Scientific paper

Multivariate Regression Modelling of Antifungal Activity of Some Benzoxazole and Oxazolo[4,5-*b*]pyridine Derivatives

Strahinja Z. Kovačević,* Sanja O. Podunavac Kuzmanović and Lidija R. Jevrić

University of Novi Sad, Faculty of Technology, Department of Applied and Engineering Chemistry, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia

** Corresponding author: E-mail: strahinjakovacevic@hotmail.com; phone: +381642839686; fax: +38121450413*

Received: 28-03-2013

Abstract

In the present study, principal component analysis (PCA) followed by principal component regression (PCR) and partial least squares (PLS) method was applied in order to identify the most important *in silico* molecular descriptors and quantify their influence on antifungal activity (expressed as minimal inhibitory concentration) of selected benzoxazole and oxazolo[4,5-*b*]pyridine derivatives against *Candida albicans*. PLS regression showed the best statistical performance, according to the lowest value of the standard error (root mean square errors of calibration of 0.02526 and *cross*-validation of 0.04533), while PCR model was characterized by root mean square errors of calibration of 0.03176 and *cross*-validation of 0.05661. The most important descriptors in both PLS and PCR model are solubility in water, expressed as AClog*S* and ABlog*S*, and lipophilicity, expressed as Xlog*P*2 and ABlog*P*. Very good predictive ability of the established models, confirmed by corresponding statistical parameters, allows us to estimate antifungal activity of structurally similar compounds.

Keywords: QSAR analysis; Principal component regression; Partial least squares; *Candida albicans*; Heterocyclic compounds.

1. Introduction

Candida albicans is one of the most common fungal opportunistic pathogen of humans that can cause local, systemic and superficial mucosal infections (especially gastrointestinal, oral, respiratory and genital infections) in immunocompromised individuals, such as patients suffering from AIDS, leukemia or diabetes. Candidiasis, an infection caused by *Candida* species, is usually treated with antifungal drugs: amphotericin B, fluconazole, ketoconazole and nystatin.[1] Because of the increasing incidence of both fungal infections and antifungal drug resistance, synthesis and analysis of some novel antifungal compounds are welcome.

Benzoxazoles and oxazolo[4,5-*b*]pyridines, as analogues of benzimidazole, are well known to the chemists, mainly due to their wide spectrum of antimicrobial properties.[2–6] It is also determined that these molecules are present in a variety of herbicidal, antihelmintic, antioxidant and antitumoral agents.[7–10] Antimicrobial activity of studied compounds is expressed as minimal inhibitory concentration (MIC) defined as the lowest concentration of the compound at which no growth of the strain is observed in time and under specified experimental conditions.

Prediction of antimicrobial activity of compounds, based on their structural characteristics, is a very important and fundamental issue of pharmaceutical chemistry. Quantitative structure-activity relationships (QSAR) analysis allows us to estimate the biological activity of novel molecules prior to their synthesis, according to statistically significant mathematical models based on a large number of already synthesized molecules.[11,12] QSAR studies are widely applied in quantitative description of relationships between the chemical structure of a drug molecule and its biological activity, aiming at defining optimal values for some physicochemical properties of the molecule and providing the fundamentals for design of new substances as drug candidates, according to current

needs.[13,14] Beside importance of QSAR modelling in drug design, *in silico* methods are important contributors to drug discovery processes.[15]

In the present paper, antifungal activity of eighteen benzoxazole derivatives and six oxazolo[4,5-*b*]pyridine derivatives was estimated according to their *in silico* molecular descriptors using principal component analysis (PCA), followed by principal component regression (PCR) and partial least squares (PLS) chemometric methodologies. In our previous work[11], we have already studied the influence of some molecular descriptors of benzoxazoles on their *in vitro* antifungal activity against *Candida albicans* using multiple linear regression (MLR), therefore the novelty of the present study is the extended series of the studied compounds and the application of PCR and PLS for the same purpose.

2. Materials and Methods

The QSAR analysis was performed in the following several steps: molecular structure optimization by computer software, structural descriptors computation, structural descriptors selection, structure-activity model generation using PCR and PLS methods, and statistical validation.

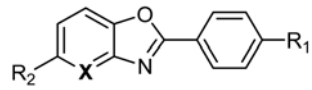
2. 1. Studied Compounds

The structures of benzoxazoles and oxazolo[4,5-*b*] pyridines investigated in this paper are presented in Table 1. The results of their *in vitro* antifungal activity against *Candida albicans* (MTCC 183) are presented in literature.[16] The logarithm of molar MIC ($\log(1/c_{MIC})$) was used for further calculations (Supporting information, Table S1).

2. 2. Molecular Modelling and Molecular Descriptors

In silico modeling of examined molecules was performed by using following software: CS ChemBioDraw Ultra 12.0 for drawing 2D structures of molecules, and CS ChemBio3D Ultra 12.0 for 3D molecular modelling running on AMD Sempron Processor 3000+.[17] The constructed 3D models were subjected to energy minimization using molecular mechanics force field method (MM2). The cutoff for structure optimization was set at a gradient of 0.1 kcal/Åmol. The Austin Model 1 (AM1) was used for full geometry optimization of all structures until the root mean square (RMS) gradient reached a value smaller than 0.0001 kcal/Åmol using Molecular Orbital Package (MOPAC) program.[18]

Table 1. The IUPAC names and chemical structures of the compounds studied.



Comp.	IUPAC name	X	R_1	R_2
1	2-phenyl-1,3-benzoxazole	CH	H	H
2	2-(4-tert-butylphenyl)-1,3-benzoxazole	CH	$C(CH_3)_3$	H
3	4-(1,3-benzoxazol-2-yl)aniline	CH	NH_2	H
4	4-(1,3-benzoxazol-2-yl)-N-methylaniline	CH	$NHCH_3$	H
5*	5-chloro-2-(4-ethylphenyl)-1,3-benzoxazole	CH	C_2H_5	Cl
6	N-[4-(5-chloro-1,3-benzoxazol-2-yl)phenyl]acetamide	CH	$NHCOCH_3$	Cl
7	4-(5-chloro-1,3-benzoxazol-2-yl)-N-methylaniline	CH	$NHCH_3$	Cl
8	5-chloro-2-(4-chlorophenyl)-1,3-benzoxazole	CH	Cl	Cl
9*	5-chloro-2-(4-nitrophenyl)-1,3-benzoxazole	CH	NO_2	Cl
10	2-(4-ethylphenyl)-1,3-benzoxazol-5-amine	CH	C_2H_5	NH_2
11*	2-(4-fluorophenyl)-1,3-benzoxazol-5-amine	CH	F	NH_2
12	5-methyl-2-(4-methylphenyl)-1,3-benzoxazole	CH	CH_3	CH_3
13	2-(4-ethylphenyl)-5-methyl-1,3-benzoxazole	CH	C_2H_5	CH_3
14	2-(4-methoxyphenyl)-5-methyl-1,3-benzoxazole	CH	OCH_3	CH_3
15	2-(4-fluorophenyl)-5-methyl-1,3-benzoxazole	CH	F	CH_3
16	N-[4-(5-methyl-1,3-benzoxazol-2-yl)phenyl]acetamide	CH	$NHCOCH_3$	CH_3
17*	N-methyl-4-(5-methyl-1,3-benzoxazol-2-yl)aniline	CH	$NHCH_3$	CH_3
18	N,N-dimethyl-4-(5-methyl-1,3-benzoxazol-2-yl)aniline	CH	$N(CH_3)_2$	CH_3
19	2-(4-methylphenyl)-[1,3]oxazolo[4,5-*b*]pyridine	N	CH_3	H
20	2-(4-ethylphenyl)-[1,3]oxazolo[4,5-*b*]pyridine	N	C_2H_5	H
21	2-(4-methoxyphenyl)-[1,3]oxazolo[4,5-*b*]pyridine	N	OCH_3	H
22	2-(4-ethoxyphenyl)-[1,3]oxazolo[4,5-*b*]pyridine	N	OC_2H_5	H
23*	4-{[1,3]oxazolo[4,5-*b*]pyridin-2-yl}aniline	N	NH_2	H
24*	2-(4-nitrophenyl)-[1,3]oxazolo[4,5-*b*]pyridine	N	NO_2	H

*External test set

The values of molecular descriptors for each compound in the data set were calculated using the software CS ChemBio3D Ultra 12.0 and ALOGPS 2.1.[17,19] These data are presented in Table S2 (Supporting information). Determined descriptors of the examined molecules were physicochemical descriptors (relative molecular mass – Mr, boiling point – BP [K], melting point – MP [K], critical temperature – CT [K], critical pressure – CP [bar], critical volume – CV [cm^3/mol], partition coefficients for *n*-octanol/water bi-phase system – Alog*P*s, ACDlog*P*, ABlog*P*, milog*P*, Alog*P*, Mlog*P*, Xlog*P*2 and Xlog*P*3, solubility in water – ABlog*S* and AClog*S*) and molecular bulkiness descriptors (molar refractivity – MR [cm^3/mol], van der Waals surface area – vdWSA [$Å^2$]). Other types of molecular descriptors of benzoxazoles and oxazolo[4,5-*b*]pyridines (electrostatic and topological descriptors) were applied as predictors of their antifungal activity in our other study that includes application of artificial neural networks as chemometric method (not published).

2. 3. Multivariate Statistical Methods: PCA, PCR and PLS

The main objective of PCA is to substitute the representation of the objects, from the initial representation in the form of the *n* original intercorrelated variables, into the new principal component coordinate space.[20] Therefore, PCA can be defined as a statistical technique for reducing the amount of data when there is a correlation present. It is worth stressing that it is not a useful technique if the variables are uncorrelated.[21] In PCA, objects or analytes are represented in a multidimensional space, where the variables define the axes, and are projected into a few principal components (PCs) which are linear combination of the original variables and describe the maximum variation within the data. Each PC is characterized by scores, which actually are the new coordinates of the projected objects, and by loadings which reflect the direction with respect to the original variables.[22] In addition, PCA is a very useful tool in providing data overview and determination of the outliers among the analytes (data lying outside the Hotelling T^2 ellipse).

The aim of PCR is to reduce the number of predictor variables by using first few PCs rather than the original variables. This statistical method works well when there is a considerable degree of correlation between the predictor variables.[21] That can cause mathematical problems with MLR, resulting in unreliable predictions. PCR usually implies three main steps: (1) running the PCA on the table of the explanatory variables, (2) running an ordinary least squares regression (linear regression) on the selected components: the factors that are most correlated with the dependent variable will be chosen, and (3) computation of the parameters of the model for the selected explanatory variables.[20]

Like PCR, PLS regression uses linear combinations of the predictor variables rather than the original variables. In PCR the PCs are chosen so that they describe as much of the variation in the predictor variables as possible, irrespective of the strength of the relationships between the predictor and the response variables. However, in PLS variables that show a high correlation with the response variables are given extra weight because they will be more effective at prediction.[21] As a consequence, PLS finds components that both show high variation and are highly correlated with dependent variable. Therefore, PCR and PLS enable analysis of strongly collinear data, reducing the high-dimensional data matrix to a much smaller and interpretable set of latent variables (LVs).

The optimal complexity and predictivity of the models are usually determined by *cross*-validation. In the case of both PCR and PLS regressions, statistical and predictive quality of the models are evaluated by cumulative sum of squares of the *Y*s explained by all extracted components (R^2Y_{cum}), root mean square error of calibration (*RMSEC*), root mean square error of prediction (*RMSEP*), root mean square error of *cross*-validation (*RMSECV*), determination coefficient of calibration (R^2_{cal}), determination coefficient of prediction (R^2_{pred}) and determination coefficient of *cross*-validation (R^2_{cv}).

The complete PCA, PCR and PLS calculation procedures were conducted by using a demo version of PLS Toolbox statistical package for MATLAB version 7.12.0.635 R2011a.[23] The data were separated in two subsets: calibration set with 18 molecules, and external test set with 6 randomly selected molecules (Table 1).

3. Results and Discussion

3. 1. PCA

PCA was carried out in order to determine the presence of outliers among the analytes with 0.95 confidence level for T^2 Hotelling limit for outliers and to overview the examined compounds for similarities and dissimilarities. PCA resulted in a three-component model explaining 92.56% of the data variation. The cumulative variance explained by the first two PCs is 87.39% (PC1 comprises 60.45% and PC2 comprises 26.94% of the total data variability). The addition of more PCs did not significantly change the distribution of the molecules on the score plot. Figure 1 shows score values and the mutual projections of the loading vectors for the first two PCs. It is obvious that all the compounds are lying inside the Hotelling T^2 ellipse, suggesting that there are no outliers among the analytes. Unfortunately, score plot did not reveal any classification of the compounds, except significant distance of compound 1, 6 and 16 from the other compounds going along the PC2 direction. It could be assumed that this separation of mentioned compounds is caused by their structural characteristics: compound 1 is unsubstituted, and compounds 6 and 16 are the only acetamides in the series with significantly different values of certain mole-

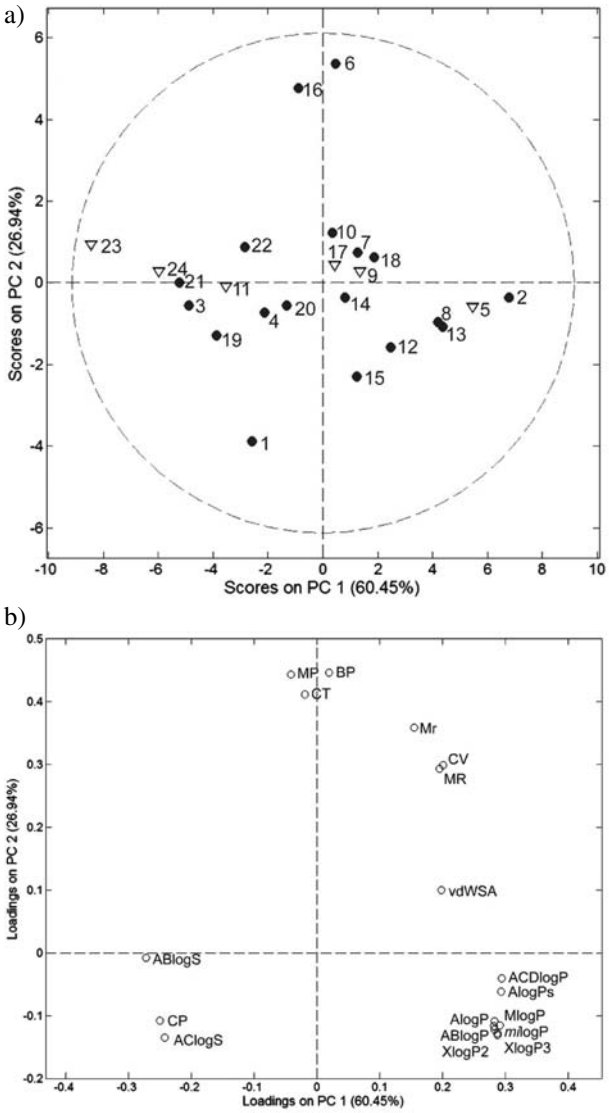


Figure 1. Score plot (a) and factor loadings (b) of molecular descriptors for the first two PCs.

cular descriptors. Loading plot shows that partition coefficients have the highest positive influence going along the PC1 axis, while water solubility descriptors and CP express the highest negative influence. The highest positive influence on the PC2 score values have MP, BP and CT.

3. 2. PCR

Simple methods for selection of a good set of PCA scores for PCR are (1) selection of the first PCA scores that cover a certain percentage of the total variance of *X* (i.e. 95%) and (2) selection of the PCA scores with maximum correlation to *Y*.[24] In this paper, we applied first approach and selected the first six PCA scores which cover 98.62% of the total variance of *X*. The obtained PCR model shows $R^2Y_{cum} = 92.99\%$, *RMSEC* = 0.03176, *RMSECV* = 0.05661, $R^2_{cal} = 0.9299$, $R^2_{cv} = 0.8105$. *Cross*-validation was preformed by splitting the entire calibration set into three random subsets and one iteration. After building the calibration model and application of *cross*-validation, the obtained model was applied on external data set and calculated *RMSEP* amounts to 0.04143 followed by R^2_{pred} of 0.9180. Figure 2a shows the plot of the experimentally measured $\log(1/c_{MIC})$ *versus* predicted $\log(1/c_{MIC})$ for PCR model.

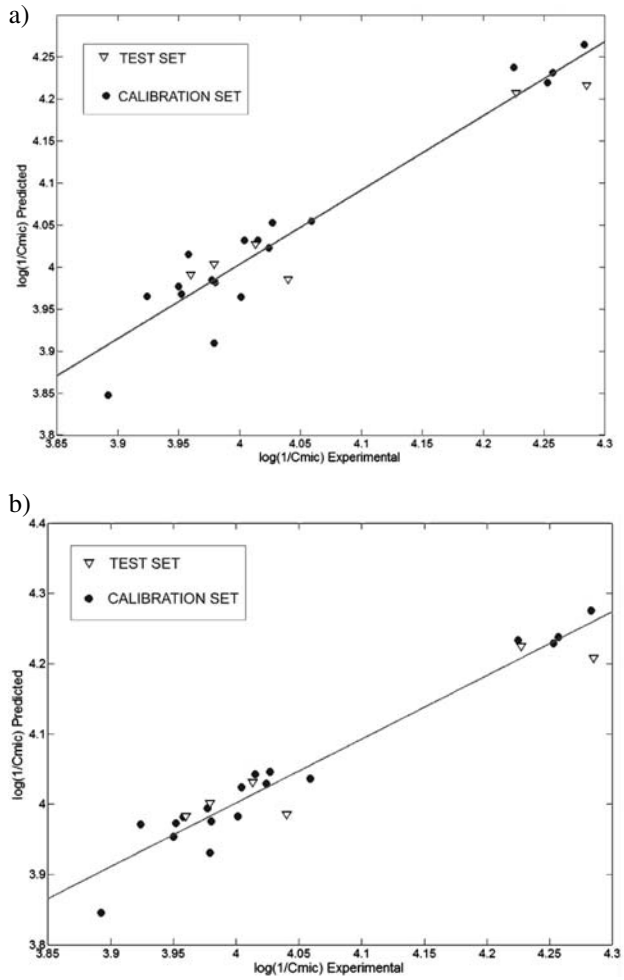


Figure 2. The plot of the experimentally measured versus predicted $\log(1/c_{MIC})$ values obtained from (a) the PCR and (b) PLS model.

Low scattering of the points around the linear relationship curve, low values of errors and high determination coefficients indicate a very good concurrence between experimental and predicted $\log(1/c_{MIC})$ values, as well as a very good predictive power of the model. This statement is also confirmed by the residual *versus* predicted $\log(1/c_{MIC})$ values plot, which is presented in Figure 3a. This plot is very informative regarding model fitting to a data set. If the residual values are randomly distributed (low correlation), then it implies that the model fits the data well.[25] Therefore, Figure 3a indicates that PCR model fits adequately to all the data.

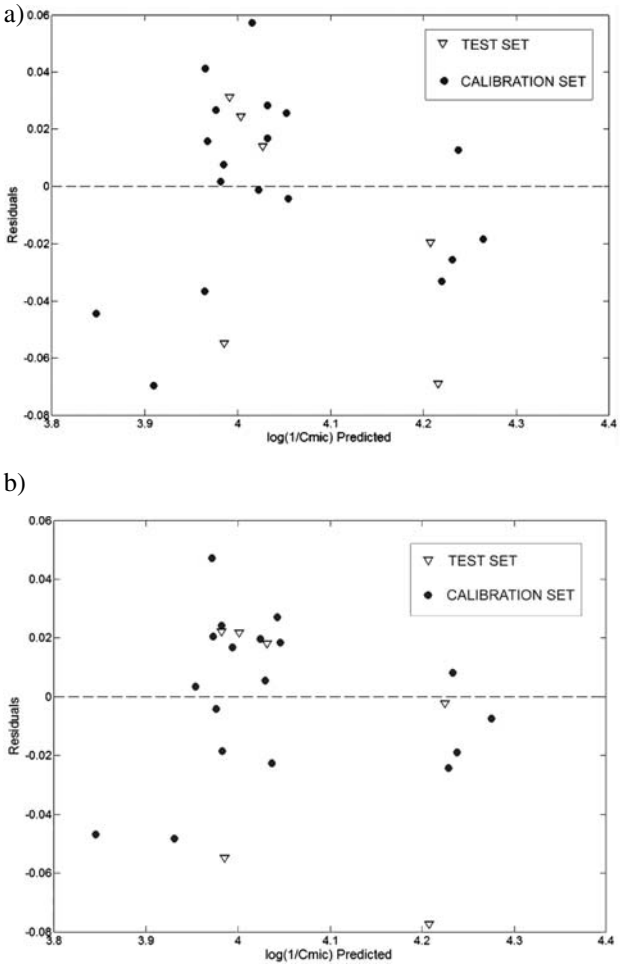


Figure 3. The residual versus predicted log(1/c_{MIC}) values plot for (a) the PCR and (b) PLS model.

3. 3. PLS

The PLS model calibration was carried out on the entire calibration data set. The obtained model was evaluated using random subsets (three splits and one iteration) *cross*-validation method. The number of the LVs was chosen on the basis of the minimum *RMSECV* value, which was obtained for five LVs model. Five LVs capture 92.89% of the variance in the descriptor variables space, indicating that the information contained in the descriptors are effectively used in the calibration model. The obtained PLS model is also characterized by R^2_{cal} = 0.9557 and R^2_{cv} = 0.8606. The explained calibration variance R^2Y_{cum} for the dependent variable log(1/c_{MIC}) was 94.16% and *RMSEC* was 0.02526 units of log(1/c_{MIC}). The error of validation, expressed as *RMSECV*, was 0.04533 units of log(1/c_{MIC}). After calibration, established model was tested by external test set. The obtained *RMSEP* was 0.04068 and R^2_{pred} was 0.9328. Comparison of the experimentally obtained log(1/c_{MIC}) values and predicted log(1/c_{MIC}) values is showed in Figure 2b. It shows low scattering of the points around the linear relationship curve indicating very good concurrence between predicted and measured data. Residual *versus* predicted log(1/c_{MIC}) values plot for PLS model is presented in Figure 3b. Randomly distributed residual values imply excellent fitting of the data.

Depicted results indicate a very good predictive power of the established PLS model, which is better than PCR model. The reason why the PLS model has better statistical measures and predictive power than PCR model lies in the fact that PLS technique uses both criteria: the direction of the highest variance in the data set and the best correlation with the dependent variable scores in its search for the latent variables in the *X*-variable and *Y*-variable domain. However, both methods PCR and PLS are very important in chemometrics because of ability to model a large number of highly intercorrelated variables. The appearance of outliers in PCR and PLS models could be explained with significant differences in molecular structure as well as with prediction error of the model.

3. 4. Contribution of the Molecular Descriptors to the MIC of Studied Compounds

The assessment of the contribution of variables (molecular descriptors) on *Y* (log(1/c_{MIC})) in both the PCR and PLS models, was done based on variable importance in the projection (VIP) scores. The descriptors with a VIP score higher than 1 were considered as relevant for explaining *Y*, and those significantly lower than 1 (arbitrarily, the value lower than 0.5 was taken) had little or almost no influence. The descriptors characterized by VIP > 1.5 were considered as the most significant. The variables *versus* VIP scores for *Y* plot is presented in Figure 4.

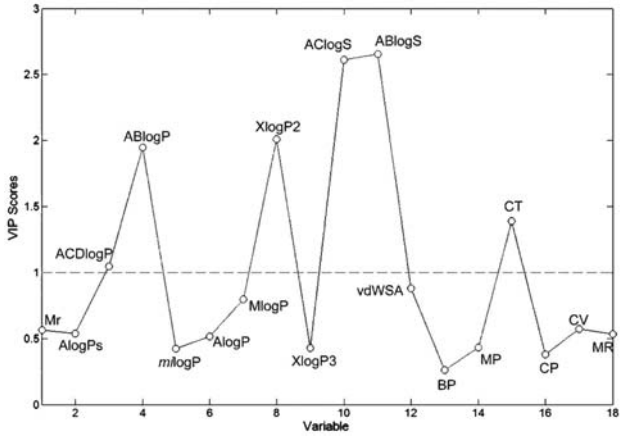


Figure 4. Plot of the variables versus VIP scores for log(1/c_{MIC})

The plots of the regression coefficients of descriptors in PCR and PLS models are showed in Figure 5, where the descriptors with VIP > 1.5 are denoted by asterisks.

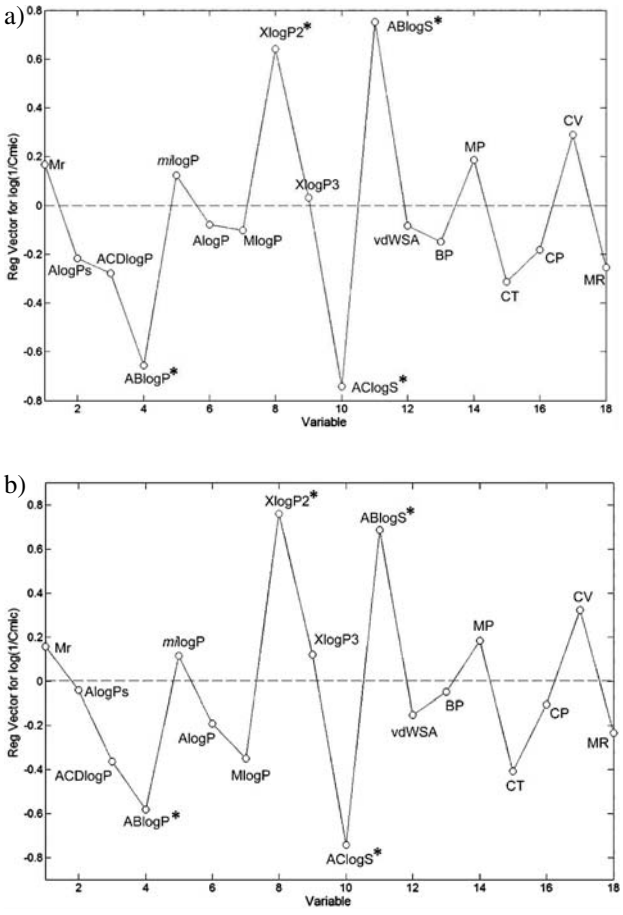


Figure 5. Plot of the coefficients of molecular descriptors in (a) the PCR and (b) PLS model.

Hence, the most significant descriptors influencing the $\log(1/c_{MIC})$ values of studied benzoxazoles and oxazolo[4,5-*b*]pyridines are: solubility in water (AClog*S*, ABlog*S*) and partition coefficients (ABlog*P*, Xlog*P*2). It is evident that the same descriptors are included in both PCR and PLS models. These descriptors are characterized by the highest regression coefficients and VIP values.

Solubility in water and lipophilicity of molecules are very important factors of biological activity,[26] in this case antifungal activity. Therefore, on the basis of mentioned descriptors, antifungal activity of studied compounds could be successfully predicted by using appropriate mathematical model.

4. Conclusions

The present paper focuses on identifying the most significant molecular descriptors that have the highest influence on antifungal activity of some benzoxazole and oxazolo[4,5-*b*]pyridine derivatives against *Candida albicans*. For this purpose, PCA followed by PCR and PLS regression chemometric methodologies was applied. Unfortunately, PCA did not reveal any significant classification of studied compounds according to calculated molecular descriptors. However, PCR and PLS methods showed that antifungal activity of benzoxazoles and oxazolo[4,5-*b*]pyridines is mostly influenced by their lipophilicity, expressed as ABlog*P* and Xlog*P*2, as well as solubility in water, expressed as AClog*S* and ABlog*S*. In both PCR and PLS models, mentioned descriptors are the most important, based on VIP factors and regression coefficients. Taking into account the statistical parameters that represent the quality and predictivity of the model, it can be concluded that PLS regression resulted as better than PCR. The present study allows us to estimate antifungal activity for similar compounds and to understand their biological behaviour.

5. Acknowledgement

These results are the part of the projects No. 172012, No.172014 and TR 31055 supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and the project No. 114-451-2373/2011, financially supported by the Provincial Secretariat for Science and Technological Development of Vojvodina.

6. References

1. O. Ursu, A. Costescu, M. V. Diudea, B. Parv, *Carpatian J. Math.* **2004**, *20*, 267–274.
2. O. Temiz-Arpaci, *Turk. J. Med. Sci.* **2001**, *31*, 493–497.
3. Z. A. Kaplancikli, G. Turan-Zitouni, G. Revial, K. Guven, *Arch. Pharm. Res.* **2004**, *27*, 1081–1085.
4. O. Temiz, I. Oren, E. Sener, I. Yalçin, N. Uçartürk, *Farmaco*. **1998**, *53*, 337–341.
5. M. Arisoy, O. Temiz-Arpaci, I. Yildiz, F. Kaynak-Onurdag, E. Aki, I. Yalçin, U. Abbasoglu, *SAR QSAR Environ Res.* **2008**, *19*, 589–612.
6. T. Ertan, I. Yildiz, B. Tekiner-Gulbas, K. Bolelli, O. Temiz-Arpaci, S. Ozkan, F. Kaynak, I. Yalçin, E. Aki, *Eur. J. Med. Chem.* **2009**, *44*, 501–510.
7. L. Ouyang, Y. Huang, Y. Zhao, G. He, Y. Xie, J. Liu, J. He, B. Liu, Y. Wei, *Bioorg Med Chem Lett.* **2012**, *22*, 3044–3049.
8. E. I. Khizhan, A. I. Khizhan, A. N. Nikolaevskii, O. P. Kniga, T. N. Ivleva, G. A. Tikhonova, *Russ J Gen Chem*, **2011**, *81*, 122–127.
9. I. Yildiz-Oren, B. Tekiner-Gulbas, O. Temiz-Arpaci, I. Yalçin, E. Aki-Sener, *Asian J. Chem.* **2004**, *16*, 1359–1366.
10. R. Gudipati, R. N. Anreddy, S. Manda, *J Enzyme Inhib Med Chem.* **2011**, *26*, 813–818.
11. S. O. Podunavac-Kuzmanović, L. R. Jevrić, S. Z. Kovačević, N. D. Kalajdžija, *APTEFF* **2012**, *43*, 273–282.
12. S. O. Podunavac-Kuzmanović, D. D. Cvetković, L. R. Jevrić, N. U. Uzelac, *Hem. Ind.* **2013**, *67*, 27–33.

13. S. W. Dietrich, in: M. E. Wolff (ed.): Burger's Medicinal Chemistry and Drug Discovery: Principles and Practice, Vol. 1, 5th ed. John Wiley, New York, USA, **1995**, pp. 415–496.
14. C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington DC, USA, **1995**.
15. S. O. Podunavac-Kuzmanović, D. D. Cvetković, L. R. Jevrić, N. J. Uzelac, *Acta Chim. Slov.* **2013**, *60*, 26–33.
16. O. Ursu, A. Costescu, M. V. Diudea, B. Parv, *Croat. Chem. Acta* **2006**, *79*, 483–488.
17. ChemBioOffice 2010, PerkinElmer Informatics, http://www.cambridgesoft.com/
18. MOPAC2012, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, http://OpenMOPAC.net
19. VCCLAB, Virtual Computational Chemistry Laboratory, http://www.vcclab.org/
20. K. H. Esbensen, Multivariate Data Analysis – In Practice: An Introduction to Multivariate Data Analysis and Experimental Design, 5th Edition, CAMO Software AS, USA, **2009**.
21. J. N. Miller, J. C. Miller, Statistics and Chemometrics for Analytical Chemistry, 6th Edition, Pearson Education Limited, Harlow, UK, **2010**.
22. R. G. Brereton, Chemometrics, Data Analysis for the Laboratory and Chemical Plant, Wiley, Chichester, England, **2003**.
23. The MathWorks Inc, Natick, MA, USA, http://www.mathworks.com
24. K. Varmuza, P. Filzmaster, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, Taylor & Francis Group, Boca Raton, **2008**.
25. J. Trifković, F. Andrić, P. Ristivojević, D. Andrić, Ž. Lj. Tešić, D. M. Milojković-Opsenica, *J. Sep. Sci.* **2010**, *33*, 2619–2628.
26. R. Mannhold, Molecular Drug Properties, Measurement and Prediction, Vol. 37, Wiley-VCH, Darmstadt, Germany, **2008**.

Povzetek

QSAR metoda je bila uporabljena v seriji benzoksazolskih in oksazolo[4,5-*b*]piridinskih derivatov različnih struktur s ciljem ugotavljanja njihovih inhibitorskih aktivnosti na glivo *Candida albicans*. Analiza glavnih komponent (*Principal Component Analysis* – PCA), regresija glavnih komponent (*Principal Component Regression* – PCR) in metod delnih najmanjših kvadratov (*Partial Least Squares* – PLS) so bili uporabljeni za identifikacijo najustreznejše deskriptorjev in za modeliranje odnosa med različnimi fizikalno-kemičnimi molekularnimi deskriptorji in antifungalne aktivnosti preizkušenih derivatov. Ugotovljeno je da imajo na inhibitorsko aktivnost derivatov benzoksazola in oksazolo[4,5-*b*]piridina na *C. albicans* največji vpliv deskriptorji lipofilnosti (ABlog*P* in Xlog*P*2) in topnost v vodi (AClog*S* in ABlog*S*). PLS regresija je pokazala statistično najboljšo učinkovitost (*RMSEC* = 0.02526 in *RMSECV* = 0.04533) od PCR (*RMSEC* = 0.03176 in *RMSECV* = 0.05661).

SUPPORTING INFORMATION

Multivariate Regression Modelling of Antifungal Activity of Some Benzoxazole and Oxazolo[4,5-*b*]pyridine Derivatives

Strahinja Z. Kovačević,* Sanja O. Podunavac Kuzmanović and Lidija R. Jevrić

University of Novi Sad, Faculty of Technology, Department of Applied and Engineering Chemistry, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia

** Corresponding author: E-mail: strahinjakovacevic@hotmail.com; phone: +381642839686; fax: +38121450413*

Received: 28-03-2013

Table S1. *In vitro* antifungal activity ($\log(1/c_{MIC})$) of studied compounds against *Candida albicans* (MTCC 183)[16]

Comp.	IUPAC name	$\log(1/c_{MIC})$
1	2-phenyl-1,3-benzoxazole	3.892
2	2-(4-tert-butylphenyl)-1,3-benzoxazole	4.001
3	4-(1,3-benzoxazol-2-yl)aniline	3.924
4	4-(1,3-benzoxazol-2-yl)-N-methylaniline	3.952
5*	5-chloro-2-(4-ethylphenyl)-1,3-benzoxazole	4.013
6	N-[4-(5-chloro-1,3-benzoxazol-2-yl)phenyl]acetamide	4.059
7	4-(5-chloro-1,3-benzoxazol-2-yl)-N-methylaniline	4.015
8	5-chloro-2-(4-chlorophenyl)-1,3-benzoxazole	4.024
9*	5-chloro-2-(4-nitrophenyl)-1,3-benzoxazole	4.040
10	2-(4-ethylphenyl)-1,3-benzoxazol-5-amine	3.979
11*	2-(4-fluorophenyl)-1,3-benzoxazol-5-amine	3.960
12	5-methyl-2-(4-methylphenyl)-1,3-benzoxazole	4.005
13	2-(4-ethylphenyl)-5-methyl-1,3-benzoxazole	3.950
14	2-(4-methoxyphenyl)-5-methyl-1,3-benzoxazole	3.977
15	2-(4-fluorophenyl)-5-methyl-1,3-benzoxazole	3.980
16	N-[4-(5-methyl-1,3-benzoxazol-2-yl)phenyl]acetamide	3.958
17*	N-methyl-4-(5-methyl-1,3-benzoxazol-2-yl)aniline	4.027
18	N,N-dimethyl-4-(5-methyl-1,3-benzoxazol-2-yl)aniline	3.979
19	2-(4-methylphenyl)-[1,3]oxazolo[4,5-b]pyridine	4.225
20	2-(4-ethylphenyl)-[1,3]oxazolo[4,5-b]pyridine	4.253
21	2-(4-methoxyphenyl)-[1,3]oxazolo[4,5-b]pyridine	4.257
22	2-(4-ethoxyphenyl)-[1,3]oxazolo[4,5-b]pyridine	4.283
23*	4-{[1,3]oxazolo[4,5-b]pyridin-2-yl}aniline	4.227
24*	2-(4-nitrophenyl)-[1,3]oxazolo[4,5-b]pyridine	4.285

Table S2. The values of the *in silico* molecular descriptors for studied benzoxazole and oxazolo[4,5-*b*]pyridine derivatives

Compound No.	Mr	AlogPs	ACDlogP	ABlogP	milogP	AlogP	MlogP	XlogP2	XlogP3	AClogS	ABlogS	vdWSA	BP	MP	CT	CP	CV	MR
1	195.230	3.860	3.630	3.960	3.650	3.280	3.250	3.160	3.430	–4.610	–3.950	271.088	651.580	434.700	861.810	37.040	561.500	58.040
2	251.350	5.560	5.140	5.590	5.350	4.680	4.260	4.980	5.100	–5.770	–5.230	398.610	744.850	494.720	870.130	24.430	774.500	77.720
3	210.250	3.310	2.910	3.160	2.720	2.530	2.680	2.340	2.740	–4.680	–3.140	286.066	743.700	530.480	898.050	39.010	590.500	62.870
4	224.280	3.680	3.200	4.050	3.510	3.090	2.950	2.810	3.410	–4.670	–3.230	322.879	729.610	511.150	871.650	32.990	652.500	67.230
5*	257.730	5.390	4.910	5.280	5.220	4.890	4.290	4.680	4.850	–5.840	–4.830	350.472	744.730	512.200	879.310	27.640	722.500	73.140
6	286.730	3.940	3.850	3.480	3.520	3.040	3.240	3.040	3.230	–5.680	–4.420	363.812	869.640	678.410	1006.720	31.140	750.500	75.810
7	258.720	4.280	3.810	4.470	4.160	3.750	3.470	3.430	4.040	–5.400	–4.120	339.165	772.020	553.590	886.040	31.110	701.500	71.840
8	264.120	5.200	4.860	4.910	4.980	4.610	4.310	4.400	4.680	–6.080	–5.250	304.035	736.400	519.580	888.830	32.770	659.500	67.250
9*	274.670	4.410	4.250	4.120	4.260	3.840	3.720	3.670	3.880	–5.630	–5.260	327.916	690.678	621.550	890.371	39.407	656.500	69.466
10	238.310	4.250	3.580	4.570	3.620	3.480	3.200	3.240	3.540	–5.180	–4.170	348.719	794.440	565.540	902.480	30.590	702.500	73.360
11*	228.240	3.330	2.970	2.710	2.860	2.740	3.080	2.500	2.840	–5.000	–4.430	292.867	747.950	543.590	882.050	36.250	608.500	63.270
12	223.290	4.530	4.260	4.780	4.520	4.250	3.770	4.030	4.150	–5.290	–4.160	390.131	707.300	482.280	864.230	28.750	673.500	69.840
13	237.320	5.140	4.620	5.270	4.990	4.710	4.020	4.500	4.590	–5.450	–4.510	335.674	730.180	493.550	867.390	26.000	729.500	74.440
14	239.290	4.070	3.840	4.260	4.130	3.750	3.200	3.510	3.760	–4.970	–4.050	366.123	729.720	504.510	869.350	28.230	691.500	71.190
15	227.250	4.250	4.000	4.410	4.240	3.970	3.910	3.760	3.890	–5.260	–4.380	351.537	683.690	471.600	846.660	30.390	653.500	62.820
16	266.320	3.700	3.550	3.470	3.290	2.890	2.970	2.850	2.970	–5.290	–3.860	310.080	855.090	659.760	994.870	29.190	757.500	77.100
17*	238.310	3.990	3.520	4.460	3.930	3.580	3.200	3.240	3.770	–5.010	–3.560	379.273	757.470	534.940	874.280	29.160	708.500	73.130
18	252.340	4.120	3.940	4.590	4.180	3.930	3.450	3.810	3.910	–4.990	–3.740	354.709	742.620	526.020	874.950	25.950	747.500	79.120
19	210.250	3.110	2.880	3.150	3.200	2.810	2.830	3.070	3.050	–5.000	–2.810	295.458	710.260	518.760	870.140	36.420	610.500	61.230
20	224.280	3.880	3.230	3.640	3.660	3.270	3.100	3.540	3.490	–5.160	–3.170	326.415	733.140	530.030	872.370	32.540	666.500	65.832
21	226.250	2.870	2.460	2.630	2.810	2.310	2.300	2.550	2.660	–4.670	–2.610	311.576	732.680	540.990	874.470	35.690	628.500	62.653
22	240.280	3.430	2.890	3.110	3.180	2.660	2.560	2.970	3.030	–4.970	–2.780	342.948	755.560	552.260	877.360	31.920	684.500	67.401
23*	211.240	2.250	1.840	1.940	1.830	1.580	2.020	1.820	2.010	–4.730	–2.410	278.255	774.520	590.750	905.320	44.210	583.500	60.890
24*	241.220	2.310	2.570	2.480	2.710	2.220	2.560	2.530	2.520	–4.940	–3.240	303.447	677.421	639.380	881.089	47.959	600.500	63.514

* external test set