

Applications of Chemometric Methods to Elucidate Physicochemical Requirements for Binding of Cyp51 Inhibitors to Its Target

Aarti Singh

Department of Pharmacy, Banasthali University Rajasthan

Email: modernchem.aarti@gmail.com

(Received January 28, 2014)

Abstract: Invasive fungal infection has prevailed in past two decades causing high mortality and morbidity rate, especially in immunocompromised patients. Classical quantitative structure-activity relationship (QSAR analysis) was executed to study correlation between molecular structure and CYP51 inhibitor activity of novel aminotetralin derivatives. A di-parametric model with high statistical values $r = 0.84$, $r^2 = 0.71$, $r^2_{cv} = 0.60$, $s = 0.38$ and $f = 23.56$ was generated and validated using leave one out technique and external set of molecules to confirm the predictive power of the generated model. The resultant QSAR model revealed the importance of first atom E-State index and VAMP polarization YZ parameters in inhibition of fungal CYP51 enzyme.

Keywords: CYP51, QSAR, MLR, PLS

1. Introduction

Since 1980s, momentous complication and prevalence for invasive fungal disease have augmented, especially in immunocompromised patients and/or those hospitalized with serious underlying diseases.^{1,2} The opportunistic human pathogen *Candida albicans* and non-albicans have acquired considerable clinical significance as infectious agents, being important causes of morbidity and mortality.³ In eukaryotic metabolic pathways sterol biosynthesis is essential in animals (cholesterol biosynthesis), fungi (ergosterol biosynthesis) and plants (biosynthesis of sitosterol and an array of phytosterols). Based on the surveillance, inhibition of ergosterol biosynthesis in fungi coincided with an accumulation of 14 α -methylsterols.⁴ It was recommended that the primary azole target is the microsomal cytochrome P450 protein implicated in 14 α -sterol demethylation. CYP51 is targeted primarily due to its eminent role in fungal growth.

Lanosterol 14 α -demethylase (P450, CYP51) is a member of the cytochrome P450 superfamily, which catalyses the elimination of 14-methyl

group (C-32) of lanosterol via three successive monooxygenation reactions.⁵ P450 participates in ergosterol biosynthesis and is prominent for fungal viability. Selective inhibition of this enzyme would cause depletion of ergosterol and accumulation of lanosterol and some other 14-methyl sterols resulting in the growth inhibition of fungal cell.

During the past three decades azole compounds have been developed as medical and agricultural agents to fight fungal diseases.⁶ They are classified as imidazoles or triazoles on the basis of number of nitrogen in five membered azole ring.⁷ The varying antifungal potency and spectrum of activity is the result of difference in affinity of azoles for the 14 α -demethylase enzyme. Various evidences indicates that the triazole ring in the scaffold of triazole antifungal is perpendicularly placed to the porphyrin plane with a ring nitrogen atom coordinated to the heme ring of CYP51 and is significant for the antifungal activity.

Quantitative structure-activity relationships (QSAR) has been emerged as helping hand in understanding various aspects of chemical-biological interactions in drug and other scientific research. In QSAR, a mathematical equation which relates chemical structure to their biological activity imparts useful information in view of drug design and medicinal chemistry.⁸ An attempt was made to develop a robust 2D QSAR model for 2-amminotetralins derivatives in order to design potent antifungal drugs. Multiple linear regression (MLR) and partial least square (PLS) analysis models with classical descriptors were implemented to carry out 2-dimensional QSAR.

2. Material and methods

Generation of elementary structure and three-dimensional optimized structure preparation. A dataset of 43 compounds of 2-aminotetralins analogues was selected to perform QSAR studies.⁹ Chemical structures of respective compounds present in the series were sketched using standalone module of discovery studio (version 2.0) and were finally placed in the TSAR work sheet (version 3.3; Accelrys Inc., Oxford, England).¹⁰ Five substituents for each compound were identified using “define substituent” option. Two-dimensional structures were converted into three- dimension in order to understand the ligand-receptor interactions well. For each input structure the CORINA generates one low energy conformation and predicts 3D-coordinates. The information regarding stereo chemical property and the presence of rings along with chain has been considered in order to quantify the model to be generated. Nextlty, “Charge2-derive charges” option was employed to enumerate the partial charge which is a prerequisite for various

structural manipulations. The 3D optimization of structures has been carried out by using “Cosmic- optimize 3D” which includes valence terms and non-bonded terms. It evaluates the total molecular energies by considering the summation of bond angle, bond length, torsional angle, vanderWaals and columbic terms for suitable sets of atoms. The consideration of the conformations with high energy was mark certain with the force field supplied by Cosmic for energy calculations.

Data set preparation and data reduction 43 compounds from the series were distinguished into training set comprising of 28 compounds and test set with 15 compounds. Training set has been used for multiple regression analysis (MLR) and partial least square (PLS) model development whereas, test set was kept to cross check the predictive power of the model.¹¹

TSAR methodology involves around 150 descriptors in order to generate QSAR model. Being an integrated analysis, it aims at interactive investigation of quantitative structure-activity relationships. Assumption of TSAR methodology depicts the importance of suitable sampling of these structural descriptors in understanding their biological properties. The calculation of numerical descriptors and employment of statistics to have a correlation has been automatically enumerates by TSAR software.

The correlation between two of the descriptors may lead to false prediction of the QSAR model due to statistical instability. Also, it results in over prediction with difficult mechanistic interpretation. The descriptor with correlation coefficient of greater than 0.5 and poor correlation with biological activity was eliminated from the work sheet in order to reduce data and inter-correlation between two successive descriptors. Therefore, two independent descriptors, first atom E-state index and VAMP polarization YZ with good correlation with biological activity were retrieved.

3. Model Development

The multiple linear regression method used the biological data as dependent and the calculated descriptors as independent variable to derive finest QSAR model. The generated model is used to compute the relationship between X and Y variable. The positive and negative sign of regression coefficient in the attained regression equation directs the relation of descriptor and biological activity. Conventional regression coefficient (R^2), Fischer's statistic (F) and the standard deviation (s) facilitate the statistical significance of the regression equations.¹²

Two methods internal and external validation techniques have been employed to assure the authenticity of the generated model. The model is internally validated by Leave-one-out (LOO) method in which one molecule is eradicated from the training set followed by recalculation of the model. This has been repeated with each compound until each of them is once omitted. The actual activity of each compound is then compared with its predicted activity. The actual activities of the test set compound were predicted to validate the generated model externally.

Partial least square (PLS) analysis works as an alternative approach to avoid risk of over fitting and enlarges the information regarding each model.¹³ Partial least square (PLS) analysis has been carried away over the same training set which was used to generate the model by multiple linear analysis (MLR) in order to check the robustness and predictive ability of the model. Leave out one row (LOO) was used to validate the generated model using partial least square (PLS). Finally, molecular docking was employed on the most active compound to enumerate related receptor-ligand interactions.

Result and Discussion

Multiple linear regression (MLR) model developed with around 150 descriptors initially resulted in poor statistical values and high inter-descriptors correlation. The model had $R^2 = 0.43$, $s = 0.55$, $f = 9.53$, $r^2_{cv} = 0.31$ and $r = 0.65$. The statistical equation was represented as

Original Data:

$$Y = 12.895951 * X_1 - 0.099354386 * X_2 - 16.089043$$

The improvement of the statistical values is attained by data reduction. Data reduction leads to refinement of descriptors which gave an outcome of better model generation with good statistical values and low inter-correlation. The two independent descriptors, first atom E-state index and VAMP polarization YZ were selected with high t value and good correlation with biological activity. These have been chosen after marking stepping as zero with elimination of less significant variables. The statistical significance was polished by identifying potential 6 outliers and was deleted. These outliers were having high residual values.¹⁴

Finally, the model developed after descriptor refinement and deletion of outliers has been recognized as robust model and fulfilled statistical criteria. The statistical values obtained through this model with two descriptors are $R^2 = 0.71$, $s = 0.38$, $f = 23.56$, $r^2_{cv} = 0.60$ and $r = 0.84$.

The regression equation for this respective model is

Original Data:

$$Y = 15.166237 * X_1 - 0.12042444 * X_2 - 18.631102$$

where X_1 = first atom E-state index, X_4 = VAMP polarization YZ (Whole molecule).

The remaining compounds of the dataset were used as test set for the external validation of the generated model.¹⁵ A correlation coefficient graph between actual activity and estimated activity shows R^2 value of 0.60. Value of R^2 clearly validates the generated model and illustrates its good predictive ability along with high statistical significance.

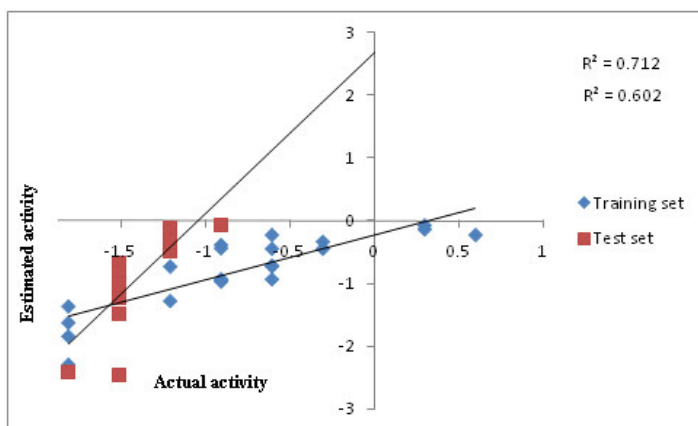


Figure 1: Plot between actual and estimated value of training and test set of MLR

Table II: Actual and predicted activity data obtained from multivariate analysis of the training set compounds

Compound name	Actual activity -log MIC (μM)	Predicted activity (μM)	
		MLR	PLS
6c	-1.80618	-1.63259	-1.65667
8b	-1.20412	-1.2844	-1.37145
9a	-1.80618	-1.84922	-1.93196
9d	-0.90309	-0.96681	-0.93293
10a	0.30103	-0.13516	-0.13109
10b	0.30103	-0.07191	0.046824
10c	-0.60206	-0.22602	-0.18101
11b	-0.60206	-0.44689	-0.44265

11c	-1.20412	-0.73179	-0.73992
11d	-0.90309	-0.93466	-1.02095
12a	-0.90309	-0.38608	-0.34108
12c	0.60206	-0.22684	-0.13628
12d	-0.30103	-0.33444	-0.11665
13b	-0.30103	-0.44744	-0.36257
13c	-0.60206	-0.72934	-0.71239
13d	-0.60206	-0.93614	-0.95241
15b	-0.90309	-0.43868	-0.42401
15d	-1.50515	-0.92488	-0.94114
17c	-0.60206	-0.7187	-0.70055
21b	-1.80618	-1.37101	-1.44358
25b	-1.80618	-2.30388	-2.46567
27a	-0.90309	-0.96495	-1.03273

Table III: Actual and predicted activity data obtained from multivariate analysis for test set compounds

Compound name	Actual activity -log MIC (μM)	Predicted activity (μM)	
		MLR	PLS
8a	-1.20412	-0.40985	-0.44985
8d	-1.50515	-1.4831	-1.53982
9b	-1.20412	-0.47865	-0.45832
9c	-1.50515	-0.76022	-0.70946
11a	-1.50515	-0.66744	-0.61806
12b	-0.90309	-0.07226	-0.12686
13a	-1.50515	-0.68982	-0.64513
14a	-1.20412	-0.1281	-0.17026
14d	-1.20412	-0.32784	-0.3503
16d	-1.20412	-0.33369	-0.35934
25a	-1.50515	-2.45027	-2.3635
25c	-1.80618	-2.40545	-2.31681
27b	-1.50515	-0.84715	-1.04413
27c	-1.50515	-1.00535	-1.16851
27d	-1.50515	-1.21992	-1.33349

Partial least square (PLS) was performed over the same model to confirm the predictive ability and enhancement of the generated model.¹⁶ The r^2 value of 0.71 demonstrates the high predictive ability of the model generated by MLR and PLS.

The statistical equation is

$$Y = 15.166237 * X_1 - 0.12042444 * X_2 - 18.631102$$

where X1 = First Atom E-State index (subst.1), X2 = VAMP Polarization YZ (whole molecule)

Statistical significance = 1.09, Residual sum of square = 6.03, Predictive sum of squares = 7.48, $r^2_{cv} = 0.64$, $r^2 = 0.71$.

The R^2 value for both the analysis was found to be 0.71 which signifies the strong relation. The actual and predicted biological activity values of MLR and PLS analysis for training and test set are shown in Table 3 and the respective plots are shown in Fig. 2.

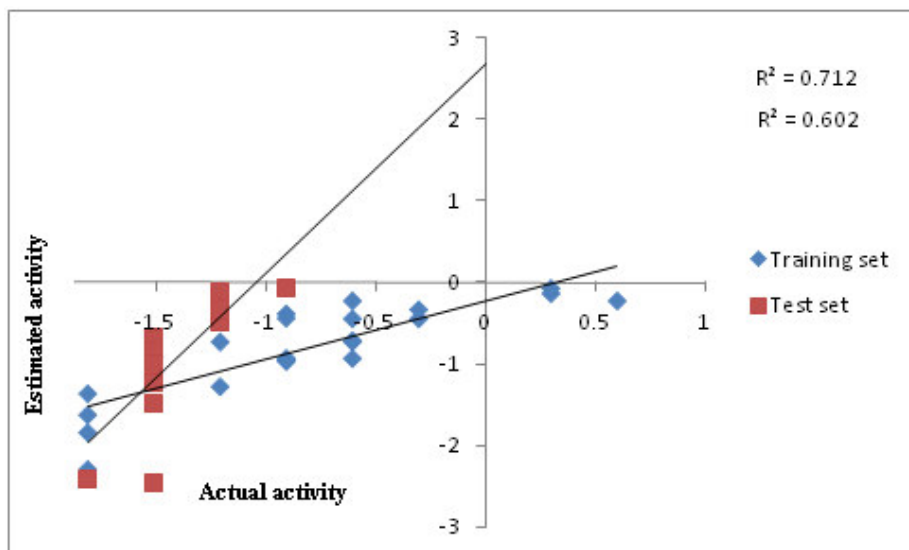


Figure 2: Plot between actual and estimated value of training and test set of PLS

Interpretation of Descriptors

E-state indices incorporate information related to atom types and electron accessibility that are influenced by all of the structural features of a molecule in combination with electronic, topological, and valence state.¹⁷ The aptitude of these indices to categorize a set of molecular structures and to organize them in a chemically meaningful manner, emphasizing electronic molecular information has been demonstrated. This parameter illustrates the positive relation with the biological activity which is well described by 12c (-log MIC: 0.60206) and 25b (-log MIC: -1.80618) with E-state value of 1.1795 and 1.1232.

Vamp is a semiempirical molecular orbital package in TSAR Version 3.3 which illustrates the nuclear repulsion energy and reckons the

electrostatics properties. The vamp nuclear energy explains the nuclear repulsion-driven processes in the molecule and might acquire allied to the conformational changes or atomic reactivity in the molecule.¹⁸ Vamp depicts negative correlation with the biological activity thereby concluding that with the decrease in electrostatic nature of the substituents there would be an impetus in biological activity of lead molecule. If there are more electron withdrawing groups there will be additional atoms to contribute contributing to the energy terms resulting in increased nuclear energy. The increase in nuclear energy goes with the reduction of biological activity. The correlation is further explained by the descriptor value for the most active 12c (-4.2832) and the least active 25b (5.9032).

Table IV: Correlation of biological activity of active and inactive molecules of training set with all three descriptors

	Name of Compound	Biological activity -log MIC (μ M)	First Atom E-State index (subst.1)	VAMP Polarisation XZ (whole molecule)
Active Compound	12c	0.60206	1.17949	-4.28316
Inactive Compound	25b	-1.80618	1.12317	5.87175

Fig. 3 shows the alignment of most active compound (12c) within the active site of CYP51 receptor. The compound showed hydrogen bond interaction with HIS 447 and van der Waal interaction with ILE 450.

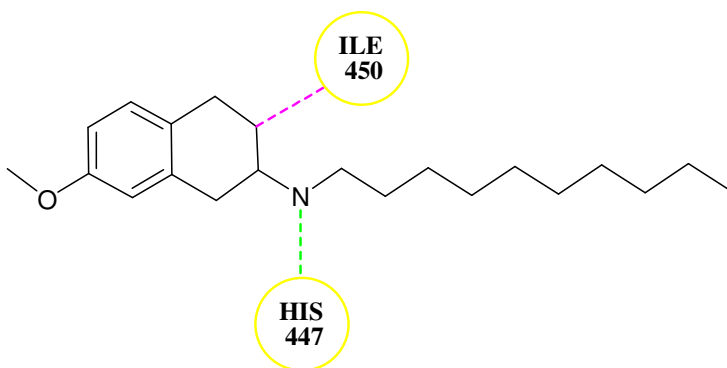


Figure 3: Binding mode of most active compound into the crystal structure of CYP51.

Green dotted line represents hydrogen bond interaction and pink dotted line represents Van Der Waal interaction.

Conclusion

In our study, multiple linear regression (MLR) and partial least square (PLS) was used to generate model which have been further validated. The MLR and PLS analysis showed the robustness of the model. The acquired model can be used to design novel 2-aminotetralins derivatives.

References

1. M M McNeil, S L Nash, R A Hajjeh, M A Phelan, L A Conn, B D Plikaytis, D W Warnock. Trends in mortality due to invasive mycotic diseases in the United States, 1980–1997. *Clin. Infect. Dis.*, **33** (2001) 641–647, *Res Toxicol.*, **14**(2001) 1284–1295
2. S K Fridkin, The changing face of fungal infections in health care settings. *Clin. Infect. Dis.*, **41** (2005) 1455–1460.
3. P R Hsueh, Y J Lau, Y C Chuang, J H Wan, W K Huang, J M Shyr, J J Yan, K W Yu, J J Wu et al. Antifungal susceptibilities of clinical isolates of *Candida* species, *Cryptococcus neoformans*, and *Aspergillus* species from Taiwan: Surveillance of Multicenter Antimicrobial Resistance in Taiwan program data from 2003. *Antimicrob Agents Chemother*, **49**(2005) 512–517.
4. M S A Khan, I Ahmad, F Aqil et al., Virulence and Pathogenicity of Fungal Pathogens with Special Reference to *Candida albicans*. *Combating Fungal Infections*, **21** (2010) 45.
5. Y Aoyama, Y Yoshida, R Sato. Yeast cytochrome P-450 catalyzing lanosterol 14 - demethylation. II. Lanosterol metabolism by purified P-450_{14DM} and by intact microsomes. *J. Biol. Chem.*, **259** (1984) 1661–1666.
6. D Lamb, D Kelly, S Kelly. *Molecular aspects of azole antifungal action and resistance.*, **2**(6) (1999) 390–402.
7. K P Conner, P Vennam, C M Woods et al., 1,2,3-Triazole-Heme Interactions in Cytochrome P450: Functionally Competent Triazole-Water- Heme Complexes, *Biochemistry*, **51** (2012) 6441–6457.
8. H Kubinyi. Quantitative structure-activity relationships (QSAR) and molecular modelling in cancer research. *J Cancer Res Clini Oncol*, **116** (1990) 529-537.
9. B Yao, H Ji, Y Cao et al., Synthesis and Antifungal Activities of Novel 2-Aminotetralin Derivatives, *J. Med. Chem.*, **50** (2007) 5293-5300.
10. E Besalu, Trends and plot methods in MLR studies. *J Chem Inf Model*, **47**(2007) 751–760.
11. D Debieu, J Bach, A Lasseron et al., Effects of sterol biosynthesis inhibitor fungicides in the phytopathogenic fungus, *Nectria haematococca*: ergosterol depletion versus

precursor or abnormal sterol accumulation as the mechanism of fungitoxicity, *Pesticide Science*, **54**(1998) 157–167.

12. K H Kim, Outliers in SAR and QSAR: Is a flexible binding site a possible source of outliers? *J Comput Aided Mol Des.*, **21**(2007) 421–435.
13. H Kubinyi, QSAR and 3D-QSAR in drug design part 1: methodology, *Drug Discov Today*, **2**(1997) 457–467.
14. R D Cramer, Partial least squares (PLS): its strengths and limitations. *Perspect Drug Discov Des.*, **1** (1993) 269–278.
15. S K Paliwal, M Pal, A A Siddiqui, Quantitative structure activity relationship analysis of angiotensin II AT1 receptor antagonists, *Med Chem Res.*, **19** (2010) 475–489.
16. S Paliwal, A Narayan, S Paliwal, Quantitative structure activity relationship analysis of dicationic diphenylisoxazole as potent anti-trypanosomal agents. *QSAR Comb Sci.*, **28** (2009) 1367–1375.
17. I Jabeen, P Wetwitayaklung, P Chiba, M Pastor, G F Ecker, 2D and 3D-QSAR studies of a series of benzopyranes and benzopyrano[3,4b][1,4]- a series of oxazines as inhibitors of the multidrug transporter P-glycoprotein. *Journal of Computer Aided Molecular Design*, **27** (2013) 161-71.
18. M T D Cronin, T W Schultz, Development of quantitative structure–activity relationships for the toxicity of aromatic compounds to *Tetrahymena pyriformis*: comparative assessment of the methodologies, *Chem. Res. Toxicol*, **14** (2001) 1284 – 1295.