Journal of Enzyme Inhibition and Medicinal Chemistry, 2010; 25(5): 696–701

informa
healthcare

RESEARCH ARTICLE

QSAR model for predicting the fungicidal action of 1,2,4-triazole derivatives against *Candida albicans*

Rajeev K Singla[1], and Varadaraj Bhat G[2]

[1]*Department of Pharmaceutical Chemistry, Shri Gopichand College of Pharmacy, Ahera-Baghpat, 250609, U.P, India, and* [2]*Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal University, Manipal-576104, Karnataka, India*

Abstract

QSAR analysis of a series of previously synthesised 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols(TDFPP) as analogues of fluconazole were tested for growth inhibitory activity against *Candida albicans* using computer assisted multiple regression analysis. This was in order to explore the selectivity requirements for fungicidal activity against *C. albicans* among these congeners. A training set comprising 40 analogues and a test set comprising ten analogues of 1-(1H-1, 2, 4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols were selected for the present investigation by using the sphere exclusion method embedded in the Vlife MDS 3.5 software. With respect to the modelling of the growth inhibitory activity of the reported compounds, the regression analysis shows that even in the mono-parametric correlations the topological and physicochemical parameters give significant regression coefficients. The validation of the QSAR models was performed by cross-validation and external test set prediction. The model is not only able to predict the activity of new compounds but also explains the important region in the molecules in a quantitative manner.

Keywords: *2D-QSAR; 1; 2; 4-triazole; MDS; multiple regression analysis;* Candida albicans

Introduction

In recent years, it has been observed that life threatening systemic fungal infections have become increasingly common, especially in the immunocompromised hosts suffering from tuberculosis, cancer or AIDS and in organ transplant cases [1,2]. Candidiasis caused by a group of microscopic fungi or yeast, the most common being *Candida albicans*. Current research can explain the frequency with which unexplained oral candidiasis has led to unequivocal acquired immunodeficiency syndrome (AIDS) in patients at risk. Twenty-two previously healthy adults with unexplained oral candidiasis (19 of those tested had a reversed T4/T8 ratio and 20 had generalised lymphadenopathy) were compared with 20 similar patients with a reversed T4/T8 ratio and generalised lymphadenopathy that did not have oral candidiasis [3]. Moreover Candida is the leading cause of oesophagitis (throat inflammation) in people with AIDS. Such a causative agent requires an equally effective focus of research.

Quantitative structure–activity relationship (QSAR) studies are undoubtedly of great importance in modern chemistry and biochemistry. Ideally, the activities and properties are connected by some known mathematical function, F where biological activity = F [structure (in the present study topological and physicochemical descriptors are used as the structural parameters)]. A 2D-QSAR study makes it simple to interpret the biological data in terms of different descriptors obtained from the two dimensional structures of the compounds without involving energy minimization procedures. In a congeneric series, where a relative study is being carried out, the 2D-descriptors may play an important role in deriving significant correlations with the biological activities of the compounds. Thus the novelty and importance of a 2D-QSAR study is mainly due to its simplicity of the calculations of different descriptors and their interpretation (in a physical sense) to explain the inhibition actions of compounds in a congeneric series.

1, 2, 4-Triazole and its derivatives represent one of the most biologically active classes of compounds, possessing a wide spectrum of activities. The 1, 2, 4-triazole derivatives have been reported to have diverse pharmacological activities such as antibacterial, antifungal, hypoglycaemic, antihypertensive and analgesic properties [4–6]. Triazoles

Address for Correspondence: Rajeev K Singla, E-mail id: rajeevsingla26@gmail.com, Mobile No. +91-9255000962.

(Received 10 August 2009; revised 01 December 2009; accepted 02 December 2009)

ISSN 1475-6366 print/ISSN 1475-6374 online © 2010 Informa UK, Ltd.
DOI: 10.3109/14756360903524296

have been used widely and efficiently as antifungal agents. These antifungal drugs act by inhibiting CYP51, a necessary enzyme in the biosynthesis of ergosterol, through a mechanism in which the heterocyclic nitrogen atom (N-4 of triazole) binds to the haem iron atom. However, the increasing administration of antifungal agents has led to the development of fungal resistance. The design of new TDFPP derivatives requires a more detailed knowledge of the mechanism of CYP51 inhibition by this class of compounds. Hence in the present study, a 2D-quantitaive SAR (2D-QSAR) study on these analogues (Tables 1, 2, 3, 4 and 5) along with their generalized structures (Figures 1 and 2) has been conducted to provide the rationale for drug design using Vlife Molecular Design suite 3.5 (Baghpat, U.P). This software offers several hundred descriptors from different perspectives relating to the physico-chemical, topological and geographical characteristics to the molecules under a multi-descriptor class environment.

Material and methods

Data set

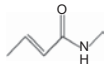
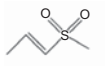
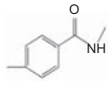
The compound derived from1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanol moiety (Tables 1, 2, 3, 4 and 5 together with Figures 1 and 2) was selected from the literature[7,8] along with their *in vitro* antifungal activity against *Candida albicans*. The activity expressed as minimum inhibitory concentration (MIC_{80}), was defined as the first well with an approximate 80% reduction in growth compared to the growth of the drug-free well and is expressed as $logMIC_{80}$ on a micromolar basis to improve the normal distribution of the experimental data points. Two dimensional quantitative structure–activity relationship studies of these 1, 2, 4-triazole derivatives were carried out by using the Vlife Molecular Design Suite 3.5. (Vlife Sciences Technologies, India) [9–11]. The structures of the compounds under study have been drawn in the 2D drawing application (2D Draw app) of MDS 3.5 using the standard procedure. These structures were converted into 3D objects by using the default conversion procedure implemented in the MDS 3.5 software. The generated 3D structures of the compounds were subjected to batch energy minimization in the force field module using Merck Molecular Force Field (MMFF) and MMFF charge followed by considering distance dependent dielectric constant of one and convergence criteria of 0.01 kcal/mol. This ensures a well defined conformer relationship across the

Table 1. Observed and modelled fungicidal action of 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols against *Candida albican* (Refer to Figure 1).

Data code	R	Obsd. $logMIC_{80}$ (μM)*	Calcd. $logMIC_{80}$ (μM) Eq(2)	Residual value
TCA1a	H	−2	−1.47112	−0.52888
TCA1b	2-F	−2.013	−1.76353	−0.24947
TCA1c	3-F	−1.411	−1.36533	−0.04567
TCA1d	4-F	−1.411	−1.5112	0.1002
TCA1e	2-Cl	−1.429	−1.38854	−0.04046
TCA1f	3-Cl	−1.429	−1.27952	−0.14948
TCA1g	4-Cl	−0.224	0.0148789	−0.23888
TCA1h	2-Br	−0.872	−1.3442	0.4722
TCA1i	4-Br	−0.872	−1.23382	0.36182
TCA1j	2-CH_3	−1.406	−1.45568	0.04968
TCA1k	4-CH_3	−1.406	−1.34872	−0.05728
TCA1l	4-NO_2	−1.44	−1.23597	−0.20403
TCA1m	4-CH_2CH_3	−1.218	−1.21684	−0.00116
TCA1n	2-Cl, 4-Cl	0.334	0.105121	0.228879
TCA2a	CH_3	−0.8506	−1.05045	0.19985
TCA2b	CH_2CH_3	−2.0706	−0.814338	−1.25626
TCA2c	$CH(CH_3)_2$	−0.2746	−0.562774	0.288174
TCA2d	$(CH_2)_3CH_3$	−0.2874	−0.313005	0.025605
TCA2e	$CH_2CH(CH_3)_2$	0.3146	−0.313005	0.627605
TCA2f	$(CH_2)_4CH_3$	0.3022	−0.0482162	0.350416
TCA3a	H	0.2971	−0.603219	0.900319
TCA3b	3-NO_2	−0.342	−0.359714	0.017714
TCA3c	4-NO_2	−0.342	−0.359714	0.017714
TCA3d	4-Cl	0.2684	−0.404133	0.672533
TCA3e	2-$COOCH_3$	−0.9546	−0.288324	−0.66628
TCA3f	4-$COO(CH_2)_3CH_3$	0.2184	0.550738	−0.33234

*Taken from Takuya [8] Bastikar [9]

Table 2. Observed and modelled fungicidal action of 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols against *Candida albicans* (Refer to Figure 2).

Data code	X	Obsd. $logMIC_{80}$ (μM)*	Calcd. $logMIC_{80}$ (μM) Eq(2)	Residual value
TCA4a		≤−1.8239	−1.68851	−0.13539
TCA4b		≤−1.8416	−1.96728	0.12568
TCA4c		≤−1.8124	−2.15553	0.34313
TCA4d		≤−1.8664	−1.7934	−0.073
TCA4e		≤−1.8108	−1.93186	0.12106

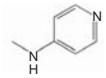
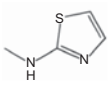
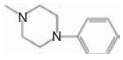
*Taken from Takuya [8] Bastikar [9]

Table 3. Observed and modelled fungicidal action of 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols against *Candida albicans* (Refer to Figure 2).

Data code	R	Obsd. $logMIC_{80}$ (μM)*	Calcd. $logMIC_{80}$ (μM) Eq(2)	Residual value
TCA5a	4-Cl	≤−1.8761	−1.7166	−0.1595
TCA5b	4-Br	≤−1.9066	−1.6758	−0.2308
TCA5c	4-Me	≤−1.8601	−1.77852	−0.08158
TCA5d	4-CN	≤−1.8696	−1.76055	−0.10905
TCA5e	4-Ac	≤−1.8827	−1.60173	−0.28097
TCA5f	4-OH	≤−1.5622	−1.78703	0.22483
TCA5g	4-$OCH_2CF_2CHF_2$	≤−1.6382	−1.6594	0.0212
TCA5h	4-SCF_3	≤−1.9208	−1.83399	−0.08681
TCA5i	3,4-$(CN)_2$	≤−1.886	−1.70818	−0.17782
TCA5j	3-Cl, 4-CN	≤−1.8928	−1.66667	−0.22613
TCA5k	2-F, 4-CN	≤−1.8827	−1.8877	0.005

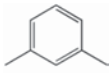
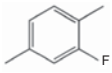
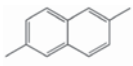
*Taken from Takuya [8] Bastikar [9]

Table 4. Observed and modelled fungicidal action of 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols against *Candida albicans* (Refer to Figure 2).

Data code	NR_1R_2	Obsd. $logMIC_{80}$ (μM)*	Calcd. $logMIC_{80}$ (μM) Eq(2)	Residual value
TCA5l		≤ −1.8538	−1.91385	0.06005
TCA5m		≤ −1.8557	−1.99869	0.14299
TCA5n		−0.9492	0.971213	−1.92041
TCA5o		−1.5575	−1.96159	0.40409
TCA5p		−1.5986	−1.63684	0.03824

*Taken from Takuya [8] Bastikar [9]

Table 5. Observed and modelled fungicidal action of 1-(1H-1,2,4-triazole-1-yl)-2-(2,4-difluorophenyl)-3-substituted-2-propanols against *Candida albicans* (Refer to Figure 2).

Data code	A	Obsd. $logMIC_{80}$ (μM)*	Calcd. $logMIC_{80}$ (μM) Eq(2)	Residual value
TCA6a		≤ −1.8696	−1.76055	−0.10905
TCA6b		≤ −1.8827	−2.03357	0.15087
TCA6c		≤ −1.9066	−1.71195	−0.19465

*Taken from Takuya [8] Bastikar [9]

compounds of the study. All the energy minimized structures of the respective compounds were ported to the 2D-QSAR of QSARPLUS module (MDS 3.5) for computing the parameters corresponding to the physico-chemical [Individual, Chi, Chiv, Path count, Chi Chain, Chiv Chain, Chain path count, Cluster, Path cluster, Kappa, Element count, Polar surface area] and alignment independent topological [Attributes- 2,T, C, N, O, F, S, Cl] descriptor classes. As the total number of the descriptors involved in this study is very large (around 278) after removal of invariable descriptors, only the name of the descriptor classes and the actual descriptor involved in the model have been listed. The multiple linear regression model with stepwise forward-backward variable selection (MLR-SFB) procedure used in developing QSAR models is briefly described below.

Model development

Vlife MDS 3.5 is very robust software. The model used here for the QSAR equation generation was MLR-SFB. MLR-SFB is a "filter" based variable selection procedure. It is a procedure to examine the impact of each variable step by step. An important task in building a QSAR equation is to evaluate the required descriptors for the molecules under

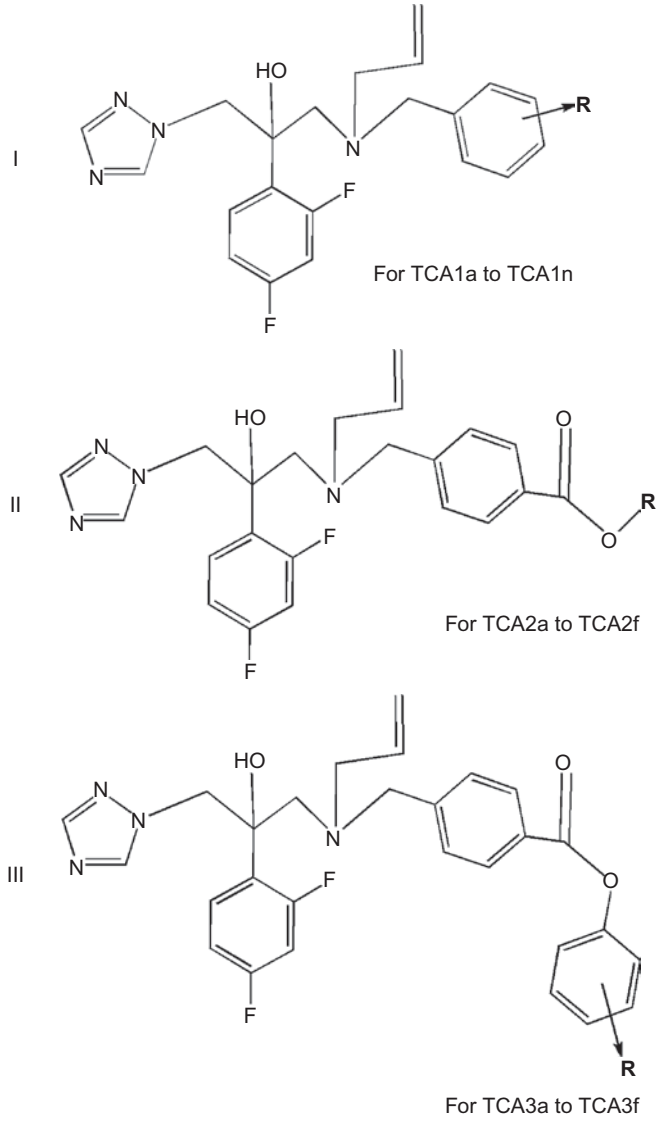


Figure 1. Part I: Basic structure of the analogues.

consideration and the Vlife MDS worksheet was provided for this purpose. $LogMIC_{80}$ was selected as a dependent variable as its value depends on various descriptors whereas the physico-chemical and topological descriptors were selected as independent variables. The sphere exclusion procedure was adopted as the training data set selection method (dissimilarity value filter set by 1.3, +ve) and selected the following: TCA1c, TCA1m, TCA2f, TCA3c, TCA3e, TCA4d, TCA5e, TCA5k, TCA5n, TCA6b as the test set molecules. The dissimilarity value gives the sphere exclusion radius and ensures that points in both the sets are uniformly distributed with respect to chemical and biological space. For the generation of QSAR models using 2D descriptors, the procedure employed was a stepwise forward-backward strategy with multiple linear regression which resulted in the selection of a subset of regressions for the extraction of the diverse structure–activity models. The "filters" set in MLR-SFB are set at 1) cross-correlation limit as 0.9, 2) F test in [4.0] and out [3.99], 3) number of variables in the final equation [2 to

10], 4) r^2 as term selection criteria, 5) 0.0 as variance cut off with autoscaling, 6) number of random iterations as 100, 7) number of groups for cross validation [40- default acceptable value]. All these filters make the variable selection process efficient and lead to a unique solution. After the completion of this process, the fitness plot for training and test set were viewed and the contribution of the individual descriptor for activity was examined. The MLR-SFB protocol has been applied with default filter thresholds to identify all possible models that could emerge from the descriptors of the compounds.

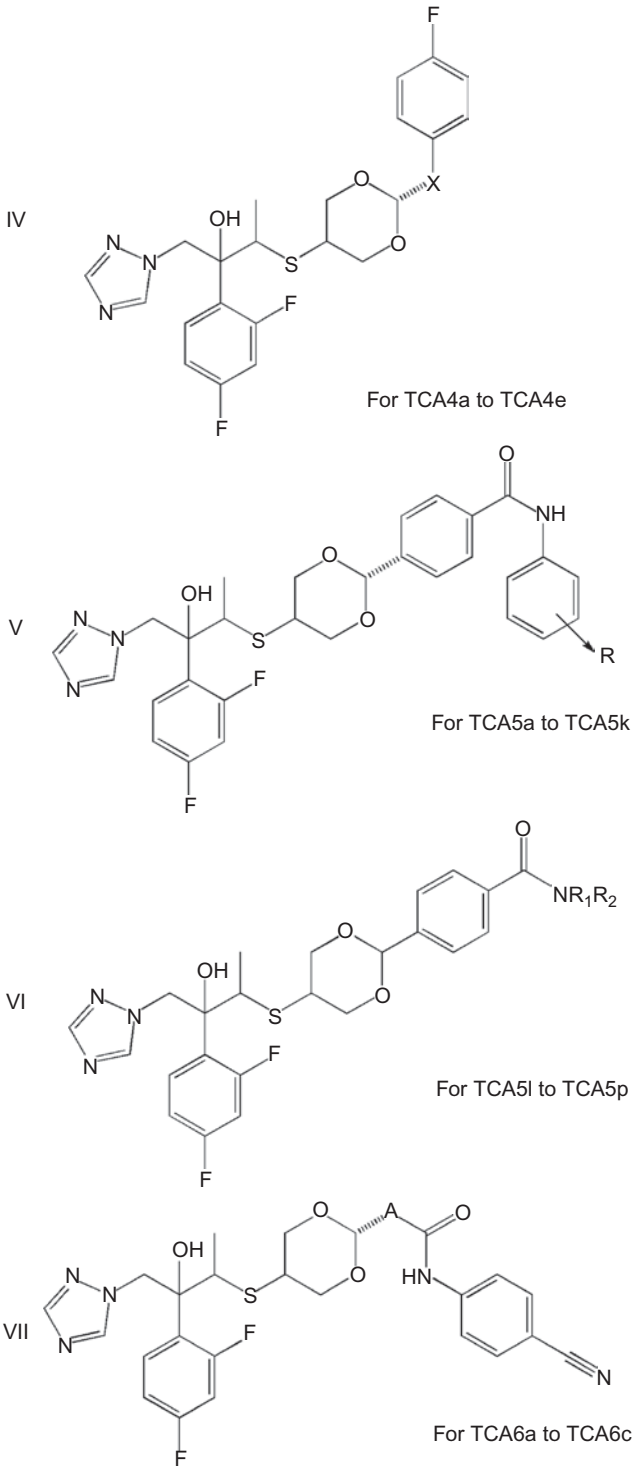


Figure 2. Part II: Basic structure of the analogues.

Results

In the current study, a QSAR model is presented for $LogMIC_{80}$ of fifty 1,2,4-triazole antifungal agents involving theoretical descriptors, which have been calculated from the molecular structure. For the selection of the most important descriptors, the aforementioned stepwise forward-backward multiple

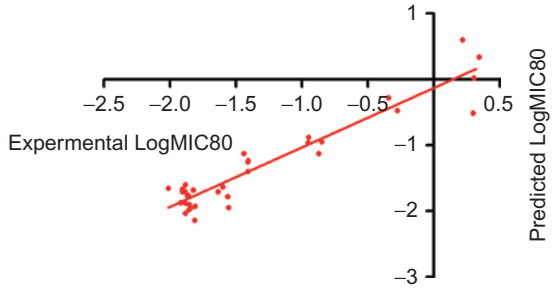


Figure 3. Graphical plot of experimental versus $logMIC_{80}$ according to the model in Equation 2.

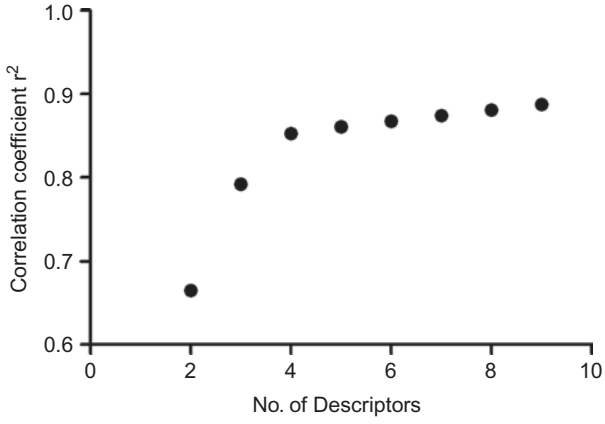


Figure 4. Plot of the correlation coefficient r^2 versus the number of descriptors.

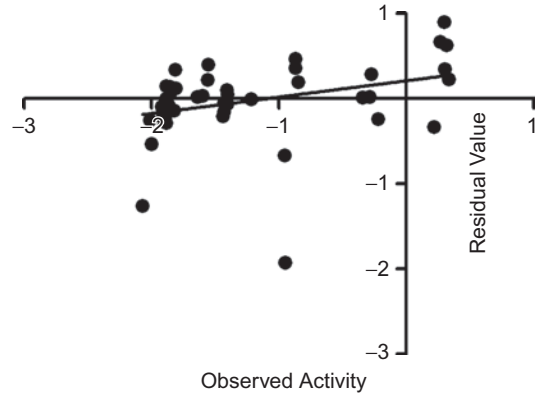


Figure 5. Graphical plot of the observed MIC80 versus residual value according to the model in Equation 2.

linear regression technique was used. A graphical presentation of the relationship between the experimental and predicted $LogMIC_{80}$ values of 1,2,4-triazoles is shown in Figure 3. The predicted property was $LogMIC_{80}$ expressed in μM.

An explanation of the descriptors involved in the multiple linear regressions can be found at the end of the paper. A graphical presentation of the plot between the descriptors and correlation coefficient is shown in Figure 4. As represented by the graph, R^2 is directly proportional to the number of descriptors used. At the same time the N/5 rule should be adhered to. As our data set comprises of only 50 molecules, the number of descriptors should not exceed by 5.

The following regression equations represent the collective class structure activity model of the compounds, together with the statistical parameters of the regression.

$$\begin{aligned} LogMIC_{80}(\mu M) = {} & 1.1282(+0.1018)\,T\_N\_O\_3 \\ & +0.5490(+0.0462)\,K3alpha \\ & +1.2730(+0.2093)\,T\_N\_Cl\_6 \\ & -0.1566(+0.0233)\,T\_C\_F\_5 \\ & -0.3785(+0.1467)\,T\_O\_O\_4 - 6.4056 \end{aligned} \quad (1)$$

$N = 40$, $r^2 = 0.8766$, $q^2 = 0.7876$, F test = 48.3096, pred_$r^2 = 0.6896$

$$\begin{aligned} LogMIC_{80}(\mu M) = {} & 1.0711(+0.1074)\,T\_N\_O\_3 \\ & +0.5117(+0.0513)K3alpha \\ & +1.2944(+0.2248)\,T\_N\_Cl\_6 \\ & -0.1459(+0.0263)\,T\_C\_F\_5 - 6.1136 \end{aligned} \quad (2)$$

$N = 40$, $r^2 = 0.8533$, $q^2 = 0.7989$, F test = 50.8824, pred_$r^2 = 0.6831$

$$\begin{aligned} LogMIC_{80}(\mu M) = {} & 1.2351(\pm 0.1221)\,T\_N\_O\_3 \\ & +0.4281(\pm 0.0532)\,K3alpha \\ & +1.2668(\pm 0.2636)\,T\_N\_Cl\_6 - 6.0886 \end{aligned} \quad (3)$$

$N = 40$, $r^2 = 0.7920$, $q^2 = 0.7282$, F test = 45.7013, pred_$r^2 = 0.6656$

$$\begin{aligned} LogMIC_{80}(\mu M) = {} & 1.3265(\pm 0.1519)\,T\_N\_O\_3 \\ & +0.3650(\pm 0.0787)\,K3alpha - 5.7325 \end{aligned} \quad (4)$$

$N = 40$, $r^2 = 0.6651$, $q^2 = 0.5995$, F test = 36.7408, pred_$r^2 = 0.6482$

As can be observed from these equations, by increasing number of descriptors in the model, the r^2 value increases significantly; however, further additions after four descriptors did not lead to any significant improvement in the r^2 value.

Based on the training set of 40 molecules, the best correlation with the minimum descriptors was between the $logMIC_{80}$ calculated and the $logMIC_{80}$ experimental obtained using Equation 2. From the statistical point of view, this is a robust model. Statistical parameters for this equation were as follows: $N = 40$, $r^2 = 0.8533$, $q^2 = 0.7989$, F test = 50.8824, pred_$r^2 = 0.6831$, alpha rand $r^2 = 0.00000$, alpha rand $q^2 = 0.00000$, alpha rand pred $r^2 = 0.01000$. Cross-validation (q^2) was performed on the training group (40) and showed that 79.89% of the 40 compounds i.e. 32 were correctly classified.

Discussion

The main objective of this study was to build a QSAR model which enables us to identify antifungal compounds from molecular databases using stepwise forward-backward MLR.

From the database of 50 antifungal compounds, we have obtained four stepwise forward-backward MLR equations capable of separating compounds according to their biological activity against *Candida albicans* and an MLR equation which enables a correlation between the MIC_{80} and chemical structure of each drug. Of the compounds 85.33% were correctly classified by r^2 in the training set. The overlapping between the groups was small, which guarantees the quality of the selected discriminant function.

All these results confirm the effectiveness of the topological and physicochemical model proposed here for the search and selection of new potential antifungal drugs against *Candida albicans*. In this work we presented a valid model for the prediction of the antifungal activity of potential molecules against *Candida albicans*.

Conclusion

Our results lead to the conclusion that the fungicidal activity of TDFPP against C*andida albicans* can be successfully modelled with physicochemical and topological descriptors. The separation of the data into two independent sets (training and test) shows that our MLR model can predict external data with great accuracy. The proposed method is a useful aid to the costly and time-consuming experiments for determining fungicidal activity as the method has a high predictive ability. The proposed method can be used to screen existing databases or virtual libraries in order to identify novel potent compounds.

Descriptors

K3alpha: This descriptor signifies the third alpha modified shape index: (n-1) (n-3)2/p32 for odd no. and (n-3)(n-2)2/p32 for even no. where s = n+ a [12].

T_N_O_3: This is the count of number of nitrogen atoms (single, double or triple bonded) separated from any oxygen atom (single, double, triple bonded) by 3 bonds in a molecule[13].

T_N_Cl_6: This is the count of number of nitrogen atoms (single, double or triple bonded) separated from any chlorine atom (single, double, triple bonded) by 6 bonds in a molecule[13].

T_C_F_5: This is the count of number of carbon atoms (single, double or triple bonded) separated from any fluorine atom (single, double, triple bonded) by 5 bonds in a molecule[13].

T_O_O_4: This is the count of number of oxygen atoms (single, double or triple bonded) separated from any other oxygen atom (single, double, triple bonded) by 4 bonds in a molecule[13].

Evaluation of the model

N, number of molecules
R2, coefficient of determination (>0.7)
Q2, cross-validated r2(>0.5)
Pred_r2, predicted r2 for external test(>0.5)
Alpha, statistical significance parameter by randomization test(<0.01)
F-test, F-test for statistical significance of the model (the higher the better, for the same set of descriptors of compounds)

Acknowledgements

The authors would like to thank Vlife Company for providing such user friendly software. The authors would also like to thank Xiaoyun Chai et al, Takuya Uchida and colleagues for providing the study literature.

Declaration of interest

The authors would like to thank Manipal University for funding of the software. The authors have no conflicting financial interests.

References

1. Fridkin SK, Jarvis WR. Epidemiology of nosocomial fungal infections. Clin Microbiol Rev 1996;9:499–511.
2. Wingard JR, Leather H. A new era of antifungal therapy. Biol Blood Marrow Transplant 2004;10:73–90.
3. Klein RS, Harris CA, Small CB, Moll B, Lesser M, Friedland GH. Oral candidiasis in high risk patients as the initial manifestation of the acquired immunodeficiency syndrome. N Engl J Med 1984;311:354–358.
4. Jones DH, Slack R, Squires S, Wooldridge KRH.Antiviral chemotherapy. I. The activity of pyridine and quinoline derivatives against neurovaccinia in mice. J Med Chem 1965;8:676–680.
5. Goswami BN, Kataky JCS, Baruah JN. Synthesis and antibacterial activity of 1-(2,4-dichlorobenzoyl)-4-substituted thiosemicarbazides, 1,2,4-triazoles and their methyl derivatives. J Heterocycl Chem 1984;21:1225–1229.
6. Holla BS, Kalluraya B, Sridhar KR. Synthesis of 4-(5-nitro-2-furfurylidene) amino-3-mercapto-5-(substituted)-l,2,4-triazoles as possible antibacterial agents. Curr Sci 1987; 56: 236–238.
7. Xiaoyun C, Zhang J, Hu H, Yu S, Sun S, Dan Z, Jiang Y, Wu Q. Design, synthesis and biological evaluation of novel triazole derivatives as inhibitors of cytochrome P450 14alpha-demethylase. Eur J Med Chem 2009;44:1913–1920.
8. Takuya U, Yoshiko K, Toshiyuki K. Amide analogs of antifungal dioxane-triazole derivatives: Synthesis and in vitro activities. Bioorganic and medicinal chemistry letters 2009;19:2013–2017.
9. Bastikar VA, Fulsundar SR, Nair JS. In silico docking analysis of peptide deformylase(PDF)- A novel target for prophylaxis of leptospirosis. Nature Precedings 2008.
10. Reena G, Subhash A, Sudhir AK. Modelling and interactions of *Aspergillus fumigatus* lanosterol 14-alpha demethylase "A" with azole antifungals. Bioorg Med Chem 2004;12:2937–2950.
11. Subhash A, Sudhir AK. A dual response partial least squares regression QSAR model and its applications in design of dual activators of PPARα and PPAR γ. QSAR and combinatorial science 2008;27(11–12):1291–1304.
12. Hall LH, Mohney BK, Kier LB. The electrotopological state: structure information at the atomic level for molecular graphs. J Chem Inf Comput Sci 1991;31:76–82.
13. Baumann K. An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features. J Chem Inf Comput Sci 2002;42:26–35.