CARPATHIAN J. MATH.
20 (2004), No. 2, 267 - 274

QSARs of some Novel Isosteric Heterocyclic with Antifungal Activity

OLEG URSU, ADINA COSTESCU, MIRCEA V. DIUDEA and BAZIL PARV

ABSTRACT. QSAR analysis of a set of 2,5,6-trisubstituted benzoxazoles, benzimidazoles and 2-substituted oxazolo(4,5-b)pyridines, tested for growth inhibitory activity against Candida albicans, was performed by using a multiple regression procedure. The activity contributions for either heterocyclic ring systems or substituent effects of these compounds were determined from the correlation equations and the lead optimization is described. The obtained QSAR revealed that the oxazolo(4,5-b)pyridine ring, substituted by a benzyl moiety at position 2, was the most favorable structure among the heterocyclic nuclei. Moreover, the fifth position in the fused ring system is found more significant than the other positions in rising up the activity.

1. INTRODUCTION

Mycosis diseases are affections due to the body infest with pathogenic mycosis parasites of pathogenic potential. Some of these fungi live in nature like saprophytes (exogenetic saprophytes) or in the human body or animals (endogenetic saprophytes) and they accidentally become parasites.

Affections produced by the mycosis parasites are widely spread in the whole world, with predominance in some areas. Superficial mycosis affections are frequently encountered in our country. *Candidiasis*, the infection caused by *Candida*, a yeast-like fungus, is such affection.

Candida species are ubiquitous fungi, existing as a normal body flora. *Candidiasis* infection causes a variety of diseases, ranging from superficial disorders such as diaper rash to invasive, rapidly fatal infections in immunocompromised hosts. *Candida albicans* is commonly responsible for candidiasis, but others, such as *Candida tropicalis, Candida parapsilosis, Candida guilliermondi,* and *Torulopsis glabrata* are also causative organisms [1].

Candidiasis affects a wide variety of organ systems. In immunocompetent persons, any warm, moist part of the body exposed to the environment is susceptible to infection. In immunocompromised patients, systemic illnesses such as myocarditis, hepatosplenic abscess occur and antifungal therapy should be started immediately after necessary cultures have been obtained from all suspected sites of infection.

Drug Category: *Antifungals* Amphotericin B, fluconazole, ketoconazole, and nystatin are the drugs most commonly used to treat candidiasies. Most fungi are completely

Received: 26.09.2004; In revised form: 17.01.2004
2000 *Mathematics Subject Classification.* 92E10
Key words and phrases. *QSAR,, PCA, antifungal, topological descriptors*

resistant to the action of antimicrobial drugs. Only a few substances have been discovered which exert an inhibitory effect on the fungi pathogenic for man, and most of these are relatively toxic [2].

In this context, the synthesis of some novel derivatives of benzoxazoles, benzimidazoles, oxazolo (4,5-b) pyridines and benzothiazoles during the last few years was welcome [3-7].

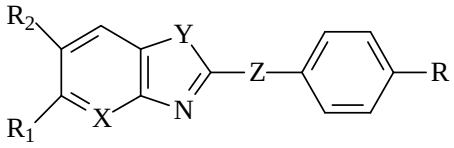
2. DATA SET

The compounds **1–68**, taken in this *QSAR* analysis, showed antifungal activity of MIC (minimum inhibitory concentration) values, between 6.25–50 g/ml, against *Candida albicans*. Compare these data with the activity of of commonly used antifungals, such as Clotrimazole, Oxiconazole and Haloprogin (below) [5].

Compound	Observed *MIC* values
1–68	6.25–50 g/ml
Clotrimazole	6.25 g/ml
Oxiconazole	6.25 g/ml
Haloprogin	3.12 g/ml

In this study, QSAR analysis of some novel antifungal benzoxazoles, benzimidazoles and oxazolo (4,5-b) pyridines **1–68** is presented. Table 1 lists the formulas and the activity data for the above mentioned structures.

Table 1. The structure and *in vitro* antifungal activity of compounds 1–68 against *C. albicans*



	X	Y	Z	R	R_1	R_2	$\log(1/C)$
1	CH	O	-	H	H	H	3.892
2	CH	O	-	$C(CH_3)_3$	H	H	4.001
3	CH	O	-	NH_2	H	H	3.924
4	CH	O	-	$NHCH_3$	H	H	3.952
5	CH	O	-	C_2H_5	Cl	H	4.013
6	CH	O	-	$NHCOCH_3$	Cl	H	4.059
7	CH	O	-	$NHCH_3$	Cl	H	4.015
8	CH	O	-	Cl	Cl	H	4.024
9	CH	O	-	NO_2	Cl	H	4.04
10	CH	O	-	H	NO_2	H	4.282
11	CH	O	-	CH_3	NO_2	H	4.308
12	CH	O	-	$C(CH_3)_3$	NO_2	H	4.375
13	CH	O	-	NH_2	NO_2	H	4.31

Table 1 (continued)

14	CH	O	-	Cl	NO_2	H	4.342
15	CH	O	-	Br	NO_2	H	4.406
16	CH	O	-	C_2H_5	NH_2	H	3.979
17	CH	O	-	F	NH_2	H	3.96
18	CH	O	-	$N(CH_3)_2$	CH_3	H	4.005
19	CH	O	-	CH_3	CH_3	H	3.95
20	CH	O	-	C_2H_5	CH_3	H	3.977
21	CH	O	-	OCH_3	CH_3	H	3.98
22	CH	O	-	F	CH_3	H	3.958
23	CH	O	-	$NHCOCH_3$	CH_3	H	4.027
24	CH	O	-	$NHCH_3$	CH_3	H	3.979
25	CH	O	-	$N(CH_3)_2$	CH_3	H	4.004
26	N	O	-	CH_3	H	H	4.225
27	N	O	-	C_2H_5	H	H	4.253
28	N	O	-	OCH_3	H	H	4.257
29	N	O	-	OC_2H_5	H	H	4.283
30	N	O	-	NH_3	H	H	4.227
31	N	O	-	NO_2	H	H	4.285
32	CH	O	CH_2	H	H	H	4.223
33	CH	O	CH_2	OCH_3	H	H	4.282
34	CH	O	CH_2	Cl	H	H	4.29
35	CH	O	CH_2	NO_2	H	H	4.308
36	CH	O	CH_2	H	Cl	H	4.29
37	CH	O	CH_2	OCH_3	Cl	H	4.34
38	CH	O	CH_2	Br	Cl	H	4.41
39	CH	O	CH_2	NO_2	Cl	H	4.363
40	CH	O	CH_2	H	NO_2	H	4.609
41	CH	O	CH_2	OCH_3	NO_2	H	4.657
42	CH	O	CH_2	Br	NO_2	H	4.725
43	CH	O	CH_2	Cl	NO_2	H	4.664
44	CH	O	CH_2	NO_2	NO_2	H	4.68
45	CH	O	CH_2O	H	CH_3	H	3.98
46	CH	O	CH_2O	H	H	NO_2	3.732
47	CH	O	CH_2O	H	CL	NO_2	3.785
48	CH	O	CH_2O	Cl	CL	NO_2	3.831
49	CH	O	CH_2S	H	NO_2	H	4.359
50	CH	O	CH_2S	H	CH_3	H	4.009
51	N	O	CH_2O	H	H	H	4.26
52	N	O	CH_2O	Cl	H	H	4.319
53	CH	NH	CH_2O	Cl	CH_3	H	4.037
54	CH	NH	CH_2S	H	NO_2	H	4.358
55	CH	NH	CH_2S	H	CH_3	H	4.009
56	CH	O	CH_2O	H	$COOCH_3$	H	4.054
57	CH	O	CH_2O	Cl	$COOCH_3$	H	4.104
58	CH	NH	CH_2O	Cl	$COOCH_3$	H	4.102
59	CH	NH	CH_2S	H	$COOCH_3$	H	4.076
60	CH	O	C_2H_4	H	NO_2	H	4.331
61	N	O	C_2H_4	H	H	H	4.253
62	CH	O	-	Br	NH_2	H	4.11

Table 1 (continued)

63	CH	O	CH_2	Br	H	H	4.36
64	CH	O	CH_2O	Cl	H	H	4.016
65	CH	NH	CH_2O	H	NO_2	H	4.283
66	CH	NH	CH_2O	Cl	H	H	4.015
67	CH	NH	CH_2S	H	Cl	H	4.041
68	CH	NH	C_2H_4	H	H	H	4.078

3. DATA ANALYSIS

The QSAR analysis consists of the following steps: (i) structure optimization using semiempirical method PM3; (ii) calculation of molecular descriptors; (iii) classification by using Good similarity index; (iv) principal component analysis (*PCA*); (v) correlation analysis using step-forward selection of descriptors; (vi) multiple regression analysis by selected principal components; (vii) evaluation of the significance level of the model; (viii) validation of the model and (ix) interpretation of the model.

The structures were optimized by using the semiempirical *PM3* Hamiltonian, available in HyperChem. A large pool of descriptors, provided by *TOPOCLUJ* software package, was calculated for every molecule. The descriptors list includes: topological descriptors (Wiener [8], Randić [9], etc.); constitutional indicator descriptors (*e.g.*, absence or presence of certain functional groups in the molecular structure), quantum calculated descriptors (HOMO and LUMO energy, dipole moment, total electronic energy); force field parameters *F* (field effect) for electronic description and steric descriptors (L and B4) for the substituents R and R1 [10-13].

L is defined as the length (in Å) of a substituent along the axis of its substitution to the parent skeleton. B4 is defined as the maximum width of the substituent and might provide a better understanding of steric requirements in ligand-receptor interactions. Electronic effect of the substituents, expressed in terms of *F*, is found to be important in determining the activity, as it is predictive in electrophilic reactions of the drugs with the nucleophilic functions of biomolecules.

Similarity expresses the relatedness of two molecules, with a large number if their molecular descriptions are closely related and with a number going to zero in case they are unrelated. *Good* index is one of the most used coefficients of similarity [14-15]:

$$SIM(A,B)_k = 1 - \left(\frac{|I_{Ak} - I_{Bk}|}{\max|I_{Ak}, I_{Bk}|} \right) \quad (1)$$

It takes values in the range [0,1]. Considering all the *n* descriptors, the index becomes:

$$SIM(A,B) = \frac{1}{n} \sum_{k=1}^{n} w_k SIM(A,B)_k \quad (2)$$

where w_k is the weighting factor for the descriptor *k*. The descriptors selected as similarity criteria are statistically the most significant ones. In other words, they have the best correlation coefficient, in simple linear regression model, with *BA* (*e.g.*, the antifungal activity, $\log(1/C)$). The initial data set (68 antifungal compounds) is split in 3 subsets: two training subsets, (clustering the ordered *BA* values) and one external (randomly selected) validation subset (see the left column in Table 2). The quality of prediction

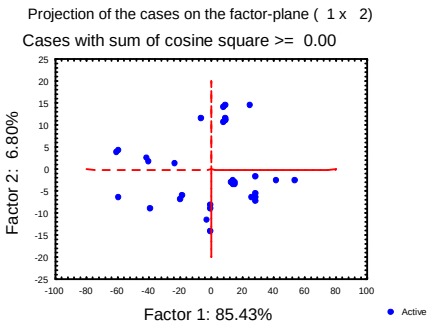
depends on correct classification of presumably "unknown" compounds with respect to the training subsets.

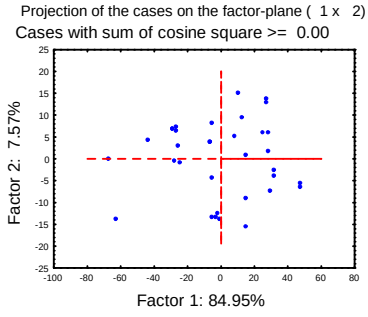
Table 2. Good frequencies for compounds in the external validation subset.

Compd.	***Subset 1*** $\sum$ *Good Frequencies*	***Subset 2*** $\sum$ *Good Frequencies*
a17	**0.7269**	0.6503
a27	0.6792	**0.6816**
a33	0.6175	**0.6222**
a35	0.6663	**0.7189**
a47	**0.6124**	0.5725
a53	**0.7599**	0.6821
a58	**0.6429**	0.5574
a65	0.7423	**0.7729**
a07	**0.7789**	0.7444

The values in Table 2 are Good index frequencies calculated in the following manner: for each compound in the external validation subset and from both training subsets, Good index is calculated according to eq. 2. The boldface values indicate the maximum and thus the subset to which belongs each compound of the external validation subset.

Principal components analysis *PCA* is a very powerful statistical technique useful to reduce the noise of the data set and to eliminate in-between correlated variables. Due to the high complexity of interactions between the receptor molecule and potential inhibitor molecules, it is often very difficult to model *BA* using simple linear regression models. In this study we used a large pool of descriptors (985) to derive *QSAR* model; thus *PCA* analysis gives a number of principal components which describes the most of variance in the data set (see Figure 1) [16].





Subset 1	*Subset 2*
Quality of representation 100%	Quality of representation 100%
no. of active cases 29	no. of active cases 30
no. of components 33	no. of components 33

Figure 1. Principal component analysis, projection on factor coordinates.

Multiple linear regressions are the simplest and mostly used models to derive a quantitative relationship between the *PC* independent variables and *BA*. Automated

selection of descriptors is performed by *forward stepwise* procedure. The derived *QSAR multivariate* models are presented in Table 3 (*b* the regression coefficients of the variables) and plots in Figure 2. Both models show very good correlation coefficient ($R^2 > 0.93$) and were tested for statistical significance, by a random exchange of *BA* column entries, within each subset; a dramatic drop-down of the correlation coefficient proved that no chance correlation occurred in our models. Next validation of the models, on the external validation dataset, tested the prediction ability of the derived QSAR models (see Figure 3). The predicted values of *BA*, with the model equations assigned cf. the classification by the Good frequencies (see above), again showed very good correlation ($R^2 > 0.92$) with the observed *BA* (available, in this case). Thus, the derived models can be used in prediction of *BA* for new compounds, in homologous series.

Table 3. Statistics for derived QSAR models for antifungal compounds in Table 1.

Subset 1

	b	*Std. Error*	*t*	*p-level*
Intercept	3.985	3.014E-03	1322.355	-
Factor 2	2.777E-02	2.993E-03	9.279	1.324E-07
Factor 3	-3.472E-02	3.119E-03	-11.132	1.198E-08
Factor 4	4.493E-02	3.232E-03	13.900	5.660E-10
Factor 5	-4.216E-02	3.784E-03	-11.144	1.181E-08
Factor 6	-7.157E-03	2.539E-03	-2.819	1.297E-02
Factor 7	2.740E-02	2.608E-03	10.504	2.605E-08
Factor 8	-1.791E-02	2.544E-03	-7.038	4.016E-06
Factor 9	-2.416E-02	3.705E-03	-6.522	9.648E-06
Factor 13	1.755E-02	2.630E-03	2.357	3.241E-02
Factor 14	1.336E-02	4.731E-03	3.710	2.095E-03
Factor 16	-8.008E-03	3.464E-03	3.857	1.551E-03
Factor 17	1.567E-02	3.016E-03	-2.655	1.800E-02

Subset 2

	B	*Std. error*	*t*	*p-level*
Intercept	4.355	8.614E-03	505.54	-
Factor 1	-0.098	8.381E-03	-11.75	1.08E-10
Factor 3	-0.042	8.402E-03	-4.95	6.74E-05
Factor 4	0.027	8.343E-03	3.29	3.52E-03
Factor 7	-0.034	8.476E-03	-4.04	5.95E-04
Factor 8	-0.026	9.036E-03	-2.84	9.80E-03
Factor 9	0.047	8.585E-03	5.52	1.77E-05
Factor 10	-0.051	8.847E-03	-5.79	9.54E-06
Factor 16	0.035	9.143E-03	3.83	9.69E-04

A rather large number of *PC* is required to model the *BA* – explanation for that resides in the fact that there is no simple mechanism of interaction between receptor molecule and ligand (active antifungal molecule). To bind to the receptor site, multiple barriers must be passed by the active and different structural features of the ligand

contribute to the binding. One single descriptor can not contain such a large quantity of information, so a rather large number of descriptors is needed to describe properly the binding and finally *BA*.

Despite the published [17] *QSAR* models, based on some physico-chemical properties or structural indicator variables, showed a rather good correlation coefficient, in the training step, in prediction, however, they are not convincing.

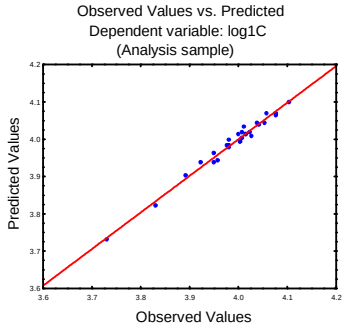
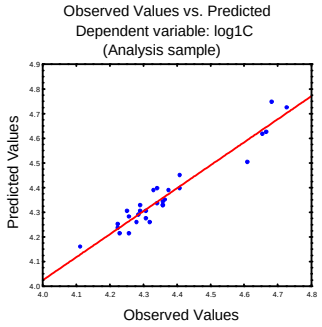
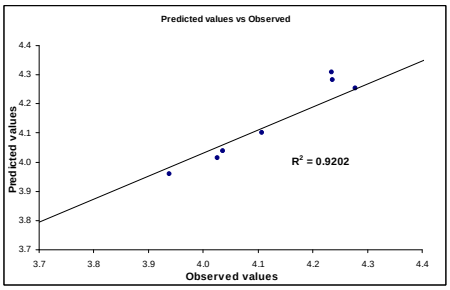
Subset 1	*Subset 2*
	
R^2 = 0.9809; s = 0.1552; F = 59.2596; p=1.79E-10	R^2 = 0.9333; s = 0.6233; F = 36.7252; p p=1.2E-10

Figure 2. Predicted vs. observed value for the training subsets.



Compd.	***Obs. BA***	***Pred. BA***	***Residual***	***CV %***
a17	3.960	3.939	0.021	0.543
a27	3.785	3.673	0.112	2.952
a33	4.037	4.036	0.001	0.017
a35	4.102	4.107	-0.005	0.117
a47	4.015	4.026	-0.011	0.268
a53	4.253	4.278	-0.025	0.591
a58	4.282	4.236	0.046	1.082
a65	4.308	4.235	0.073	1.687
a07	4.283	4.415	-0.132	3.073

Figure 3. Predicted and residual values in the validation set

4. CONCLUSIONS

In the current paper we presented a QSAR model for a set of antifungal compounds. A special attention was focused on model testing and validation. Statistical tools such as *PCA* and stepwise regression analysis have been employed. Another important aspect taken in consideration was the correct splitting of the data set into training and validation subsets, which can considerably affect the results. The derived models showed very good prediction ability and thus could be employed in future evaluation of *BA* for new compounds, in homologous series.

REFERENCES

[1] http://www.emedicine.com/emerg/byname/candidiasis.htm
[2] Meyers, F.H., Jawetz, E., and Goldfien, A., Review of Medical Pharmacology, Lange Medical Pub., California, 1976.
[3] Yalçìn, I., Ören, I., Sener, E., Akìn, A. and Uçartürk, N., Eur. J. Med. Chem. **27** (1992), 401–406.
[4] Yalçìn, I., Sener, E., Özden, T., Özden, S. and Akìn, A., Eur. J. Med Chem. **25** (1990), 705–708.
[5] Ören, I., The synthesis, structure elucidation and quantitative structure-activity relationship analysis of some novel antimicrobial active isosteric heterocyclic compounds, Ph.D. Thesis, University of Ankara, 1995.
[6] Ören, I., Temiz, Ö., Yalçìn, I., Sener, A. and Altanlar, N., Eur. J. Pharm. Sci. **7** (1998), 153–160.
[7] Mrozek, A., Trzezwinska, H., Karolak-Wojciechowska, J., Yalçìn, I., Tener, E., Polish J. Chem. **73** (1999), 625–633.
[8] Wiener, H., J. Am. Chem. Soc. **69 (**1947), 17-20.
[9] Randić, M., J. Amer. Chem. Soc. **97** (1975), 6609-6615.
[10] Hansch, C. and Fujita, T., J. Am. Chem. **86 (**1964), 1616.
[11] Hansch, C. and Fujita, T., Acc. Chem. Rev. **2** (1969)**,** 232.
[12] Verloop, A., Hoogenstraten, W. and Tipker, J., Drug Design **7 (**1976), 165.
[13] Hansch, C. and Leo, A., *Substituent Constant for Correlation Analysis in Chemistry and Biology***,** John Wiley and Sons, New York, 1979.
[14] Good, A. C., J. Mol. Graphics **10 (**1992), 114.
[15] Son, S. H., Han, Ch. K., Ahn, S. K., Yoon, J. H., No, K. T., J. Chem. Inf. Comput. Chem. **39** (1999), 601- 609.
[16] StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com
[17] Yalçìn, I., Ören, I., Temiz O. and Sener, A., Acta Biochimica Polonica **47** (2000), 481-486.

BABEŞ - BOLYAI UNIVERSITY
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING
ARANY JÁNOS 11, 40028, CLUJ,
ROMANIA

BABEŞ - BOLYAI UNIVERSITY
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
400084, CLUJ,
ROMANIA