



QSAR modeling of the antifungal activity against *Candida albicans* for a diverse set of organic compounds

Alan R. Katritzky^{a,*}, Svetoslav H. Slavov^{a,c}, Dimitar A. Dobchev^{a,b}, Mati Karelson^{b,c}

^a Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA

^b Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

^c Department of Chemistry, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia

ARTICLE INFO

Article history:

Received 14 February 2008

Revised 2 May 2008

Accepted 5 May 2008

Available online 9 May 2008

Keywords:

Antifungal activity

Candida albicans

QSAR

BLMR

CODESSA PRO

ABSTRACT

The molecular structures of 83 diverse organic compounds are correlated by a quantitative structure–activity relationship (QSAR) to their minimum inhibitor concentrations (MIC expressed as $\log(1/\text{MIC})$), involving 6 descriptors with $R^2 = 0.788$, $F = 47.140$, $s^2 = 0.130$. A novel QSAR development technique is utilized combining advantages of the two frequently applied methods. The topological, electronic, geometrical, and hybrid type descriptors for the compounds were calculated by CODESSA PRO software.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Since the 1980s, fungal infection complications have been recognized as a major cause of morbidity and mortality in immunocompromised patients including those suffering from tuberculosis, infected with HIV-1, organ-transplant patients, and those undergoing cancer chemotherapy. Over 90% of patients diagnosed as HIV-positive contract a fungal infection.¹ Increased incidence of fungal infections also follows the frequent use of antibacterial and cytotoxic drugs.

Effective antifungal agents in the market each carry several drawbacks: either they possess high toxicity (e.g., Amphotericin B, AMB) or are increasingly ineffective due to appearance of resistant strains, limited spectrum of activity, tissue distribution, central nervous system (CNS) penetration, or high cost.^{1,3} AMB remains the ‘gold standard’ drug for life-threatening fungal infections, but its use is limited due to its severe renal toxicity and the inconvenience of intravenous dosing.⁴ Fluconazole (FLC) is an orally effective azole-based antifungal drug with low toxicity, but it has a limited antifungal spectrum and is not fungicidal.⁵ FLC shows significant efficacy against *Candida albicans* and *Candida neoformans*, but not against *Aspergillus niger* and *Aspergillus fumigatus*^{6,7} and the extensive use has increased the number of FLC-resistant *C. albicans* mutations.

Mortality from fungal infections is still unacceptably high. Thus, the development of new and effective antifungal agents against life-threatening systemic mycoses is an urgent and important task for scientists working in the field of pharmacy, organic synthesis, and QSAR. The subject of the present study is the dimorphic fungus *C. albicans* (*Candida stellatoidea*) which causes a variety of superficial and deep-seated mycoses and exists in two easily identifiable morphologies, namely yeast and hyphal forms.^{8,9} The hyphal form may play an important role in the pathogenesis of *C. albicans*, particularly with respect to the ability of the organism to penetrate host cells and tissues; the hyphal form has often been considered to be important in the ability of the fungus to adhere to cell surfaces.⁹

Among recent QSAR articles concerning antifungal activity,^{10–16} many concentrate on ligand–receptor interactions, thus proposing CoMFA models with good to excellent predictive capabilities. Results obtained by ANN, and classic 2D-QSAR techniques have been carried out. However, each of these publications deals with a particular dataset: we have now combined several sets of diverse compounds and searched for a general QSAR model.

2. Dataset

We collected experimental data for a series of 83 diverse compounds (see Table 1)^{3,4,17,18} combining cyanoborane, fluconazole, carbonylaminobenzoxazole, and imidazolylmethylindole (salts excluded) derivatives. The molecules forming three of the subsets

* Corresponding author. Tel.: +1 352 392 0554; fax: +1 352 392 9199.

E-mail address: katritzky@chem.ufl.edu (A.R. Katritzky).

Table 1

Structural formulas, observed and predicted log(1/MIC) values and their difference for the studied dataset of 83 diverse compounds

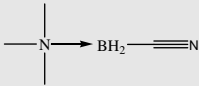
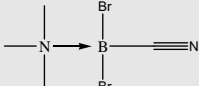
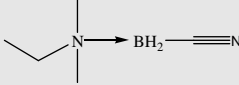
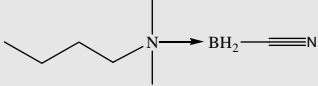
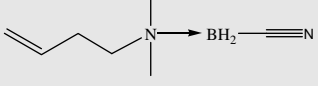
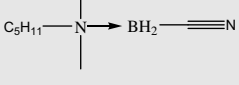
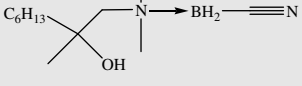
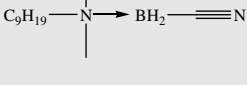
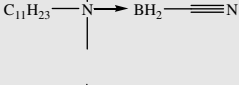
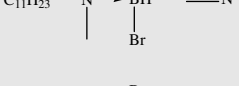
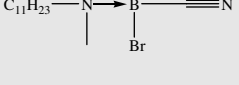
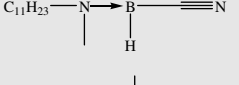


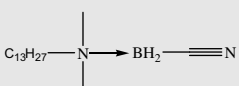
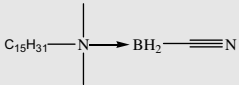
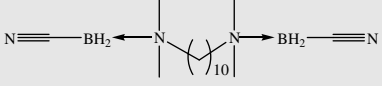
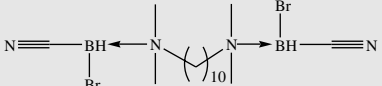
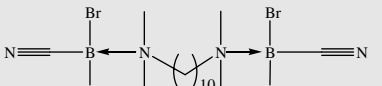
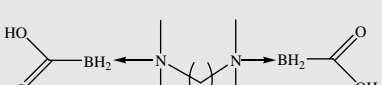
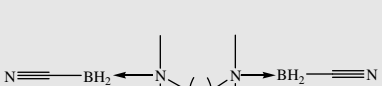
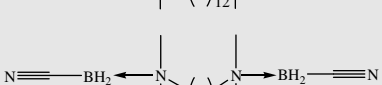
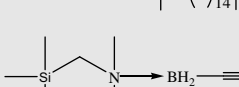
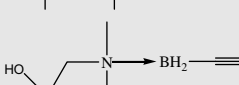
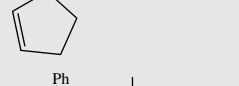
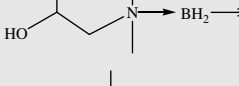
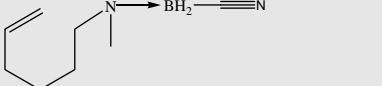
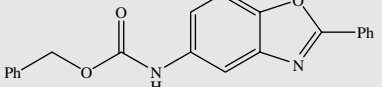
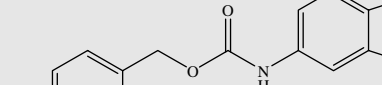
Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
1		−3.708	−3.969	0.261
2		−3.295	−3.841	0.546
3		−3.649	−3.646	−0.003
4		−3.553	−3.093	−0.460
5		−2.672	−2.486	−0.186
6		−2.909	−2.847	−0.062
7		−3.344	−2.785	−0.559
8		−2.176	−2.110	−0.066
9		−1.512	−1.865	0.353
10		−1.690	−1.713	0.023
11		−1.591	−1.790	0.199
12		−1.806	−1.985	0.179
13		−2.692	−2.324	−0.368
14		−1.785	−1.767	−0.018

Table 1 (continued)

Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
15		−1.763	−1.683	−0.080
16		−1.724	−1.560	−0.164
17		−2.611	−2.507	−0.104
18		−2.146	−2.240	0.094
19		−2.017	−2.407	0.390
20		−2.274	−2.540	0.266
21		−1.982	−1.576	−0.406
22		−1.634	−1.463	−0.171
23		−3.468	−3.346	−0.122
24		−3.142	−3.003	−0.139
25		−2.389	−2.757	0.368
26		−1.973	−2.150	0.177
27		−1.505	−1.864	0.359
28		−1.544	−1.678	0.134
29		−1.839	−1.804	−0.035

(continued on next page)

Table 1 (continued)

Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
30		−2.100	−1.995	−0.105
31		−1.820	−2.158	0.338
32		−1.532	−1.718	0.186
33		−2.127	−1.723	−0.404
34		−1.785	−1.870	0.085
35		−1.806	−2.039	0.233
36		−1.813	−1.574	−0.239
37		−1.813	−1.588	−0.225
38		−1.477	−1.699	0.222
39		−1.792	−1.850	0.058
40		−2.161	−1.875	−0.286
41		−1.820	−2.021	0.201

Table 1 (continued)

Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	Δ Obs – pred
42		–2.143	–2.224	0.081
43		–1.875	–1.808	–0.067
44		–1.556	–1.513	–0.043
45		–3.610	–3.299	–0.311
46		–3.143	–2.835	–0.308
47		–3.172	–3.016	–0.156
48		–2.846	–2.784	–0.062

(continued on next page)

Table 1 (continued)

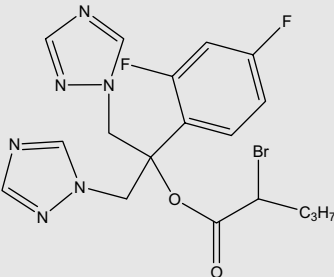
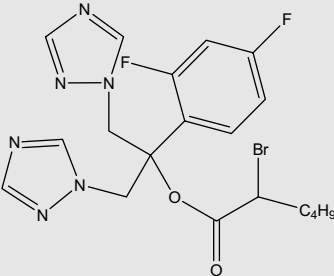
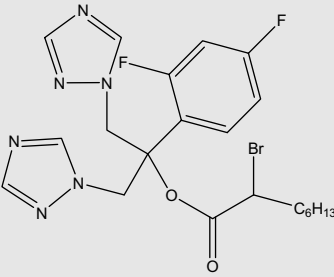
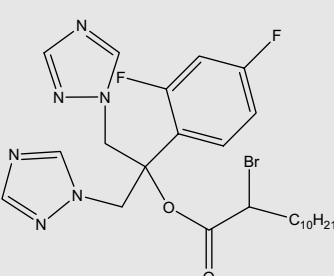
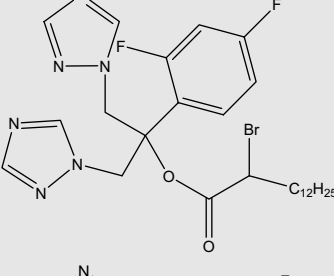
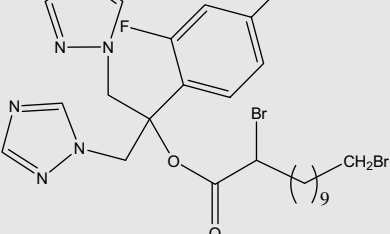
Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
49		−2.768	−2.737	−0.031
50		−2.100	−2.738	0.638
51		−1.431	−2.398	0.967
52		−3.412	−2.656	−0.756
53		−3.089	−2.729	−0.360
54		−1.342	−1.833	0.491

Table 1 (continued)

(continued on next page)

Table 1 (continued)

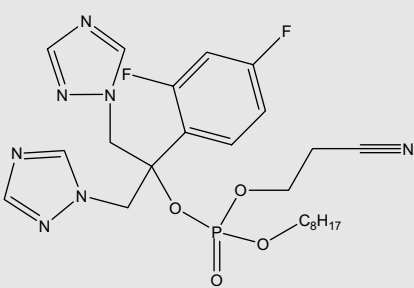
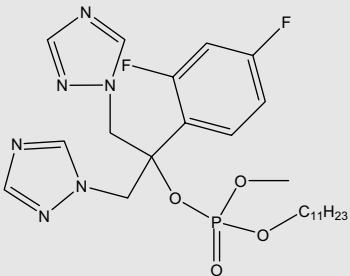
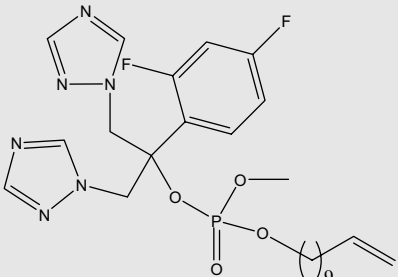
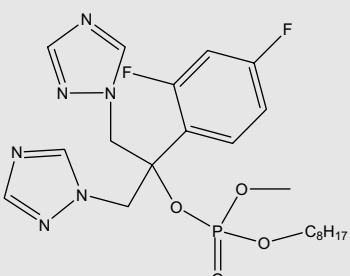
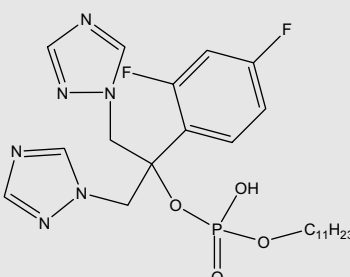
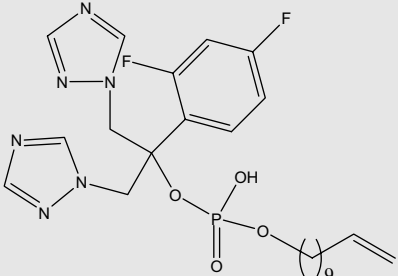
Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
61		−3.091	−2.842	−0.249
62		−3.138	−2.891	−0.247
63		−2.616	−2.936	0.320
64		−2.962	−2.841	−0.121
65		−3.148	−3.378	0.230
66		−3.445	−3.388	−0.057

Table 1 (continued)

Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
67		−3.377	−3.525	0.148
68		−3.208	−3.456	0.248
69		−1.924	−1.321	−0.603
70		−1.138	−1.622	0.484
71		−1.208	−1.530	0.322
72 [*]		−1.117	−2.083	0.966
73		−1.695	−1.757	0.062

(continued on next page)

Table 1 (continued)

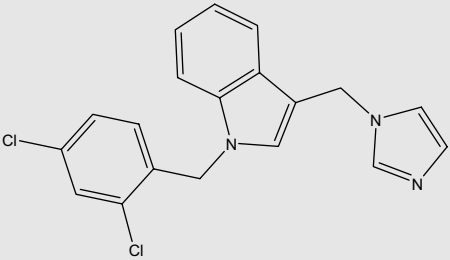
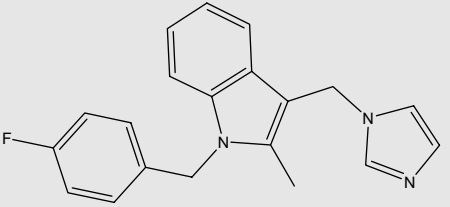
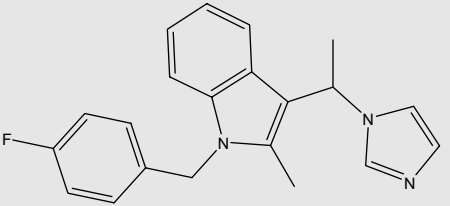
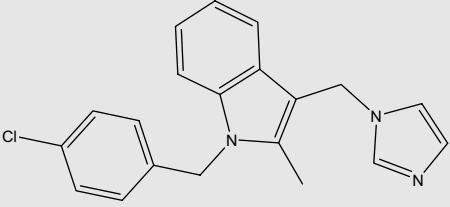
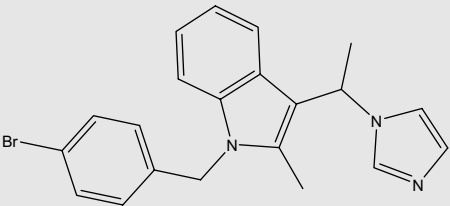
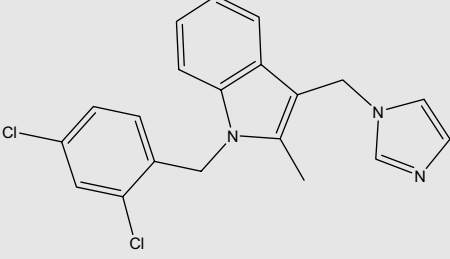
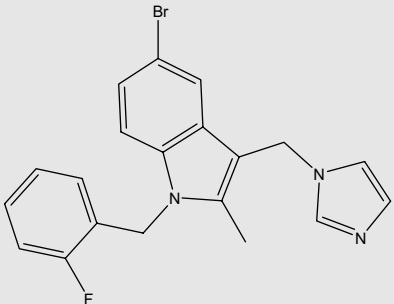
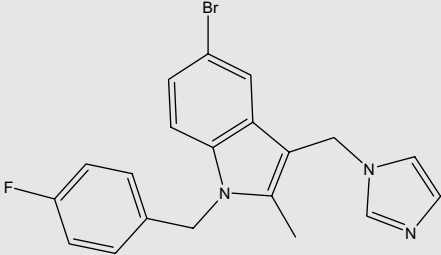
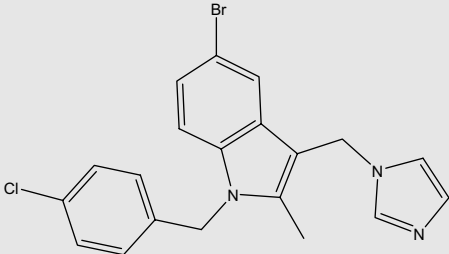
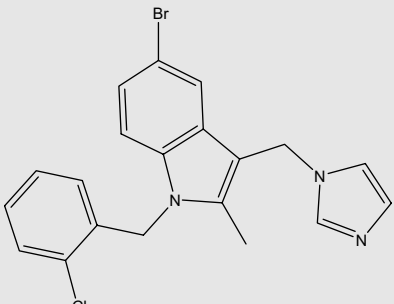
Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
74		−1.147	−1.629	0.482
75		−2.496	−1.919	−0.577
76		−1.819	−1.946	0.127
77		−0.474	−0.685	0.211
78		−2.404	−1.655	−0.749
79		−1.432	−1.497	0.065

Table 1 (continued)

Compound	Structure formula	Observed log(1/MIC)	Predicted log(1/MIC)	$\Delta\text{Obs} - \text{pred}$
80		−2.415	−2.263	−0.152
81		−2.415	−2.292	−0.123
82		−1.902	−1.883	−0.019
83		−2.397	−1.889	−0.508

Note: The compounds found to be outliers on the basis of their deviation from the regression line are marked with an asterisk ***.

(excluding cyanoboranes) are characterized by the presence of a halogenated aromatic ring, a heterocycle and in most cases a carboxylic group. The cyanoboranes also fit quite well into the model, and are included in the dataset, thus increasing the diversity. The minimum inhibitory concentration (MIC) values were modeled. The collected MIC values, in different units ($\mu\text{mol/L}$ and $\mu\text{g/mL}$) were transformed into $\mu\text{mol/L}$. A reciprocal logarithmic function was used to transform the distribution of the data to Gaussian (Fig. 1). Since not all sources provided information about the experimental error, our estimation of the relative error of the data used is based on Ref. 18. The relative errors for each individual compound, calculated from the absolute error values, were 3–20%. The reported model explains the data variance within the range of experimental error, assuming that the relative error covers approximately the same range for all investigated compounds.

3. QSAR analysis

A general QSAR approach requires several steps, usually performed by use of different program packages. Hence, the development of a QSAR model can sometimes be non-trivial and time consuming. CODESSA PRO software, used in the present study, combines all the necessary modules into one program package, enabling the calculation of numerous quantitative descriptors solely on the basis of molecular structural information.^{21,22,26} Research using CODESSA PRO has successfully correlated and predicted many physical and biological properties^{19–25} including melting and boiling points, vapor pressure, UV-intensities, and refractive indexes.

The multilinear regression utilized in the CODESSA PRO program employs more than 800 different constitutional, geometrical,

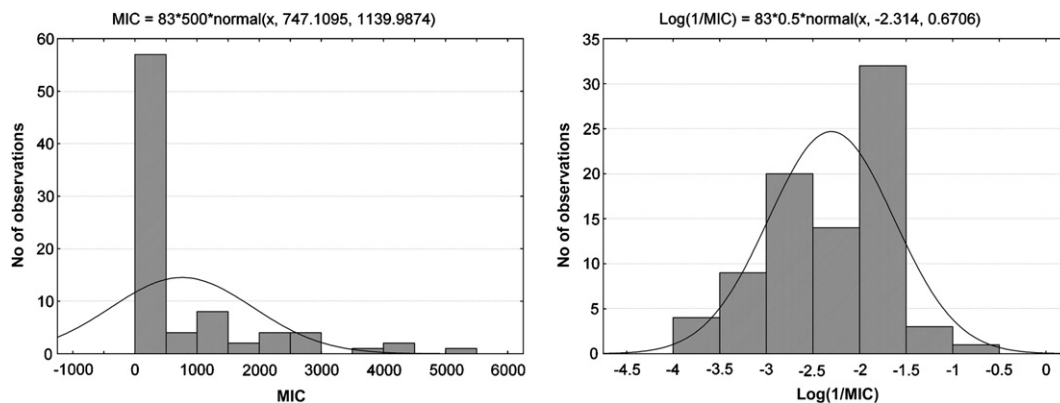


Figure 1. Original and transformed data distribution of the property values (MIC and $\log(1/\text{MIC})$).

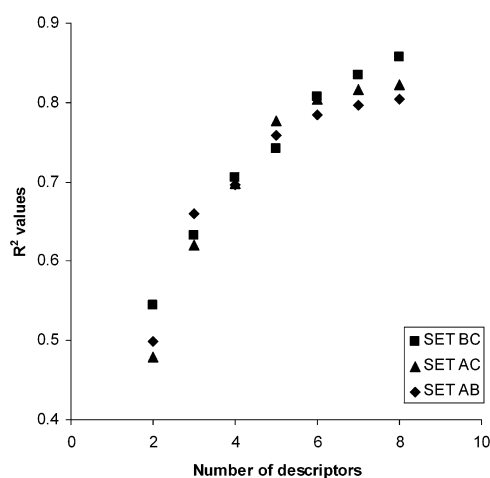


Figure 2. R^2 values versus number of descriptors used for the submodels.

topological, electrostatic, quantum chemical, and thermodynamic molecular descriptors.²⁶ Topological, electronic, and geometric descriptors contain information about (i) the number, type, and

connectivity of atoms in the molecule, (ii) the electronic aspects of the structures and (iii) the three-dimensional structure of a compound, respectively. Finally, hybrid descriptors such as charged partial surface area (CPSA) descriptors combine information about both geometric and electronic aspects of a molecule. Since they are important pharmacokinetic parameters the $\log P$ and $(\log P)^2$ values calculated by the EpiSuite package²⁷ were included as external descriptors. The $(\log P)^2$ values were introduced because lipophilicity tends to follow a parabolic relationship relative to activity as shown by Hansch.²⁸ In our treatment, all descriptors are derived solely from molecular structure and do not require experimental data.

4. Molecular modeling

The optimization procedure comprised: (i) 2D to 3D conversion using parameters included in HyperChem 7.5²⁹; (ii) initial optimization using MM+; (iii) conformational search by the 'random walk' method integrated in HyperChem; (iv) an optimization using a Polak-Ribiere conjugated gradient with RMS 0.01 kcal/mol at AM1 level of the quantum theory.

The optimized geometries were loaded into CODESSA PRO software and single point calculations were performed.³⁰ As a final

Table 2
QSAR models for AB, AC, and BC subsets

Subsets	No.	B	ΔB	t-Test	Descriptors
AB	0	-124.2	9.984	-12.44	Intercept
$R^2 = 0.785$	1	42.39	3.495	12.13	Min valency of a N atom
$R_{cv}^2 = 0.713$	2	-0.02660	0.004576	-5.813	$(\log P)^2$
$F = 29.8$	3	-9.972	2.057	-4.848	Molecular volume/XYZ Box
$s^2 = 0.145$	4	-6.273	1.407	-4.459	RNCG relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]
Test C	5	-39.88	9.158	-4.354	Minimum partial charge for a H atom [Zefirov's PC]
$R_{ext}^2 = 0.692$	6	91.54	37.26	2.457	Average electrophilic reactivity index for a C atom
AC	0	-123.7	20.10	-6.155	Intercept
$R^2 = 0.805$	1	9.677	1.113	8.692	Relative number of C atoms
$R_{cv}^2 = 0.746$	2	-0.06677	0.01025	-6.513	HDCA H-donors charged surface area [Semi-MO PC]
$F = 33.1$	3	121.5	20.04	6.065	Average valency of a H atom
$s^2 = 0.126$	4	-8.178	1.416	-5.776	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]
Test B	5	-0.01678	0.003659	-4.585	$(\log P)^2$
$R_{ext}^2 = 0.750$	6	143.6	54.13	2.653	Average electrophilic reactivity index for a C atom
BC	0	-121.9	19.31	-6.313	Intercept
$R^2 = 0.808$	1	9.618	1.143	8.416	Relative number of C atoms
$R_{cv}^2 = 0.741$	2	119.6	19.20	6.226	Average valency of a H atom
$F = 33.5$	3	-0.07316	0.01184	-6.177	HDCA H-donors charged surface area [Semi-MO PC]
$s^2 = 0.118$	4	-7.804	1.349	-5.784	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]
Test A	5	-0.01529	0.003540	-4.319	$(\log P)^2$
$R_{ext}^2 = 0.760$	6	168.7	41.79	4.037	Average electrophilic reactivity index for a C atom

Table 3

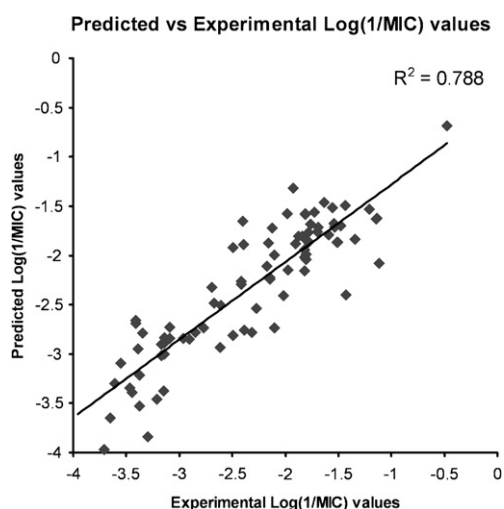
General QSAR model and its statistical parameters (the descriptors are ordered according to their significance)

No.	B	ΔB	t-Test	Descriptor name
0	−114.5	16.19	−7.073	Intercept
1	9.029	0.9475	9.530	Relative number of C atoms
2	−0.06771	0.009373	−7.223	HDCA H-donors charged surface area [Semi-MO PC]
3	112.3	16.11	6.971	Average valency of a H atom
4	−7.501	1.139	−6.587	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]
5	−0.01565	0.003062	−5.112	(logP) ²
6	146.0	35.95	4.061	Average electrophilic reactivity index for a C atom

Table 4

QSAR models for the four different classes of antifungal agents

No.	B	ΔB	t-Test	Descriptors
$R^2 = 0.796$; $R_{cv}^2 = 0.697$; $F = 18.485$; $s^2 = 0.090$; (24 fluconazoles; structures 45–68 in Table 1)				
0	94.52	51.05	1.852	Intercept
1	−40.53	5.699	−7.113	Relative number of O atoms
2	−0.2802	0.1557	−1.800	Max electron–electron repulsion for a F atom
3	−0.05198	0.007319	−7.102	(logP) ²
4	−23.57	4.556	−5.173	Relative number of N atoms
$R^2 = 0.740$; $R_{cv}^2 = 0.576$; $F = 13.257$; $s^2 = 0.016$ (18 carbonylaminobenzoxazoles; structures 27–44 in Table 1)				
0	−394.4	109.6	−3.599	Intercept
1	0.07387	0.01728	4.275	Structural information content (order 1)
2	−5.737	1.459	−3.932	RPCG relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]
3	1.203	0.3356	3.583	Max electron–nuclei attraction for a C–N bond
$R^2 = 0.950$; $R_{cv}^2 = 0.932$; $F = 100.872$; $s^2 = 0.031$ (26 cyanoboranes; structures 1–26 in Table 1)				
0	46.00	9.200	5.000	Intercept
1	−138.30	9.742	−14.20	HASA-2/TMSA [Semi-MO PC]
2	10.92	1.411	7.742	Relative number of C atoms
3	0.4811	0.08278	5.812	Number of triple bonds
4	−51.01	9.289	−5.492	Maximum bond order of a H atom
$R^2 = 0.875$; $R_{cv}^2 = 0.743$; $F = 25.692$; $s^2 = 0.063$ (15 imidazolymethylindoles; structures 69–83 in Table 1)				
0	−33.73	6.58	−5.128	Intercept
1	−1.183	0.1485	−7.963	Number of Br atoms
2	18.19	3.393	5.360	Maximum atomic orbital electronic population
3	−12.46	3.279	−3.801	Molecular volume/XYZ Box

**Figure 3.** Predicted versus observed log(1/MIC) values for the general set.

step, more than 800 theoretical descriptors were calculated and stored for further treatment.

5. Multilinear regression modeling

An important stage of the multilinear regression QSAR methodology is the search for the best multilinear equation among a given

descriptor set, especially when using large number of descriptors. In other words, the aim is to find the best correlation of the MIC values (A) with a certain number, n , of significant molecular descriptors (D_i) weighted by the regression coefficients b_i , as defined by Eq. 1

$$A = b_0 + \sum_{i=1}^n b_i D_i \quad (1)$$

The best multilinear regression technique (BMLR) encoded in CODESSA PRO software was used to select significant descriptors for building multilinear QSAR equations. In the BMLR treatment, the multilinear regression analysis commences with the whole set of two descriptors with pair $R_{ij}^2 < 0.05$, that is, with all the orthogonal or nearly orthogonal pairs of descriptors. The best two-parameter correlation from this set is found simply by performing the treatment of the property with each of the pairs. In the correlation treatment of the third rank (i.e., with three independent variables involved), each of the orthogonal descriptor pairs discussed above is combined with each of the other non-collinear ($R_{ij}^2 < 0.5$) descriptors, and again the best correlation is chosen. Analogously, in the treatment of the fourth and higher ranks, the 1000 best variable sets for the previous rank are considered in combination with each of the additional non-collinear descriptor scales. To speed up the calculations all descriptors with insignificant variance within the present dataset are rejected. This significantly decreased the probability of including unrelated descriptors by chance.

We employed a combination of multilinear regression and forward selection procedures to develop physically meaningful multi-

linear QSAR equations containing a limited number of adjustable parameters from the very large descriptor space.

6. QSAR procedure

In this work a modified approach for QSPR modeling is used, with the aim to combine the advantages of both the QSPR methods most frequently applied, that is, (i) using all available data points to build the model and to apply as the sole validation the standard internal crossvalidation procedure or (ii) to use only a part of the available data for building the model, keeping the remaining data points for external validation. Our recommended procedure to build a QSPR model is as follows:

1. All data points of the full dataset were ordered in the descending order of $\log(1/\text{MIC})$ values.
2. The initial set was separated into three subsets (conditionally denoted as A, B, and C) by selection of every third point from the original dataset in order to obtain a similar distribution of the investigated property values for each subset, A, B, C.
3. Three new datasets were constructed using the three binary sums combinations: A+B, A+C, and B+C.
4. The standard QSAR modeling procedure including best multiple linear regression method (BMLR) was applied to those three datasets obtained in Step 3.
5. The complementary parts to each of these three datasets (C, B, and A, respectively) were used as external validation datasets by considering their consistency.
6. All the descriptors that appeared in the obtained models of Step 4 were tested to obtain a general model including all the existing compounds.
7. The general model was again validated using classical internal crossvalidation and scrambling procedures.

All available information about the mechanism of action of the compounds studied and the biochemistry of the inhibitory process was taken into account when selecting the final QSAR model. In other words, to choose the most reliable model among all possible others, we relied on the physicochemical nature of the process as well as on purely mathematical criteria.

7. Results and discussion

The procedure described above was applied to the complete dataset of 83 compounds. The three subsets constructed contained, respectively, 56, 55, and 55 compounds. In each case, the remaining 27, 28, and 28 compounds, respectively, were used as external validation datasets. 'Classical' QSAR models with up to 8 descriptors were generated using the BLMR procedure. Although application of the breaking rule did not provide an unambiguous breakpoint, inspection of Figure 2 suggests two possibilities. As can be seen from Figure 2 the R^2 values for the subsets are closest to each other (R_{AB}^2 , R_{AC}^2 and R_{BC}^2) in the cases where 4 or 6 descriptors are used. The statistical parameters for the 6 descriptors models were preferred as they were significantly higher than those for the four-parameters models.

The square of the correlation coefficient (R^2), the crossvalidated correlation coefficient (R_{cv}^2), the Fisher criterion—(F), the squared standard deviation (s^2), and the R^2 predictive (R_{ext}^2) were used as criteria for the stability and the robustness of the models. B and ΔB represent the regression coefficients and their errors, respectively.

The descriptors in Table 2 for the submodels of datasets A+B, A+C, and B+C are similar, with small differences due to the procedure applied for the descriptor selection in the BLMR method.

Namely, only one of a pair or set of highly intercorrelated descriptors is used in further development of the model. Depending on the dataset, different but physically similar and highly intercorrelated descriptors may, therefore, appear in different models.

In the next step of the modeling process, only those descriptors contained in these submodels were treated further to build a general model for all 83 compounds (Table 3, Fig. 3). The calculated statistical parameters that define the quality of the final model and its predictive power are shown below.

$$R^2 = 0.788; F = 47.144; s^2 = 0.130; R_{cv}^2 = 0.749$$

To examine the sensitivity of the proposed QSAR model to chance correlations a scrambling procedure was applied, that is, the model was fitted to randomly reordered activity values and then compared with the one obtained for the actual activities.³¹ Twenty randomizations resulting in average $R^2 = 0.309$ were performed. The substantial difference between the actual R^2 and the averaged R^2 from the scrambling procedure proves the stability of the model.

In the model of Table 3, compounds **51**, **52**, **72**, and **78** had deviations slightly greater than twice the standard deviation of the data; however, their removal did not significantly influence the statistical parameters of the model; they were therefore not considered as outliers.

QSAR models including non-congener series of compounds are more complicated from the viewpoint of their interpretation than the QSARs based on congener series. Thus, we considered all apparently ambiguous facts, which can help model interpretation.

An important descriptor in our model is the 'relative number of C atoms' which is defined by the ratio between the carbon atoms and all the other atoms in the molecule. The descriptor value depends in a complex way on the number of the double and triple bonds and the presence of heteroatoms in the molecules. Hence, more double or triple bonds and less heteroatoms will enhance the property. As can be seen from Table 1, among the most active compounds are those with highly conjugated systems and very few heteroatoms (usually less than 5).

The next important descriptor for the explanation of the antifungal effect is the 'H-donors charged surface area'. As is well known, the ligand–receptor interactions are non-covalent and thus dependent on the donor–acceptor abilities of the compounds. The larger the charged area on the hydrogen bonding donor atoms the weaker is the antifungal effect. One possible explanation is that larger overlapping between the hydrogen donor atoms and the H atoms decreases the possibility for the creation of stronger donor–acceptor bonds.

The 'average valence of a H atom' descriptor is a measure of the stability of bonds formed between the hydrogen atoms and the molecular backbone. The plus sign of the regression coefficients in this case indicates that the more stable the compounds the higher their antifungal effect.

The chemical individuality of bioactive organic compounds mainly depends on the heteroatoms present in their molecules. Thus, the presence in the model of the 'relative negative charge' descriptor in the model is probably connected to the heteroatom binding affinity. The negative coefficient of RNCG in the model implies that the increase of the descriptor value will lower the antifungal effect; for example, for the following pair of compounds—**73**, **74** and also **75**, **77** a chlorine atom has a greater positive influence than fluorine on the biological effect.^{4,18}

The presence of the $(\log P)^2$ descriptor in the model suggests that the transport of the compounds through the cellular membranes contributes significantly to the explanation of the data variance. According to the proposed model, higher $(\log P)^2$ values lead to decreased antifungal effect. In support of our results, the same

approach (applying a parabolic function) was used by Lien and Wang³² to improve the fit of Hansch equations for the antifungal activities of homologous aliphatic amines.

The ‘average electrophilic reactivity index for a C atom’ measures the stabilization in energy when the system acquires an additional electronic charge ΔN , thus describing the electron flow between the donor and the acceptor. Due to the positive sign of the regression coefficient, the higher the descriptor value (the molecule will act as a stronger electrophile) the higher its antifungal effect.

Individual multilinear models were also obtained for the cyanoborane, fluconazole, carbonylaminobenzoxazole, and imidazolylmethylindole datasets. The number of the descriptors in the generated submodels was limited, respectively, to 3 or 4 according to the ‘maximum one descriptors per five objects’ rule. Since, these submodels consider specific features for each of the four chemical classes presented they give a more detailed picture of the existing interactions than the general model. On detailed consideration, the general model and the submodels presented in Table 4 are quite similar. As can be seen (Table 4) most of the descriptors are related to specific atoms types: O, F, N, C, and Br. As discussed above, the heteroatoms define the unique features of the organic compounds, and due to their electronegativity/electron pairs are usually responsible for the surface recognition and for establishing the donor–acceptor bonds. The remaining descriptors are electrostatic and likely describe the donor–acceptor (non-covalent) interactions between ligands and receptors. The last two descriptors $(\log P)^2$ and ‘molecular volume/XYZ Box’ depict, respectively, the transport properties of the compounds and the steric effects.

It can be concluded that the individual models follow the same tendency as the general model showing the prevailing influence of electrostatic effects (all descriptors with high ‘t-test’ values) over the transport abilities and the steric effects.

8. Conclusion

Classical QSAR approach was applied successfully to a series of 83 compounds with well-expressed antifungal activity. A general QSPR model with six theoretical molecular descriptors was obtained. An external validation was performed aiming to evaluate the predictive ability of the model. All descriptors involved were calculated solely from the chemical structure of the compounds and have definite biochemical meaning corresponding to the nature of the antifungal drugs action. Our work fully supports and extends the results obtained earlier by other workers.^{4,18,30}

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.05.014.

References and notes

- Walsh, T. J.; Gonzalez, C.; Roilides, E.; Mueller, B. U.; Ali, N.; Lewis, L. L.; Whitcomb, T. O.; Marshall, D. J.; Pizzo, P. A. *Clin. Infect. Dis.* **1995**, 20, 900–906.
- Jalilian, A. R.; Sattari, S.; Bineshmarvasti, M.; Shafiee, A.; Daneshdatab, M. *Arch. Pharm. Pharm. Med. Chem.* **2000**, 333, 347–354.
- Takrouiri, K.; Oren, G.; Polacheck, I.; Sionov, E.; Shalom, E.; Katzhendler, J.; Srebnik, M. *J. Med. Chem.* **2006**, 49, 4879–4885.
- Nam, N.; Sardari, S.; Selecky, M.; Parang, K. *Bioorg. Med. Chem.* **2004**, 12, 6255–6269.
- Kauffman, C. A.; Carver, P. L. *Drugs* **1997**, 53, 539–549.
- Schmitt, H. J.; Edwards, F.; Andrade, J.; Niki, Y.; Armstrong, D. *Chemotherapy* **1992**, 38, 118–126.
- Zakula, D.; Capobianco, J. O.; Goldman, R. C. *J. Antimicrob. Chemother.* **1997**, 39, 261–264.
- Hawser, S.; Islam, K. J. *Antimicrob. Chemother.* **1999**, 43, 411–413.
- Jones, T.; Federspiel, N. A.; Chibana, H.; Dungan, J.; Kalman, S.; Magee, B. B.; Newport, G.; Thorstenson, Y. R.; Agabian, N.; Magee, P. T.; Davis, R. W.; Scherer, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 7329–7334.
- Hasegawa, K.; Morikami, K.; Shiratori, Y.; Ohtsuka, T.; Aoki, Y.; Shimma, N. *Chemom. Intell. Lab. Syst.* **2003**, 69, 51–59.
- Purushottamachar, P.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2003**, 11, 3487–3497.
- 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, 1227–1235.
- Waisser, K.; Perina, M.; Buchta, V.; Kubanova, P. *Folia Pharm. Univ. Carolinae* **2003**, 29, 21–33.
- Souad, M.; Didier, V.; Mohammed, M.; Abderrahim, J.; Driss, C. *Fresenius Environ. Bull.* **2000**, 9, 734–745.
- Soni, S. K.; Agrawal, V. K.; Khadikar, P. V.; Srivastava, A. K. *Indian J. Chem., Sect A* **1999**, 38, 745–752.
- Mahfouz, N. M.; Moghazy, S. M. *Alex. J. Pharm. Sci.* **1995**, 9, 21–27.
- Arpaci, Ö. T.; Ören, İ.; Altanlar, A. *Il Farmaco* **2002**, 57, 175–181.
- Na, Y.-M.; Le Borgne, M.; Pagniez, F.; Le Baut, G.; Le Pape, P. *Eur. J. Med. Chem.* **2003**, 38, 75–87.
- Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T. J. *Chem. Inf. Comput. Sci.* **1998**, 38, 720–725.
- Katritzky, A. R.; Karelson, M.; Lobanov, V. S. *Pure Appl. Chem.* **1997**, 69, 245–248.
- Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, 24, 279–285.
- Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. *Collect. Czech. Chem. Commun.* **1999**, 64, 1551–1557.
- Katritzky, A. R.; Taemm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 136–142.
- Basak, S. C.; Mills, D. *ARKIVOC* **2005**, 2, 60–76.
- Thakur, A. *ARKIVOC* **2005**, 14, 49–58.
- Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.
- <http://www.epa.gov/opptintr/exposure/pubs/episuite.htm>.
- Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, 86, 1616–1626.
- Hyperchem, v. 7.5; Hypercube Inc., Gainesville, FL. <http://www.hyper.com/>.
- CODESSA PRO Software, University of Florida, 2002. www.codessa-pro.com/.
- Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. *Environ. Health Perspect.* **2003**, 111, 1361–1375.
- Lien, E. J.; Wang, P. H. *J. Pharm. Sci.* **1980**, 69, 648–650.