per model using genetic algorithm, as available in MOBYDIGS 1.0 software (Talette SRL, Milan, Italy). The MLR models were evaluated using the following fitting criteria: QUIK rule (0.005), asymptotic Q2 rule (−0.005), redundancy RP rule (0.05) and overfitting RN rule (0).[19,20] Due to the stochastic nature of the genetic algorithm, the

Table 1. Structural scaffold and pMIC$_{80}$ values of training and test set compounds

Training set

N°	R$_1$	R$_2$	R$_3$	pMIC$_{80}$
1	$-CH_2(CH_2)_3CH_3$	$-OH$	$-OH$	-1.39
2	$-CH_2CH_2CH=C(CH_3)_2$	$-OH$	$-OH$	-1.67
3	$-CH_2(CH_2)_3CH_3$	$=NOH$	$-OH$	-1.07
4	$-CH_2(CH_2)_3CH_3$	$=O$	$-OH$	-1.69
5	$-CH_2(CH_2)_7CH_3$	$=O$	$-OH$	-1.91
6	$-CH_2CH(CH_3)_2$	$=O$	$-OH$	-2.02
7	$-CH_2(CH_2)_5CH_3$	$=O$	$-OCH_3$	-1.93
8	$-CH_2CH_2CH=C(CH_3)_2$	$=O$	$-OCH_3$	-1.95
9	$-CH_2(CH_2)_3CH_3$	$-OH$	$-OCH_3$	-1.67
10	$-CH_2CH_2CH=C(CH_3)_2$	$-OH$	$-OCH_3$	-1.65

N°	R$_1$	R$_2$	pMIC$_{80}$
11	$-CH_2(CH_2)_5CH_3$	$-OCH_3$	-1.95
12	$-CH_2CH_2CH=C(CH_3)_2$	$-OCH_3$	-1.98

N°	R$_1$	R$_2$	pMIC$_{80}$	N°	R$_1$	R$_2$	pMIC$_{80}$
13	7–OCH$_3$	$-H$	-0.90	40	6–OCH$_3$	$-(CH_2)_{17}CH_3$	-0.90
14	8–OCH$_3$	$-H$	-0.90	41	7–OCH$_3$	$-(CH_2)_{17}CH_3$	-0.90
15	5–OH	$-H$	-0.90	42	8–OCH$_3$	$-(CH_2)_{17}CH_3$	-0.90
16	6–OH	$-H$	-0.90	43	6–OH	$-(CH_2)_{17}CH_3$	-0.90
17	7–OH	$-H$	-0.90	44	8–OH	$-(CH_2)_{17}CH_3$	-0.90
18	7–OCH$_3$	$-(CH_2)_3CH_3$	-0.90	45	5–OCH$_3$		-0.90
19	8–OCH$_3$	$-(CH_2)_3CH_3$	-0.90	46	6–OCH$_3$		-0.90
20	5–OH	$-(CH_2)_3CH_3$	-0.90	47	8–OCH$_3$		-0.90
21	7–OH	$-(CH_2)_3CH_3$	-0.90	48	6–OH		-0.90
22	5–OCH$_3$	$-(CH_2)_7CH_3$	-0.30	49	8–OH		-0.60
23	7–OCH$_3$	$-(CH_2)_7CH_3$	-0.30	50	5–OCH$_3$		-0.90
24	8–OH	$-(CH_2)_7CH_3$	0.00	51	7–OCH$_3$		-0.90
25	5–OCH$_3$	$-(CH_2)_8CH_3$	0.60	52	6–OH		-0.90

Table 1. continuation

N°	R_1	R_2	$pMIC_{80}$	N°	R_1	R_2	$pMIC_{80}$
26	6–OCH$_3$	–(CH$_2$)$_8$CH$_3$	0.30	53	7–OH	benzyl	−0.90
27	8–OCH$_3$	–(CH$_2$)$_8$CH$_3$	0.30	54	8–OH	benzyl	−0.90
28	6–OH	–(CH$_2$)$_8$CH$_3$	1.20	55	5–OCH$_3$	4-methoxyphenyl	−0.90
29	7–OCH$_3$	–(CH$_2$)$_9$CH$_3$	0.90	56	6–OH	benzyl	−0.90
30	5–OH	–(CH$_2$)$_9$CH$_3$	0.90	57	7–OCH$_3$	4-methoxyphenyl	−0.90
31	7–OH	–(CH$_2$)$_9$CH$_3$	0.30	58	8–OCH$_3$	4-methoxyphenyl	−0.90
32	8–OH	–(CH$_2$)$_9$CH$_3$	0.90	59	5–OH	4-methoxyphenyl	−0.90
33	6–OCH$_3$	–(CH$_2$)$_{11}$CH$_3$	0.30	60	7–OH	4-methoxyphenyl	−0.90
34	8–OCH$_3$	–(CH$_2$)$_{11}$CH$_3$	−0.30	61	8–OH	4-methoxyphenyl	−0.90
35	6–OH	–(CH$_2$)$_{11}$CH$_3$	0.60	62	5–OCH$_3$	4-chlorophenyl	−0.90
36	7–OCH$_3$	–(CH$_2$)$_{13}$CH$_3$	−0.30	63	7–OCH$_3$	4-chlorophenyl	−0.90
37	8–OCH$_3$	–(CH$_2$)$_{13}$CH$_3$	−0.60	64	8–OCH$_3$	4-chlorophenyl	−0.90
38	5–OH	–(CH$_2$)$_{13}$CH$_3$	0.00	65	7–OH	4-chlorophenyl	−0.30
39	7–OH	–(CH$_2$)$_{13}$CH$_3$	0.00	66	8–OH	4-chlorophenyl	−0.60

N°	R_1	R_2	R_3	R_4	$pMIC_{80}$
67	–OCH$_3$	–OCH$_3$	–(CH$_2$)$_3$CH$_3$	–H	−1.51
68	–H	–F	–(CH$_2$)$_3$CH$_3$	–H	−1.57
69	–OCH$_3$	–OCH$_3$	–(CH$_2$)$_7$CH$_3$	=O	−0.52
70	–OCH$_3$	–OCH$_3$	–(CH$_2$)$_7$CH$_3$	–H	−0.53
71	–OH	–OH	–(CH$_2$)$_7$CH$_3$	–H	0.04
72	–H	–F	–(CH$_2$)$_8$CH$_3$	=O	0.05
73	–H	–F	–(CH$_2$)$_8$CH$_3$	–H	−0.27
74	–OCH$_3$	–OCH$_3$	–(CH$_2$)$_{11}$CH$_3$	=O	−1.06
75	–OCH$_3$	–OCH$_3$	–(CH$_2$)$_{11}$CH$_3$	–H	−0.77

Table 1. continuation

Test set			

N°	R$_1$	R$_2$	R$_3$	pMIC$_{80}$
76	$-CH_2(CH_2)_5CH_3$	$-OH$	$-OH$	-1.04
77	$-CH_2CH_2CH=C(CH_3)_2$	$=NOH$	$-OH$	-1.05
78	$-CH_2CH_2CH=C(CH_3)_2$	$=O$	$-OH$	-1.67
79	$-CH_2(CH_2)_3CH_3$	$=O$	$-OCH_3$	-1.97
80	$-CH_2(CH_2)_5CH_3$	$-OH$	$-OCH_3$	-1.32

N°	R$_1$	R$_2$	pMIC$_{80}$
81	$-CH_2(CH_2)_3CH_3$	$-OCH_3$	-2.00

N°	R$_1$	R$_2$	pMIC$_{80}$	N°	R$_1$	R$_2$	pMIC$_{80}$
82	5–OCH$_3$	–H	-0.90	94	8–OH	$-(CH_2)_{13}CH_3$	-0.60
83	6–OCH$_3$	–H	-0.90	95	5–OCH$_3$	$-(CH_2)_{17}CH_3$	-0.90
84	8–OH	–H	-0.90	96	5–OH	$-(CH_2)_{17}CH_3$	-0.90
85	5–OCH$_3$	$-(CH_2)_3CH_3$	-0.90	97	7–OH	$-(CH_2)_{17}CH_3$	-0.90
86	6–OCH$_3$	$-(CH_2)_3CH_3$	-0.90	98	7–OCH$_3$	phenyl	-0.90
87	6–OH	$-(CH_2)_3CH_3$	-0.90	99	5–OH	phenyl	-0.90
88	7–OH	$-(CH_2)_7CH_3$	-0.30	100	6–OCH$_3$	benzyl	-0.90
89	7–OCH$_3$	$-(CH_2)_8CH_3$	0.30	101	5–OH	benzyl	-0.90
90	6–OH	$-(CH_2)_9CH_3$	0.90	102	6–OCH$_3$	4-methoxyphenyl	-0.90
91	5–OH	$-(CH_2)_{11}CH_3$	0.00	103	6–OH	4-methoxyphenyl	-0.90
92	5–OCH$_3$	$-(CH_2)_{13}CH_3$	-0.60	104	6–OCH$_3$	4-chlorophenyl	-0.90
93	6–OH	$-(CH_2)_{13}CH_3$	0.30	105	6–OH	4-chlorophenyl	-0.30

Table 1. continuation

			Test set		

N°	R_1	R_2	R_3	R_4	$pMIC_{80}$
106	–H	–F	$-(CH_2)_3CH_3$	=O	−1.25
107	–H	–F	$-(CH_2)_7CH_3$	–H	−0.28
108	$-OCH_3$	$-OCH_3$	$-(CH_2)_8CH_3$	=O	−0.20
109	$-OCH_3$	$-OCH_3$	$-(CH_2)_8CH_3$	–H	−0.21
110	–OH	–OH	$-(CH_2)_8CH_3$	–H	−0.25

search (Population size = 100, reproduction/mutation trade-off = 0.5, selection bias = 50%) was carried out using 10 independent populations of 2000 models each that evolved for more than 100 generations or at least 2 million models. The descriptors found in the 20 best models were pooled together, autoscaled and employed to develop partial least squares (PLS) models, as implemented in the PIROUETTE 4.0 software (Infometrix, Washington, USA).[19,21,22]

Hologram-based QSAR strategy

Statistical HQSAR modeling was carried out as previously described.[23,24] Briefly, each molecule in the dataset was decomposed into linear, branched, and overlapping fragments, which were hashed to a fixed-length array (53 to 401 bins) that is called molecular hologram. The bin occupancies encode compositional and topological molecular information used as independent variables in QSAR modeling. Parameters that affect hologram generation such as hologram length, fragment size and fragment distinction (atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor/acceptor (DA)) were evaluated during model development, using default fragment size 4-7 over the 12 default series of hologram lengths. Next, the influence of fragment size (2-5, 3-6, 5-8, 6-9, 7-10, 8-11) was further investigated for the best model. All models generated in this study were investigated using the full cross-validated r^2 (q^2) Partial Least Squares (PLS) Leave-One-Out (LOO) method and the stability of best model was evaluated using the mean value of full cross-validated r^2 (q^2) from 25 rounds of leave-5-out method.

QSAR model validation

External validation was carried out using a test set of 35 compounds, which were not considered for the purpose of QSAR model development. The predictive ability of the models was estimated as described by Schuumann and cols.[25-27]

Results and Discussion

Hepatotoxicity is one of the main concerns for long-term treatment of deep-seated fungal infections with azole drugs, especially among immune-compromised patients that suffer from recurring fungal infections.[28] As a consequence, medicinal chemistry efforts have been directed towards separating azole´s toxicophore from their pharmacophore. Some progress has been achieved by maximizing the interactions with residues of fungal lanosterol 14α-demethylase that are different from or absent in the human counterpart, while ruling out interactions with the heme group. This strategy is hampered by the lack of reliable structural information on the macromolecular target, but could benefit from ligand-based strategies, such as quantitative structure-activity relationship models that rely either on topological or hologram-based descriptors.

Accordingly, a dataset of 110 isoquinoline, chromene and 2-aminetetraline derivatives were split into training set (compounds **1-75**) and test set (compounds **76-110**) so that chemical diversity and potency range were similarly represented in both sets. (Figure 1S) The antifungal-activity against *C. neoformans* (MIC_{80}) ranges from 1370 mmol L^{-1} to 0.5 mmol L^{-1} (a factor of about 2700) and was measured under the same experimental conditions. This fact renders

this property suitable for QSAR studies. Although all assays were carried out under standardized NCCLS (National Committee for Clinical Laboratory Standards) conditions, the biological data still bear a minor inter-experiment variation, for instance MIC_{80} values for fluconazole vary from 6.5 mmol L^{-1} to 13.1 mmol L^{-1}. Aiming at overcome this dilemma and produce robust QSAR models, the $MIC_{80cpd}/MIC_{80fluconazole}$ ratio was employed to calculate $pMIC_{80}$ ($-\log MIC_{80cpd}/ MIC_{80fluconazole}$), the dependent variable used in QSAR modeling. This strategy improved the overall statistical soundness of the QSAR models due to experimental-noise reduction, proving essential to the following steps.

Next, 215 topological descriptors, available in DRAGON 5.5, were computed and employed to build preliminary multiple linear regression QSAR models with up to 6 variables, by means of genetic algorithm, using MOBYDIGS 1.0. QSAR models with up to 5 variables have q^2 values below 0.5, but improved results were achieved when 6 variables were employed ($r^2 = 0.78$, $q^2 = 0.76$). Nevertheless, the predictive ability of the MLR models was marginal ($r^2_{pred} = 0.19$). This suggests that the chemical and structural features captured in the model do not extend beyond the chemical space of training set compounds, limiting its usefulness in drug design. In order to improve QSAR model applicability domain, we resorted to more powerful statistical tools such as partial least square (PLS) and principal component analysis (PCA), available in PIROUETTE 4.0 software. Thus, 32 descriptors, found in the 20 best models were pooled together, autoscaled and used for further PCA and PLS QSAR model development. The underlying goal of using unsupervised chemometric tools, such as PCA, is to investigate whether the low predictive power is due to insufficient sampling of descriptor space (reduced number of descriptors per model) or due to inadequate representation of chemical space (i.e., 3D information is required to describe the structure-activity relationship for this dataset).

Analysis of the loading plot for the 2 first PCs, which account for 67.6% of the data variability (Figure 2), shows that the 80% of most potent inhibitors ($pMIC_{80} > 0.01$) have PC2 values below -2.5, whereas 79% of the average potency inhibitors ($-1.0 < pMIC_{80} < 0.0$) are grouped between -2.5 and 2.0 and 83% of weaker inhibitors ($pMIC_{80} < -1.0$) lie above 2.0. This result clearly shows that PC2 explain the compounds potency, suggesting that selected descriptors are capable of explaining the potency profile of this dataset and thus are suitable for QSAR model development. Accordingly, they were gathered, autoscaled and employed for QSAR models development by partial least square regression (PLS), as available in PIROUETTE 4.0. Initial models show only minor
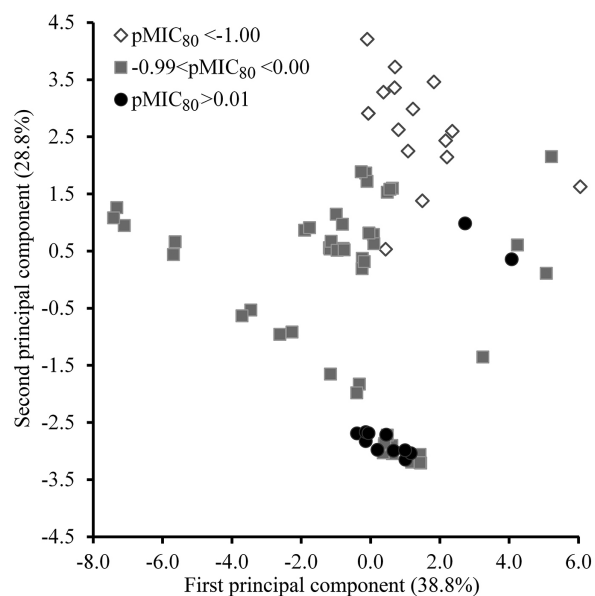

Figure 2. Loading plot for training set compounds according to PCA. Weak inhibitors (open diamond) are clearly separated from average (grey squares) and potent (black cycles) non-azole antifungals.

improvement in statistical values ($r^2 = 0.81$ and $q^2 = 0.77$, 6 LVs), but large increase in predictive ability ($r^2_{pred} = 0.72$), in comparison to the predictive power of the RLM model ($r^2_{pred} = 0.19$). In order to further improve the statistical soundness of QSAR models, an iterative exclusion of 16 descriptors that have minor contributions to the regression vector was carried out until no improvement in statistical values was achieved. The regression vector can be thought of as a weighted sum of the loadings included in the model. Thus, descriptors with small coefficients do not contribute significantly to explain the dependent variable (biological activity) and can be discarded from the model. This strategy lead to an overall improvement in model adjustment ($r^2 = 0.92$, $q^2 = 0.90$) and predictive power ($r^2_{pred} = 0.86$), when 6 latent variables (LV), 96.4% explained variance, where employed (Figure 3 and Table 2).

More than statistical soundness, useful QSAR models should also provide some insight into the physical-chemical and structural requirements for the biological activity.[29] This sort of information can be gathered from the analysis of the regression vector plot, which underscores the relative importance of descriptors towards the final QSAR model (Figure 4). According to regression vector, B09[C–N] has the greatest positive contribution to antifungal activity of these compounds.

B09[C–N] describes the presence or absence of carbon (from aliphatic side chain or methoxyl group) linked to nitrogen at topological distance of 9 bonds. It is interesting to note that 93.33% of weak inhibitors ($pMIC_{80} < -1.0$) have zero value for this descriptor, whereas this moiety is easily
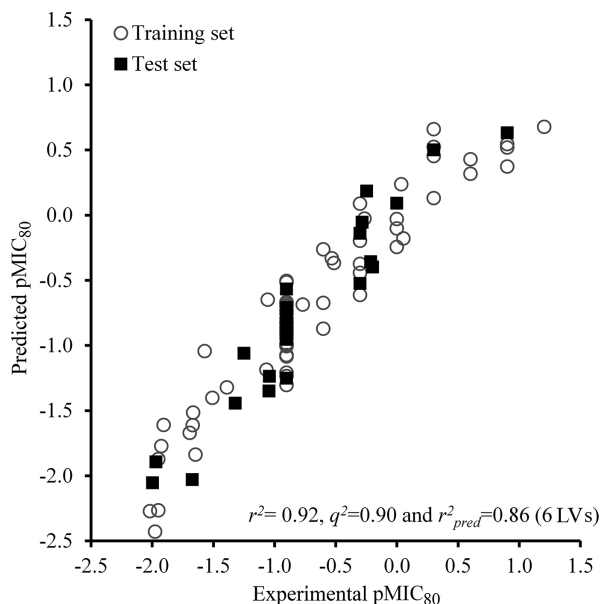
Figure 3. Predicted *vs.* Experimental values of $pMIC_{80}$ according to the best descriptor-based QSAR model.

identified in most of average (73.33%) and potent (93.33%) inhibitors (Figure 5). Although this descriptor can be traced to different molecular fragments, it is noteworthy that most of them are constant throughout the dataset, whereas some chemical diversity can be traced to fragments connecting methoxyl groups and the nitrogen atom of isoquinoline inhibitors (Figure 5). In fact, linear alkyl substituent groups with 8 or more atoms connected to nitrogen atom seems to

Figure 4. Regression vector plot. Descriptors that increase potency have positive leverage, whereas those that reduce potency display negative values.

Table 2. Predicted $pMIC_{80}$ values for test set compounds according to hologram-based (HQSAR) and descriptor-based QSAR models

Compounds	Experimental $pMIC_{80}$	HQSAR		QSAR	
		Predicted $pMIC_{80}$	Residual	Predicted $pMIC_{80}$	Residual
76	−1.04	−1.24	0.20	−1.41	0.37
77	−1.05	−1.35	0.30	−1.29	0.24
78	−1.67	−2.03	0.36	−1.75	0.08
79	−1.97	−1.89	−0.08	−1.87	−0.10
80	−1.32	−1.44	0.12	−1.60	0.28
81	−2.00	−2.05	0.06	−2.10	0.10
82	−0.90	−0.95	0.05	−0.94	0.04
83	−0.90	−0.86	−0.04	−0.87	−0.03
84	−0.90	−0.84	−0.06	−0.92	0.02
85	−0.90	−0.71	−0.19	−0.82	−0.09
86	−0.90	−0.74	−0.17	−0.75	−0.15
87	−0.90	−0.57	−0.34	−0.48	−0.42
88	−0.30	−0.14	−0.16	−0.83	−0.08
89	0.30	0.50	−0.20	0.24	0.06
90	0.90	0.63	0.27	0.53	0.38
91	0.00	0.09	−0.09	−0.04	0.04
92	−0.60	−0.26	−0.35	−0.27	−0.34
93	0.30	−0.02	0.32	0.07	0.23
94	−0.60	−0.11	−0.49	−0.28	−0.33
95	−0.90	−0.96	0.06	−0.72	−0.18
96	−0.90	−0.82	−0.08	−0.69	−0.22
97	−0.90	−0.76	−0.15	−0.50	−0.40
98	−0.90	−1.04	0.14	−0.88	−0.03
99	−0.90	−0.86	−0.05	−0.88	−0.03
100	−0.90	−0.95	0.05	−0.97	0.07
101	−0.90	−0.75	−0.15	−1.01	0.11
102	−0.90	−1.25	0.35	−0.73	−0.17
103	−0.90	−0.76	−0.14	−0.45	−0.45
104	−0.90	−0.79	−0.11	−0.69	−0.22
105	−0.30	−0.52	0.22	−0.41	0.11
106	−1.25	−1.06	−0.19	−0.94	−0.31
107	−0.28	−0.05	−0.23	−0.09	−0.19
108	−0.20	−0.40	0.20	−0.44	0.24
109	−0.21	−0.36	0.14	−0.66	0.44
110	−0.25	0.18	−0.43	−0.11	−0.14

improve the biological activity of non-azole compounds. This hypothesis is reinforced by docking studies carried out by Yao, B and cols. (2007) which suggest that such moieties might properly interact into the hydrophobic pocket of the lanosterol 14α-demethylase.[10]
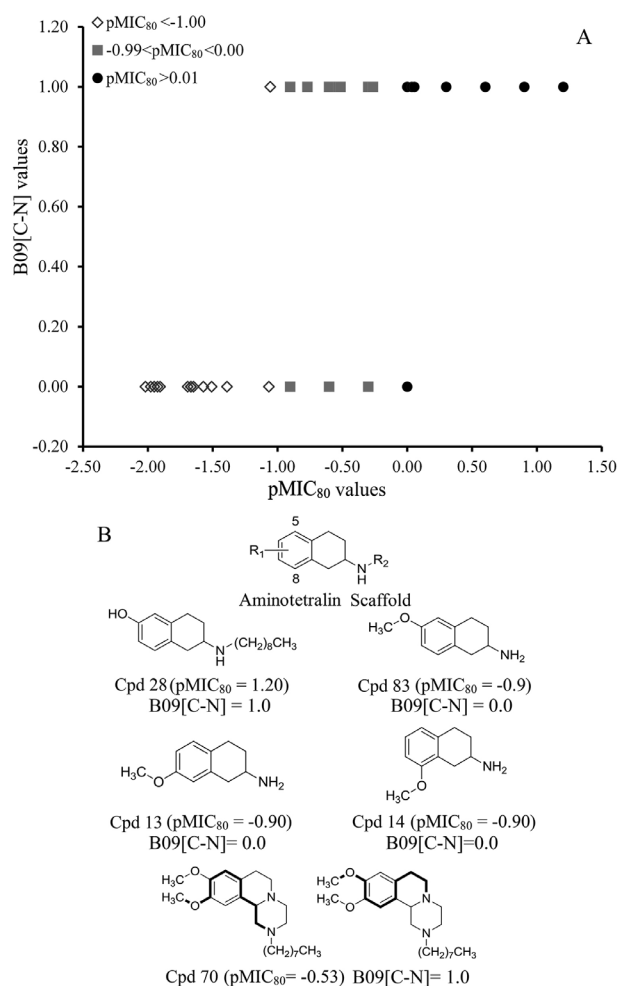

Figure 5. (A) plot of B09[C–N] values versus pMIC$_{80}$ of training set compounds; (B) B09[C–N] descriptor highlighted in different molecules of the training set.

On the other hand, ESpm08r which accounts for resonance integrals between atoms eight bonds apart[30] shows a negative contribution to the biological activity (Figure 6). Then, it is reasonable to assume that moieties that affect charge distribution on aromatic rings, such as the fused pyranone or the dihydro-pyrane ring, play an important role towards the biological activity.

Ring substituent importance to the biological property have already been highlighted by Tang and cols (2008), but for a different reason as those authors claim that these moieties would contribute to activity through hydrogen bonding to Tyr169. Our results do not contradict this hypothesis, but adds a new, quantitative, perspective to
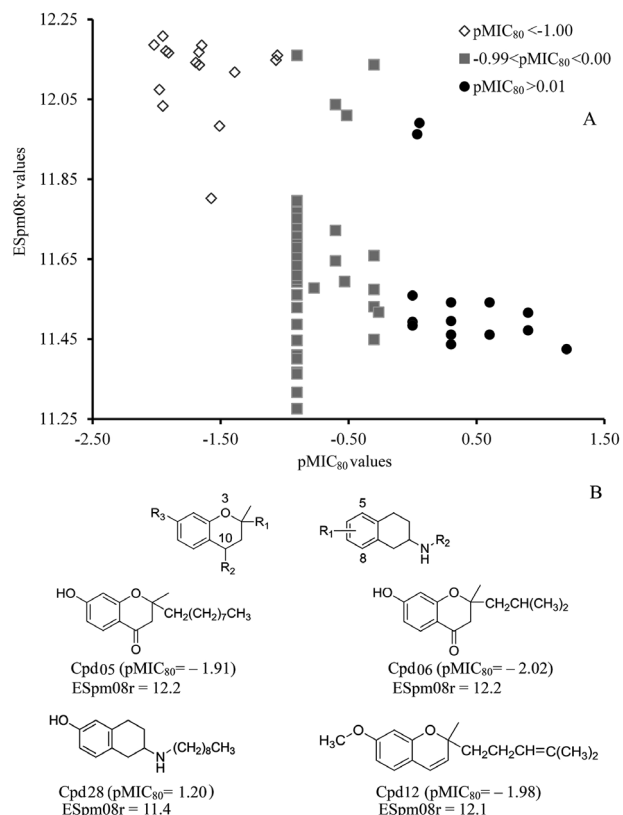

Figure 6. (A) Plot of ESpm08r values versus pMIC$_{80}$ of training set compounds; (B) Influence of fused rings over the charge density and antifungal activity of selected compounds.

the scenario: Charge distribution, which might strengths or weakens a hydrogen bond, contributes significantly to non-azoles antifungals potency.

Despite the fact that homology models have played their part in novel antifungal drug design campaigns, we decided to investigate whether 2D information only would suffice to guide the development of second-generation non-azole antifungals. Therefore, we resorted to another 2D QSAR approach, Hologram QSAR, which has been proven as effective as many 3D QSAR approaches and complements the information provided by descriptor-based 2D QSAR models.[26,27,31,32] Hence, molecular holograms were generated for training set compounds using a number of fragment distinction combinations (Table 3).

The standard fragment distinction (ABC) shows poor fit ($r^2 = 0.55$), which was not improved by the addition of hydrogen (H) or chirality (Ch) to fragment distinction (compare models 2 and 3 *vs.* 1). A slight improvement was observed when donor and acceptor atoms (DA) was considered ($r^2 = 0.58$). The low statistical quality of the models might indicate that hologram-based descriptors are not suitable to describe non-azole antifungals biological property or that the fragment distinction combinations investigated so far have contrasting effects to the model. For

Table 3. Influence of fragments distinction over HQSAR statistical parameters, using standard fragment size (4-7)*

Model	Fragment distinction	q^2	r^2	HL	LV
1	ABC	0.46	0.55	199	3
2	ABCH	0.53	0.54	71	2
3	ABCHCh	0.53	0.54	53	3
4	ABCHChDA	0.48	0.58	71	3
5	ABH	0.53	0.57	307	3
6	ACH	0.55	0.56	71	3
7	ABCh	0.53	0.67	151	5
8	ACCh	0.52	0.66	151	5
9	ABHCh	0.54	0.57	151	3
10	ACHCh	0.54	0.59	307	3
11	ABCCh	0.52	0.69	401	6
12	ACDA	0.48	0.66	53	6

q^2, cross-validated correlation coefficient; r^2, noncross-validated correlation coefficient; HL, hologram length; *LV*, optimal number of latent variable. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

instance, use of DA might improve fit, but B could have the opposite effect. Aiming at further investigate this hypothesis simpler models (5 to 8) were built, in which either B or C on the distinction was not employed. This strategy provided two models with improved fit, but still with limited internal consistency (models 7 and 8). Further addition of the fragment distinction C to model 8 resulted in the best fitted model, but no significant improvement in internal consistency was observed. Aiming at circumvent this setback, the influence of fragment size over the statistical parameter was also investigated (Table 4). Generally, only the model with highest statistical quality is evaluated, but as shown below this strategy might be misleading.

It is well known that fragment-distinction and fragment-size control the number of fragments that are generated and then hashed to build the hologram. During this step, fragment collision, due to greater number of fragments than hologram length, leads to abrupt changes in the statistical parameters.[33] Thus, it is reasonable to assume that even sub-optimal HQSAR models, built with default fragment size (4-7 atoms), might provide the best QSAR model after fragment-size optimization. Therefore, the influence of the size was investigated for all models with $q^2 > 0.50$ (Table 4). This strategy revealed a common trend among the models: larger sized fragments afford models with significantly higher statistical quality, which is the best model (Figure S2 and Table 4).

This intriguing feature, obviously, must be related to some hidden SAR pattern that was not captured by

Table 4. Influence of fragment size over the statistical parameters of the 3 best HQSAR models

Fragments' size		q^2	r^2	HL	LV
ABCH	2-5	0.47	0.62	71	6
	3-6	0.47	0.58	71	4
	5-8	0.63	0.76	71	6
	6-9	0.66	0.79	71	6
	7-10	0.75	0.83	71	6
	8-11	0.81	0.87	71	6
ABCHCh	2-5	0.49	0.56	53	3
	3-6	0.47	0.55	53	3
	5-8	0.56	0.72	53	6
	6-9	0.72	0.82	53	6
	7-10	0.69	0.81	53	6
	8-11	0.74	0.82	53	5
ABH	2-5	0.58	0.75	307	6
	3-6	0.48	0.57	307	4
	5-8	0.58	0.75	307	6
	6-9	0.75	0.83	307	6
	7-10	0.75	0.84	307	6
	8-11	0.75	0.84	307	6
ACH	2-5	0.47	0.56	71	4
	3-6	0.47	0.62	71	6
	5-8	0.60	0.75	71	6
	6-9	0.72	0.82	71	6
	7-10	0.69	0.81	71	6
	8-11	0.62	0.78	71	5
ACHCh	2-5	0.46	0.56	307	4
	3-6	0.47	0.58	307	4
	5-8	0.60	0.78	307	6
	6-9	0.69	0.83	307	6
	7-10	0.72	0.83	307	6
	8-11	0.72	0.83	307	6
ABHCh	2-5	0.48	0.55	151	4
	3-6	0.50	0.63	151	6
	5-8	0.63	0.76	151	6
	6-9	0.69	0.82	151	6
	7-10	0.75	0.84	151	6
	8-11	0.76	0.84	151	6

default-sized HQSAR models. One reasonable explanation is that 7-atom length fragments are not enough to explain the contribution of the N-alkyl side chain, which according to Yao and cols. (2007) and Tang and cols. (2010) is essential to activity.

Beside of good statistical quality, QSAR models should be able to predict the biological property of
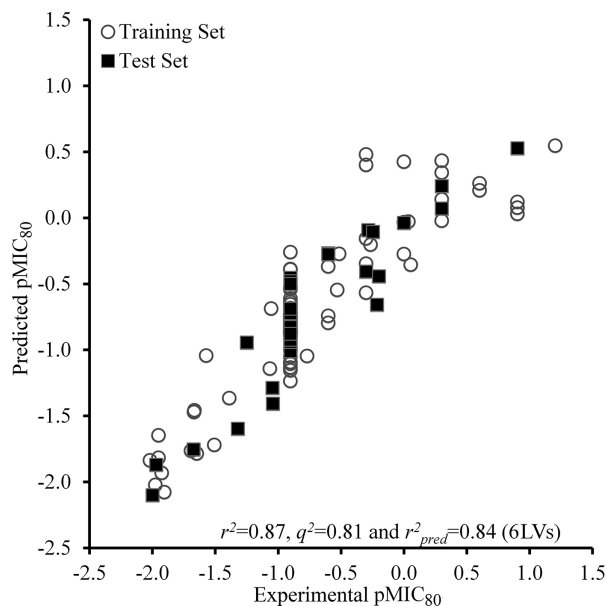
Figure 7. Predicted *vs.* experimental values of pMIC$_{80}$ according to the best Hologram QSAR model.

congeneric molecules and thus guide the development of second generation non-azole antifungals. The predictive power of the best HQSAR model (Figure 7 and Table 2), evaluated as described in Material and Methods section, ($r^2_{pred} = 0.84$) is similar to that observed for descriptor-based-QSAR model ($r^2_{pred} = 0.86$). Although it should be mentioned that compounds with pMIC$_{80}$ > 0.5, within the training-set, are frequently predicted as less potent than they really are. This trend is less obvious in descriptor-based QSAR model suggesting that hologram-based descriptors (molecular holograms) do not cover all the information required to explain the biological activity of non-azole antifungals that are more potent than fluconazole. Nevertheless, it should be possible to integrate the information provided by both models to reveal structural and chemical properties that are crucial for the antifungal activity. As stated before, regression vector plot can be used to analyze the contribution of each descriptor to the descriptor-based QSAR model. Similarly, contribution maps, which graphically represent atoms or fragments with positive (yellow or green) or detrimental (orange, red) contribution towards antifungal activity, can be used to analyze the HQSAR model. A shortcoming of this analysis it that fragments that hold pharmacophore groups into position can be colored gray (neutral contribution), once they are found in weak and potent compounds.

According to contribution map of the less potent compound (cpd 06, pMIC$_{80}$ = −2.02), the aromatic ring appears partially colored in red (Figure 8) whereas in potent compounds (cpd 35, pMIC$_{80}$ = 0.60) the aromatic ring is

widely colored in green and yellow, indicating that this ring can either improve or decrease the biological activity. Apparently, this coloring scheme can be associated with resonance effects once the most potent compound has one of the lowest Espm08r values (11.5), whereas the less potent one has the highest one (12.2).
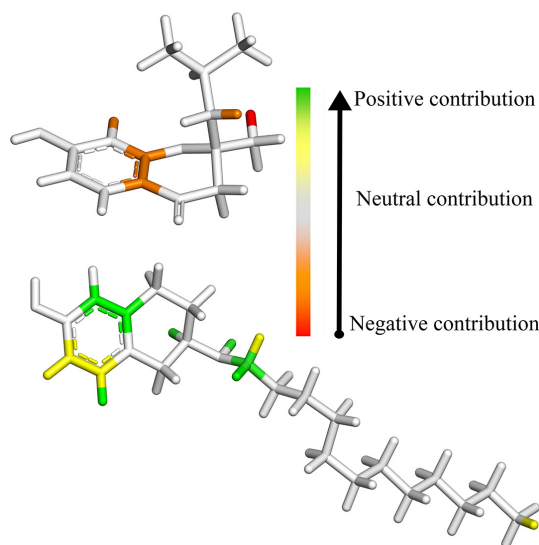

Figure 8. Contribution maps for weak (upper panel, cpd 06 pMIC$_{80}$ = −2.02) and potent (lower panel, cpd 35 pMIC$_{80}$ = 0.60) non-azole antifungals according to the best HQSAR model.

Conclusions

Although our approach is based uniquely on 2D fragments or physical-chemical descriptor, QSAR models with good statistical values were developed (descriptor-based ($r^2 = 0.92$, $q^2 = 0.90$, 6 LVs and $r^2_{pred} = 0.86$; hologram-based $r^2 = 0.87$, $q^2 = 0.81$, 6 LVs and $r^2_{pred} = 0.84$). Furthermore, the interactive analysis of 2D QSAR models showed to be useful to highlight the importance of resonance effects towards antifungal activity of isoquinoline, chromenes and 2-aminetetralines derivatives as well as pinpoint which part of the molecules are most influenced by this feature. For instance, HQSAR contribution maps suggest that fused rings play a crucial role towards the antifungal activity, due to electrostatic interactions towards lanosterol 14-alpha demethylase. Besides, the great molecular diversity of the dataset along with the good predictive ability of the QSAR models indicate that molecular features underscored in this study should apply not only to training set compounds, but also the congeneric molecules that lie within the chemical space sampled by isoquinoline, chromenes and 2-aminetetralines derivatives. Thus, QSAR models reported herein should be useful to guide the design of more potent non-azole antifungal compounds.

Supplementary Information

Acknowledgements

References

1. Grimaldi, M.; De Rosa, M.; Di Marino, S.; Scrima, M.; Posteraro, B.; Sanguinetti, M.; Fadda, G.; Soriente, A.; D'Ursi, A. M.; *Bioorg. Med. Chem.* **2010,** *18*, 7985.

2. Tang, H.; Zheng, C.; Lv, J.; Wu, J.; Li, Y.; Yang, H.; Fu, B.; Li, C.; Zhou, Y.; Zhu, J.; *Bioorg. Med. Chem. Lett.* **2010,** *20*, 979.

3. Vasanthanathan, P.; Lakshmi, M.; Babu, M. A.; Gupta, A. K.; Kaskhedikar, S. G.; *Chem. Pharm. Bull.* **2006,** *54*, 583.

4. Perfect, J. R.; Casadevall, A.; *Infect. Dis. Clin. N. Am.* **2002,** *16*, 837.

5. Perfect, J. R.; Dismukes, W. E.; Dromer, F.; Goldman, D. L.; Graybill, J. R.; Hamill, R. J.; Harrison, T. S.; Larsen, R. A.; Lortholary, O.; Nguyen, M. H.; Pappas, P. G.; Powderly, W. G.; Singh, N.; Sobel, J. D.; Sorrell, T. C.; *Clin. Infect. Dis.* **2010,** *50*, 291.

6. Polak, A.; *Mycoses* **1999,** *42*, 355.

7. Chiou, C. C.; Groll, A. H.; Walsh, T. J.; *Oncologist* **2000,** *5*, 120.

8. Sionov, E.; Lee, H.; Chang, Y. C.; Kwon-Chung, K. J.; *PLoS Pathog.* **2010,** *6*, e1000848.

9. Varma, A.; Kwon-Chung, K. J.; *Antimicrob. Agents Chemother.* **2010,** *54*, 2303.

10. Yao, B.; Ji, H.; Cao, Y.; Zhou, Y.; Zhu, J.; Lu, J.; Li, Y.; Chen, J.; Zheng, C.; Jiang, Y.; Liang, R.; Tang, H.; *J. Med. Chem.* **2007,** *50*, 5293.

11. Somchit, N.; Norshahida, A. R.; Hasiah, A. H.; Zuraini, A.; Sulaiman, M. R.; Noordin, M. M.; *Hum. Exp. Toxicol.* **2004,** *23*, 519.

12. Kumar, S.; Bawa, S.; Drabu, S.; Panda, B.; *Med. Chem. Res.* **2011,** *20*, 1340.

13. Vasanthanathan, P.; Lakshmi, M.; Babu, M. A.; Kaskhedikar, S. G.; *Med. Chem.* **2006,** *2*, 363.

14. Tani, N.; Rahnasto-Rilla, M.; Wittekindt, C.; Salminen, K. A.; Ritvanen, A.; Ollakka, R.; Koskiranta, J.; Raunio, H.; Juvonen, R. O.; *Eur. J. Med. Chem.* **2012,** *47*, 270.

15. Foo, H.; Gottlieb, T.; *Clin. Infect. Dis.* **2007,** *45*, 803.

16. Zhu, J.; Lu, J.; Zhou, Y.; Li, Y.; Cheng, J.; Zheng, C.; *Bioorg. Med. Chem. Lett.* **2006,** *16*, 5285.

17. Tang, H.; Zhou, Y. J.; Li, Y. W.; Lv, J. G.; Zheng, C. H.; Chen, J.; Zhu, J; *Chin. Chem. Lett.* **2008,** *19*, 264.

18. Ji, H.; Zhang, W.; Zhang, M.; Kudo, M.; Aoyama, Y.; Yoshida, Y.; Sheng, C.; Song, Y.; Yang, S.; Zhou, Y.; Lü, J.; Zhu, J.; *J. Med. Chem.* **2003,** *46*, 474.

19. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; *Data Handling in Science and Technology*, Elsevier: Oxford, 2003.

20. MobyDigs, Version 1, Talete srl, Milano, Italy, 2004.

21. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003,** *22*, 69.

22. Gramatica, P.; *QSAR Comb. Sci.* **2007,** *26*, 694.

23. Castilho, M. S.; Postigo, M. P.; de Paula, C. B. V.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2006,** *14*, 516.

24. Castilho, M. S.; Guido, R. V.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2007,** *15*, 6242.

25. Schuurmann, G.; Ebert, R. U.; Chen, J.; Wang, B.; Kuhne, R.; *J. Chem. Inf. Model.* **2008,** *48*, 2140.

26. Doddareddy, M. R.; Cho, Y. S.; Koh, H. Y.; Pae, A. N.; *Bioorg. Med. Chem.* **2004,** *12*, 3977.

27. de Freitas, H. F.; Castilho, M. S.; *Med. Chem.* **2012,** *8*, 252.

28. Kuroyanagi, J.; Kanai, K.; Horiuchi, T.; Takeshita, H.; Kobayashi, S.; Achiwa, I.; Yoshida, K.; Nakamura, K.; Kawakami, K.; *Chem. Pharm. Bull.* **2011,** *59*, 341.

29. Andricopulo, A. D.; Yunes, R. A.; *Chem. Pharm. Bull.* **2001,** *49*, 10.

30. Estrada, E.; *J. Chem. Inf. Model.* **1997,** *37*, 320.

31. Valadares, N. F.; Castilho, M. S.; Polikarpov, I.; Garratt, R. C.; *Bioorg. Med. Chem.* **2007,** *15*, 4609.

32. Mota, S. G. R.; Barros, T. F.; Castilho, M. S., *J. Braz. Chem. Soc.* **2009,** *20*, 451.

33. Seel, M.; Turner, D. B.; Willett, P., *Quant. Struct. –Act. Relat.* **1999,** *18*, 245.