J Young Pharm, 2020;12(1):48-54.
A multifaceted peer reviewed journal in the field of Pharmacy
www.jyoungpharm.org | www.phcog.net

Original Article

2D and 3D QSAR Studies of Saponin Analogues as Antifungal Agents against *Candida albicans*

Bhushan A Baviskar*, Sharada L Deore, Amruta I Jadhav
Government College of Pharmacy, Kathora Naka, Amravati, Maharashtra, INDIA.

ABSTRACT

Objectives: Present communication deals with two- and three-dimensional (2D and 3D) QSAR studies of twenty saponin analogue for antifungal activity against *Candida albicans*. **Methods:** The 2D-QSAR model for the prediction was obtained by applying Multiple Linear Regression (MLR) method, giving r^2 = 0.8551 and q^2 = 0.7717 and Partial Least Squares (PLS) method, giving r^2 = 0.8551 and q^2 = 0.7717. 3D-QSAR study was performed using the stepwise variable selection k-nearest neighbour molecular field analysis (kNN-MFA) approach for both electrostatic and steric fields. Two different kNN-MFA methods (SA and GA) were used for the building of 3D-QSAR models. **Results:** The best model shows interesting result in terms of internal (q^2 > 0.62) and external (predictive r^2 > 0.52) predictivity for training and test set. **Conclusion:** Thus, QSAR models showed that hydrophobic and electrostatic effects dominantly determine the binding affinities. Hence the QSAR models proposed in this work would be further useful for development of new antifungal agents from medicinal plants and can help in the design of novel potent molecule.

Key words: Saponin, QSAR, MLR, PLS, Candida.

Correspondence

Bhushan A Baviskar

Government College of Pharmacy, Kathora Naka, Amravati-444 604, Maharashtra, INDIA.

Email: baviskarbhushan@gmail.com

DOI: 10.5530/jyp.2020.12.10

INTRODUCTION

Over the past few decades, extreme use of antimicrobial agents has led to a worldwide health issue of antibacterial resistance. *Candida albicans* is the pathogen having infectious properties particularly in immunocompromised patients. As a consequence, there is constant demand of development of new anti-fungal agents.[1]

Saponins plays an important role in protection of plant against attack by pathogens and posses wide variety of biological activities such as antifungal, antibacterial, anti-inflammatory, anticancer, antiprotozoal, Immunomodulatory and hypoglycemic.[2-11] The antifungal mechanism of saponin is not well understood but it is proposed as that it forms complex with sterols in cell membrane, leading to pore formation and further loos of membrane integrity.[12,13]

The quantitative structure–activity relationship (QSAR) has been proved to be useful approach for the prediction of biological activities, particularly in computational chemistry.[14,15] These include both 2D (two dimensional) and 3D (three-dimensional) QSAR methods. The major differences of these methods can be analyzed from two viewpoints: first the structural parameters that are used to characterize molecular identities and second the mathematical procedure that is employed to obtain the quantitative relationship between a biological activity and the structural parameters.[16-18] In an effort for search of new potent antifungal agents from medicinal plants, 2D- and 3D-QSAR of saponin analogues performed to quantify necessary structural and physicochemical requirements of this series of compounds as potent antifungal agents.

MATERIALS AND METHODS

2D-QSAR methodology

Data set: In this study, a series of 20 triterpenoid saponins (Figure 1) and their MIC values against *Candida albicans* were taken from the literature.[19-24] The activity data (MIC) of each molecule were converted into logarithmic scale [pMIC = -log (MIC×10^{-6})] was used as dependent variable for 2D- and 3D-QSAR analysis and it is listed in Table 1 and Table 2. All molecular modeling studies (2D and 3D) were performed using the Molecular Design Suite (VLife MDS software package, version 3.5; from VLife Sciences, Pune, India). All structures were sketched using 2D draw application are cleaned and 3D optimized. Energy minimization and geometric optimization were conducted using the Merck molecular force field method with the root mean square gradient set to 0.01 kcal/molÅ, the maximum number of cycles was 10,000 and medium's dielectric constant of 1 by batch energy minimization method. Energy-minimized geometry was used for calculation of descriptors.

Selection of training and test set

The dataset of 20 molecules was divided into training set (12 compounds) and test set (8 compounds) for multiple linear regression (MLR) and partial least squares (PLS) model by stepwise forward backward variable selection methods and various 2D descriptors as independent variables. The unicolumn statistics of test and training sets (Table 3) showed the accurate selection of test and training sets, as the maximum of the training set was more than that of the test set and the minimum of the training set was less than or equal to that of the test set.

Molecular descriptors

The various molecular descriptors require for 2D-QSAR study. A large number of theoretical 2D individual descriptors such as Mol. Wt., Volume, XlogP, smr; physiochemical such as Estate Numbers, Estate contributions, Polar Surface Area, Element Count, Dipole moment, Hydrophobicity XlogpA, Hydrophobicity SlogpA; topological such as T_2_Cl_6, T_C_Cl_6, T_T_S_7, T_T_Cl_7 type have been computed for these geometrically optimized structures from the chemical structures of the compounds. The descriptors having the same value or highly

correlated with other descriptors were removed initially, as they do not contribute to the QSAR. The reduced set of descriptors was then treated by Forward Stepwise Variable Selection for further reduction of non-significant descriptors and finally the optimum models with four significant descriptors were considered in our 2D-QSAR analysis.

Statistical parameters

Dataset of 20 molecules was subjected to regression analysis using MLR and PLS as model building methods. QSAR models were generated using pMIC values as the dependent variable ŷ (biological activity) and various descriptors values as independent variables xi (molecular descriptor) by using linear equations. The cross-correlation limit was set at 0.5, term selection criteria as r^2, F-test 'in,' at 4 and 'out' at 3.99, r^2 and F-test. Variance cutoff was set at 0, scaling to autoscaling and number of random iterations to 10. Statistical measures were used for the evaluation of QSAR models were the number of compounds in regression n, regression coefficient r^2, number of descriptors in a model k, F-test (Fisher test value) for statistical significance F, cross-validated correlation coefficient q2, predictive squared correlation coefficients pred_r^2, coefficient of correlation of predicted data set pred_r^2 se and standard error (SE) of estimation r^2 se and q^2 se.

MLR analysis

Multiple Linear Regression (MLR) analysis is the traditional and standard approach for multivariate data analysis. It is based on ordinary least square regression (OLS) method. MLR is a method used for modelling linear relationship between a dependent variable Y (pMIC) and independent variable X (2D descriptors). MLR estimates values of regression coefficients (r^2) by applying least squares curve fitting method. The model creates a relationship in the form of a straight line (linear) that best approximates all the individual data points.

The multiple regression equation takes the form as mentioned in Equation (1)

$$Y = b1 * x1 + b2 * x2 + b3 * x3 + c \quad \text{---------- (1)}$$

Where Y is dependent variable, the 'b's are regression coefficients for corresponding 'x's independent variable, 'c' is a regression constant for intercept.[25]

PLS regression method

PLS analysis is a popular regression technique which can be used to specified linear relationship between dependent variable (Y) to several independent (X) variables even when factors are many and highly collinear. PLS creates orthogonal components using existing correlations between independent variables and corresponding outputs while also keeping most of the variance of independent variables. Main aim of PLS regression is to predict the activity (Y) from X and to describe their common structure.[26] PLS is probably the least restrictive of various multivariate extensions of MLR model.

Validation of QSAR model

The generated QSAR model was validated by the internal stability and predictive ability inside the model.

***Internal validation*:** Internal validation was carried out using leave-one-out (q^2, LOO) cross validation method. For calculating q^2, one object (one biological activity value) is eliminated from training set and training dataset is divided into subsets (number of subsets = number of data points) of equal size.

***External validation*:** The predictive ability of the selected model was also confirmed by external validation of test set compounds which is also denoted with pred_r^2.

Randomization test

To evaluate the statistical significance of the QSAR model for an actual dataset, one-tail hypothesis testing is employed. The robustness of the QSAR models for experimental training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging biological activities of the training set molecules. The significance of the models hence obtained was derived on calculated Z score.[27]

3D-QSAR studies

Data set and molecular modeling

The total set of compounds was divided into a training set (14 compounds) for generating 3D-QSAR models and a test set (6 compounds) for validating the quality of the models. The SE method was adopted for division of training and test data set comprising of 14 and 6 molecules, respectively, having dissimilarity value of 7.9 with pMIC activity field as dependent variable and various 3D descriptors calculated for the compounds as independent variables. The most active compound in the dataset is selected as the starting point for building a sphere. Six compounds, namely, 2, 3, 12, 14, 18 and 19, were used as test set while the remaining molecules were used as the training set (Table 4). The unicolumn statistics of the training and test sets are reported in Table 3.

Alignment procedure

Molecular alignment is a crucial step in 3D-QSAR based on moving of molecules in 3D space, which is related to the conformational flexibility of molecules. Conformational search was carried out by systemic conformational search method (grid search), which generates all possible conformations, by systematically varying each of the torsion angles of a molecule by some increment, keeping the bond lengths and bond angles fixed and lowest energy conformers were selected. All the compounds were aligned by template-based method. In template-based alignment method, a template structure was defined and used as a basis for alignment of a set of molecules. These aligned conformations were used to generate the predictive QSAR model. Multiple conformation of each molecule was generated using Monte Carlo conformation search method. It is a random search method for finding the conformations of molecules which uses the metropolis condition to accept or discard generated conformers.[28]

Calculation of field descriptors

Using Gasteiger-Marsili charge type[29] electrostatic and steric field descriptors were calculated with cut offs of 10.0 kcal/mol and 30.0 kcal/mol for electrostatic and steric respectively. The dielectric constant was set to 1.0, considering distance-dependent dielectric function. Probe setting was carbon atom with charge +1.0. 3D-QSAR analysis was performed after removal of all the invariable columns, as they do not contribute to QSAR.

k-Nearest neighbour molecular field analysis (kNNMFA)

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its kNN in training set.[30] The steric and electrostatic interaction energies are computed at lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules.

kNN-MFA with simulated annealing (SA)

SA is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to

the Boltzmann distribution so that at equilibrium, low energy states are the most populated.

kNN-MFA with genetic algorithm (GA)

GA mimics natural evolution and selection. In biological systems, genetic information that determines the individuality of an organism is stored in chromosomes. This method employs a stochastic variable selection procedure, combined with kNN, to optimize (i) the number of nearest neighbours (k) and (ii) the selection of variables from the original pool as described in simulated annealing.

RESULTS

MIC values against *Candida albicans* were taken from the literature and activity data (MIC) of each molecule were converted into logarithmic scale [pMIC = -log (MIC×10^{-6})] was used as dependent variable for 2D- and 3D-QSAR analysis and it is listed in Table 1 and Table 2 respectively. Table 3 and 4 are of unicolumn statistic of training and test set of 2D and 3D QSAR models respectively. The frequency of use of a particular descriptor in the population of equations indicated the relevant contributions of the descriptors (Table 5). Various statistical parameters in the 2D QSAR are shown in Table 6. The stepwise forward backward variable selection method resulted in several statistically significant models as shown in Table 7. The molecular descriptors which are contributing in 3D QSAR study are shown in Table 8.

Figure 1 is a series of 20 triterpenoid saponins selected for study from literature. Figure 2 shows Contribution plot for model 1 reveals that the Estate contribution descriptors like SsCH3E-index contributing inversely and SdsCHE-index are contributing positively as 75%, 25% respectively to biological activity. Figure 3 of the contribution plot for model 2 reveals that the Estate contribution descriptor such as SdssCE-index is contributing 100% to biological activity. Figure 4 and 5 is showing 3D-QSAR models with important steric and electrostatic points contributing to the models with range of values shown in parenthesis.

DISCUSSION

2D-QSAR equations were selected by optimizing the statistical results generated along with variation of the descriptors in these models. The fitness/pattern plots were also generated for evaluating the dependence of the biological activity on various different types of the descriptors. Statistically significant QSAR models were selected for discussion.

Model-1 (MLR)

MIC (*Candida albicans*) = -42.4282(±4.8546) SsCH3E-index +50.7691(±16.8111) SdsCHE-index +791.7581.

where n = 12 training and 8 test, DF = 9, r^2 = 0.855, q^2 = 0.771, F-test = 26.550, r^2 se = 0.452, q^2 se = 0.427, pred_r^2 = 0.639

The experimental and predicted activities with residual value are shown in Table 1. Following Figure 2 shows Contribution plot for model 1 reveals that the Estate contribution descriptors like SsCH3E-index contributing inversely and SdsCHE-index are contributing positively as 75%, 25% respectively to biological activity.

Model-2 (PLS)

MIC = +118.0474SdssCE-index+343.3258.

where *n* = 12 training and 8 test, DF = 10, r^2 = 0.7889, q^2 = 0.727, F-test = 37.365, r^2 se = 0.195, q^2 se = 0.295, pred_r^2 = 0.578

Table 1: Experimental and Predicted activity of saponin analogue in 2D QSAR study.

Sr.No.	Structure	Antifungal activity against *Candida albicans* pMIC[a]					
		MLR model			PLS model		
		Exp.	Pred.	Residual	Exp.	Pred.	Residual
1.	Arvenoside B A7	5.50	4.01	1.49	5.50[T]	3.65	1.85
2.	Barrigenol family A19	4.57	3.85	0.72	4.57[T]	3.55	1.02
3.	Barrigenol family A25	4.54	-	-	4.54	-	-
4.	Chenopodium Quinoa 2	3.299	3.42	-0.121	3.299	3.27	0.029
5.	Chenopodium Quinoa 10	3.300	3.31	-0.01	3.300	3.40	-0.1
6.	Chenopodium Quinoa 1	3.297[T]	3.488	-0.191	3.297[T]	3.313	-0.016
7.	Chenopodium Quinoa 3	3.295[T]	3.36	-0.065	3.295	3.317	-0.022
8.	Chenopodium Quinoa 4	3.294[T]	3.422	-0.128	3.294	3.276	0.018
9.	Chenopodium Quinoa 5	3.292[T]	3.361	-0.069	3.292	3.32	-0.028
10.	Chenopodium Quinoa 6	3.290	3.31	-0.02	3.290	3.279	0.011
11.	Chenopodium Quinoa 7	3.301	3.315	-0.014	3.301	3.39	-0.089
12.	Chenopodium Quinoa 8	3.296[T]	3.221	0.075	3.296	3.38	-0.084
13.	Chenopodium Quinoa 9	4[T]	3.26	0.74	4[T]	3.320	0.68
14.	Maesabalide family A17	4.50	-	-	4.50[T]	4.14	0.36
15.	Maesabalide family A21	4.78	-	-	4.78	4.76	0.02
16.	Oleanane and ursane 1	4.19[T]	3.67	0.52	4.19[T]	3.35	0.84
17.	Oleanane and ursane 2	4.79	3.707	1.09	4.79[T]	3.571	1.22
18.	Oleanane and ursane 3	3.88	3.85	0.03	3.88	3.572	0.31
19.	Phytolaccoside B	3.90[T]	3.39	0.51	3.90	3.458	0.442
20.	Sakurasosaponin A2	4.56	-	-	4.56[T]	3.851	0.709

Expt. =Experimental activity, pred. =Predicted activity; a= -log (MIC×10^{-6}); T= Test set

Figure 1: Saponins selected for QSAR study.

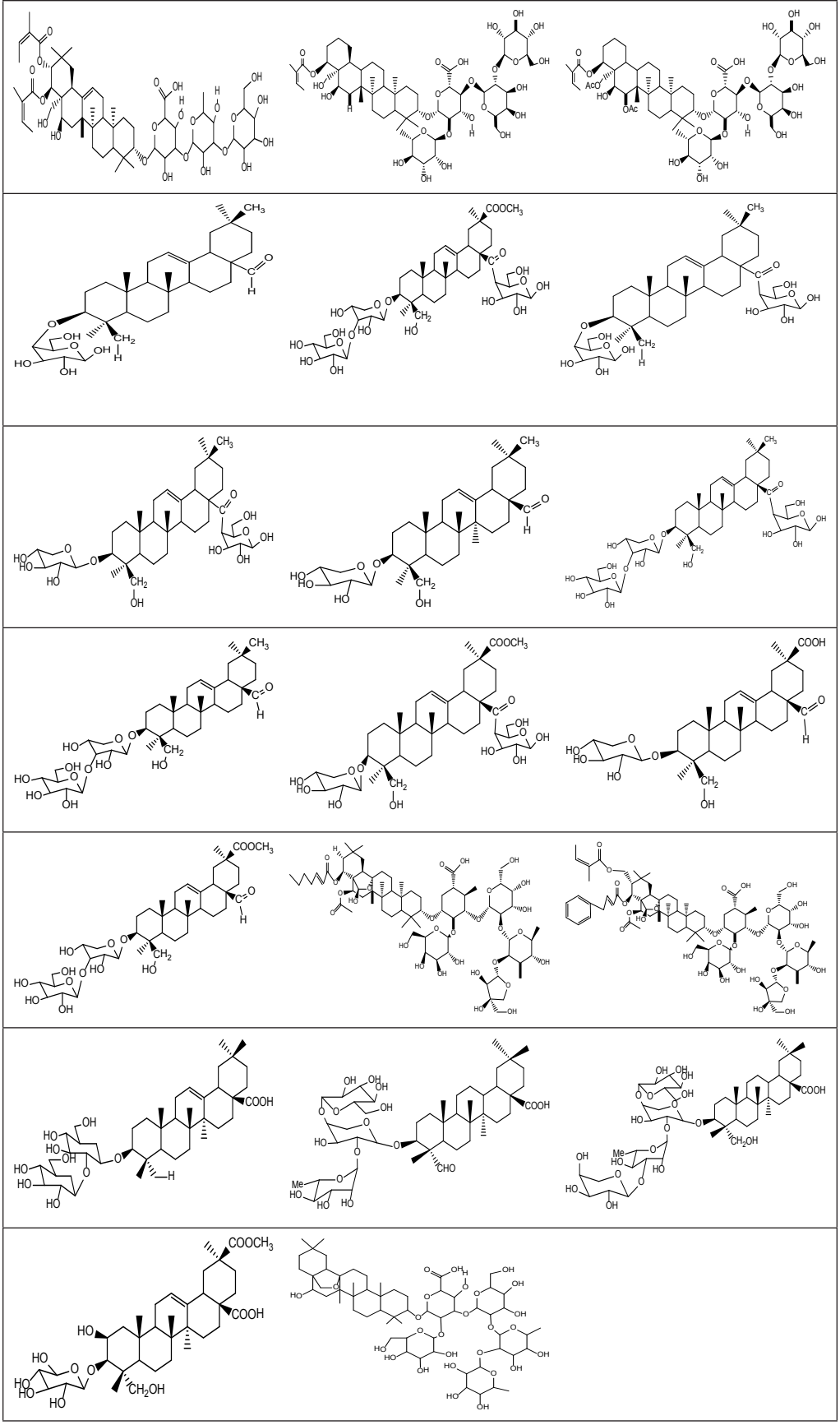

Table 2: Experimental activity and predicted activity of 3D QSAR models.

Sr.No.	Structure	Antifungal activity against *Candida albicans* pMIC[a]					
		Model 1(SA)			Model 2(GA)		
		Exp.	Pred.	Residual	Exp.	Pred.	Residual
1.	Arvenoside B A7	5.50	4.68	0.82	5.50	0	5.5
2.	Barrigenol family A19	4.57	3.58[T]	0.99	4.57	4.38	0.19
3.	Barrigenol family A25	4.54	0[T]	4.54	4.54	5.03	-0.49
4.	Chenopodium Quinoa 2	3.299	3.2962	0.0028	3.299	3.2955	0.0035
5.	Chenopodium Quinoa 10	3.300	3.2967	0.0033	3.300	3.2950	0.005
6.	Chenopodium Quinoa 1	3.297	3.2967	0.0003	3.297	3.2954	0.0016
7.	Chenopodium Quinoa 3	3.295	3.2954	-0.0004	3.295	3.2974	-0.0024
8.	Chenopodium Quinoa 4	3.294	3.2965	-0.0025	3.294	3.2982	-0.0042
9.	Chenopodium Quinoa 5	3.292	3.3005	-0.0085	3.292	3.2997	-0.0077
10.	Chenopodium Quinoa 6	3.290	3.2959	-0.0059	3.290	3.2959	-0.0059
11.	Chenopodium Quinoa 7	3.301	3.2965	0.0045	3.301	3.2950	0.006
12.	Chenopodium Quinoa 8	3.296	3.3006[T]	-0.0046	3.296	3.2949	0.0011
13.	Chenopodium Quinoa 9	4	0	4	4	0	4
14.	Maesabalide family A17	4.50	3.5079[T]	0.9921	4.50	3.2975	1.2025
15.	Maesabalide family A21	4.78	5.3561	-0.5761	4.78	4.5594	0.2206
16.	Oleanane and ursane 1	4.19	0	4.19	4.19	4.8244	-0.6344
17.	Oleanane and ursane 2	4.79	4.8642	-0.0742	4.79	4.2514	0.5386
18.	Oleanane and ursane 3	3.88	4.4178[T]	-0.5378	3.88	3.5646	0.3154
19.	Phytolaccoside B	3.90	402618[T]	-4026	3.90	3.5694	0.3306
20.	Sakurasosaponin A2	4.56	5.3781	-0.8181	4.56	4.7796	-0.2196

Expt. =Experimental activity, pred. =Predicted activity; a= -log (MIC×10^{-6}); T= Test set.

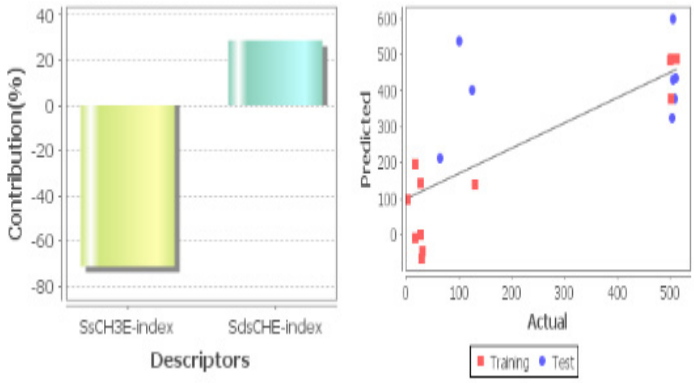

Figure 2: A. Contribution plot for model 1 by MLR method. B. Data fitness plot for model 1 by MLR method.

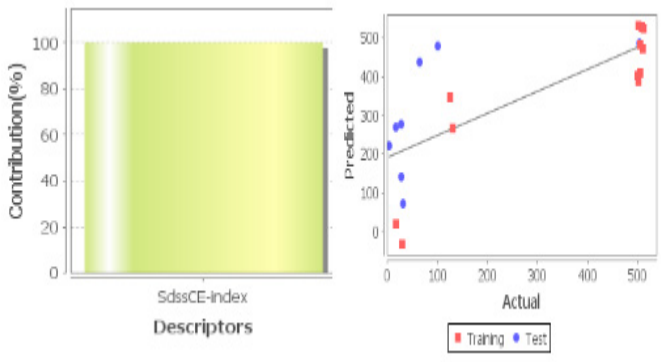

Figure 3: A. Contribution plot for model 2 by PLS method B. Data fitness plot for model 2 by PLS method.

The structures, experimental and predicted activity with residual value are shown in Table 1. Figure 3 shows the contribution plot for model 2 reveals that the Estate contribution descriptor such as SdssCE-index is contributing 100% to biological activity. In the 2D QSAR various statistical parameter (Table 6) were used to evaluate the models includes, number of compounds in regression n, Degree of freedom, the regression coefficient r^2, Cross validated coefficient correlation q^2, F test (Fischer's test) For the statistical significance F, The regression coefficient is a relative measure of fit by the regression equation. This shows the part of variation in observed data explained by regression. The r^2 value is greater than 0.7 or closer to 1.0 gives the best fit. Also, the q^2 value is always greater than 0.5 gives best fit. High value of F test shows model is significant. Predicted r^2 is for the external test set shows predictive capacity of model the Pred_r^2 greater than 0.5 shows that the model is good predicted. Also, Z score is calculated by the q^2 in randomization. Best_ran_q^2 is the highest q^2 value in the randomization test and α_ran_q^2 is the statistical significance parameter obtained by randomization test.

3D-QSAR modelling was performed using kNN-MFA method that adopts a kNN principle for generating relationships between molecular fields and antifungal activity against *Candida albicans*. The kNN-MFA models were generated using training set of 14 compounds and 3D-QSAR models were validated using a test set of six compounds. The steric (S) and electrostatic (E) descriptors specify the regions, where variation in the structural features of different compounds in training

Table 3: Unicolumn Statistic of training and test set of 2D QSAR models.

Models	Data set	Column name	Average	Max.	Min.	SD	Sum
Model 1 (MLR)	Training	MIC ug/ml	191.19	512.00	3.10	233.00	2294.30
	Test	MIC ug/ml	312.75	510.00	64.0	212.97	2822.00
Model 2 (PLS)	Training	MIC ug/ml	362.01	512.00	16.50	214.33	4344.20
	Test	MIC ug/ml	96.51	504.00	3.10	167.45	772.10

Table 4: Unicolumn Statistic of training and test set of 3D QSAR models.

Models	Data set	Column name	Average	Max.	Min.	SD	Sum
Models	Training	MIC	305.0071	512.0000	3.1000	241.2060	4270.1000
	Test	MIC	141.0333	505.0000	26.5000	184.7672	846.2000

Table 5: Molecular descriptor contributing in present 2D QSAR study.

Estate contribution Descriptor	Description
SsCH3E-index	Electrotopological state indices for number of -CH3 group connected with one single bond.
SdsCHE-index	Electrotopological state indices for number of –CH group connected with one double and one single bond.
SdssCE-index	Electrotopological state indices for number of carbon atom connected with one double and two single bonds.

Table 6: Statistical parameters of MLR and PLS methods for 2D QSAR.

Parameter	*Candida albicans*	
	MLR model	PLS model
N	12	12
DF	9	10
r^2	0.8551	0.7889
q^2	0.7717	0.7270
F test	26.550	37.365
r2se	0.452	0.195
q2se	0.427	0.295
pred_r2	0.639	0.578
pred_r2_se	0.5842	0.5457
best_ran_r2	0.430	0.4523
best_ran_q2	0.2490	0.3397
Best_ran_pred_r2	0.54706	0.31157

Where, N - Number of molecules, **K** - Number of descriptors in a model, **DF** - Degree of freedom (higher is better), $\mathbf{r^2}$ - Coefficient of determination (> 0.7), $\mathbf{q^2}$ - Cross-validated r (>0.5), pred_$\mathbf{r^2}$ - r for external test set (>0.5), **F-test** - F-test for statistical significance of the model (higher is better, for same set of descriptors and compounds).

set leads to increase or decrease in activities. The number accompanied by descriptors represents its position in 3D MFA grid. The stepwise forward backward variable selection method resulted in several statistically (Table 7) significant models, of which following models considered as the best one. Nearness of experimental and calculated activity value described in Table 2 are also adding to this fact. The molecular descriptors which contributing in 3D QSAR study is shown in Table 8. Fitness plot

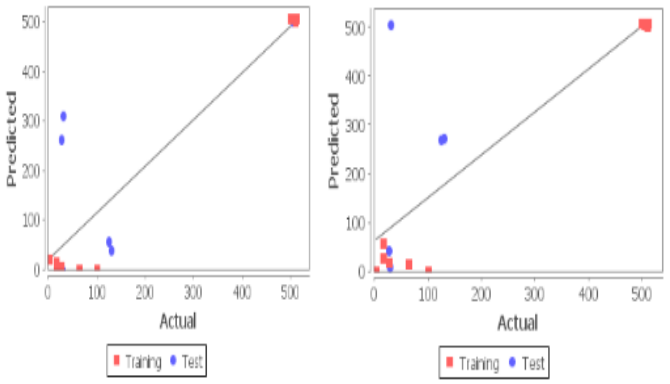

Figure 4: A. data fitness plot for model 1 by SA method B. Data fitness plot for model 2 by GA method.

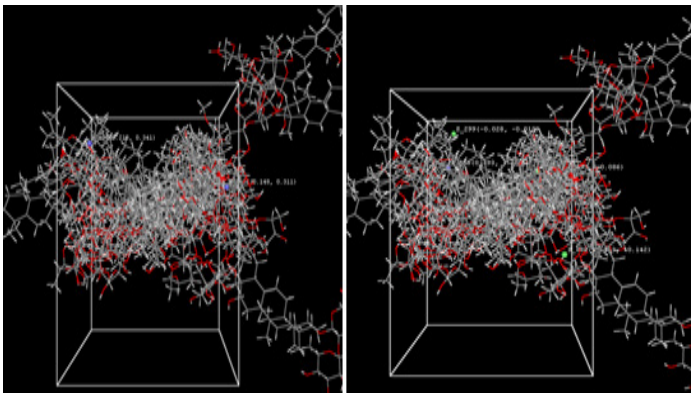

Figure 5: Contour plots of 3D-QSAR models with important steric and electrostatic points contributing to the models with range of values shown in parenthesis.

Table 7: Statistical evaluation of 3D QSAR models of simulated annealing, Geometrical algorithm.

Parameters	3d QSAR models (knn method)	
	Model 1 (Simulated Annealing)	Model 2 (Genetic Algorithm)
KNN	2	2
Degree of Freedom	11	9
q^2	0.9798	0.9808
q^2_Se	34.3014	33.4007
Pred_r^2	0.5562	0.2003
Pred_r^2Se	171.6730	230.4404

The best model selection is based on the values of q^2, pred_r^2, q^2_se and pred_r^2_se.

(q^2= internal predictive ability of the model; pred_ r^2= the ability of the model to predict the activity of external test set).

Table 8: Molecular descriptor contributing in present 3D QSAR study.

3D descriptors	
Electrostatic	Steric
E_190 (SA)	S_299 (GA)
E_1006 (SA)	S_880 (GA)
E_285 (GA)	S_802 (GA)

for 3D QSAR studies is shown in Figure 4 and contour plots of 3D-QSAR models with important steric and electrostatic points contributing to the models with range of values shown in parenthesis is shown in Figure 5.

Model 1-SA

pMIC = E_190 (0.2175 0.3411) and E_1006 (0.1482 0.3106)

The 3D QSAR model 1 gives values of k (2), q^2 (0.9798), pred_r^2 (0.5562), q^2_se (34.3014), pred_r^2_se (171.6730) prove that QSAR equation so obtained is statistically significant.

Model 2-GA

pMIC = S_299 (-0.0283 -0.0128), S_880 (-0.4826 -0.1415), S_802 (-0.0935 -0.0860) and E_285 (0.3928 1.1382)

The 3D QSAR model-2 gives the values k (2), q^2 (0.9808), pred_r^2 (0.2003), q^2_se (33.4007) and pred_r^2_se (230.4404). Prove that QSAR equation so obtained is statistically significant.

CONCLUSION

It is concluded that statistically significant 2D/3D-QSAR models were generated with the purpose of deriving structural requirements for the inhibitory activities of saponin analogues against *Candida albicans*. The best 2D-QSAR model indicates that the descriptors SsCH3E-index contributing inversely and SdsCHE-index, SdssCE-index are contributing positively as 75%, 25% and 100% respectively to biological activity. KNN-MFA investigated the substitution requirements for the receptor-drug interaction and constructed the best 3D-QSAR models providing useful information in characterization and differentiation of their binding sites. In conclusion, the information provided by the robust 2D/3D-QSAR models will be useful for the design of new molecules and hence, this method is expected to provide a good alternative for the drug design.

ACKNOWLEDGEMENT

Authors are thankful to entire team of V life MDS for their constant support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

QSAR: Quantitative structure-activity relationship; **MLR:** Multiple Linear Regressions; **PKE:** Pulsatilla koreana; **N:** Optimum number of components; **r^2:** Square of correlation coefficient; **q^2:** Cross-validated correlation coefficient; **pred_r^2:** r^2 for external test set; **F-test:** Fischer's value.

REFERENCES

1. Poole K. Overcoming antimicrobial resistance by targeting resistance mechanisms. J Pharm Pharm. 2001;53(3):283-94.
2. Morrissey JP, Osbourn AE. Fungal resistance to plant antibiotics as a mechanism of pathogenesis. Microbiol Mol Biol Rev. 1999;63(3):708-24.
3. Bruneton J. In Pharmacognosie, Phytochimie, Plantes Médicinales Technique and Documentation, Paris. 1995;1120.
4. Oleszek W, Bialy Z. Chromatographic determination of plant saponins-an update (2002–2005). J Chromatography: A. 2006;1112(1-2):78-91.
5. Elisa B, Giuliano B, Vincenzo A, Behzad Z, Ebrahim S, Felice S, *et al*. Saponins from *Allium minutiflorum* with antifungal activity. Phytochemistry. 2007;68(5):596-603.
6. Benziane M, Kambuche BN, Bellahouel S, Merah B, Fortas Z, Soulimani R, *et al*. Antimicrobial activity of the alkaloids and saponin extracts of Anabasis articulate. J Biotechnol Pharm Res. 2012;3(3):54-7.
7. Pilar N, Rosa MG, Carmen RM, Salvador M, Cerdá-Nicolásb M, José-Luis R. *In vivo* anti-inflammatory activity of saponins from *Bupleurum rotundifolium*. Life Sci. 2001;68(10):1199-206.
8. Liancai Z, Jun T, Bochu W, Chao Z. *In-vitro* Antitumor Activity and Antifungal Activity of Pennogenin Steroidal Saponins from Paris *Polyphylla* var. *Yunnanensis*. Iran J Pharml Res. 2011;10(2):279-86.
9. Deore SL, Khadabadi SS. Antiproliferative activity of saponin fractions of *Chlorophytum borivilianum*. Pharmacogn J. 2010;2(16):29-33.
10. Foubert K, Cuyckens F, Matheeussen A, Vlietinck A, Apers S, Maes L, *et al*. Antiprotozoal and antiangiogenic saponins from *Apodytes dimidiata*. Phytochemistry. 2011;72(11-12):1414-23.
11. Upendra S, Neeraj K, Bikram S, Renuka KM, Supriya B. Immunomodulatory active steroidal saponins from *Asparagus racemosus*. Med Chem Res. 2013;22(2):573-9.
12. Deng Y, He K, Ye X, Chen X, Huang J, Li X, *et al*. Saponin rich fractions from *Polygonatum odoratum* (Mill.) Druce with more potential hypoglycemic effects. J Ethnopharmacol. 2012;141(1):228-33.
13. Mert-Turk F. Saponin versus plant fungal pathogens. Canakkale Onsekiz mart university, Turkey. J Cell Mol Biol. 2006;5:13-17.
14. Pourbasheer E, Riahi S, Ganjali MR, Norouzi P. Application of genetic algorithm-support vector machine (GA–SVM) for prediction of BK-channels activity. Eur J Med Chem. 2009;44(12):5023-8.
15. Darnag R, Mazouz ELM, Schmitzer A, Villemin D, Jarid A, Cherqaoui D. Support vector machines: Development of QSAR models for predicting anti-HIV-1 activity of TIBO derivatives. Eur J Med Chem. 2010;45(4):1590-7.
16. Dlaz HG, Prado FJ. Unified QSAR and network-based computational chemistry approach to antimicrobials, part 1: Multispecies activity models for antifungals. J. Comput Chem. 2008;29(4):656-67.
17. Hemmateenejad B, Sancholi M. Substituent electronic descriptors for fast QSAR/QSPR. J Chemometr. 2007;21(3-4):96-107.
18. VLife MDS. 3.5 Molecular Design Suite. Vlife Sciences Technologies Pvt. Ltd, Pune, India. 2004. www.vlifesciences.com
19. Jeffrey JC, Ikechukwu O, George PT, Edward BH, Florence FW, *et al*. Characterization of plant derived saponin natural products against *Candida albicans*, ACS Chem. Biol. 2010;5(3):321-32.
20. Woldemichael GM, Wink M. Identification and biological activity of triterpenoidal saponin from Chenopodium Quinoa. J Agric Food Chem. 2001;49(5):2327-32.
21. Escalante AM, Santecchia CB, López SN, Gattuso MA, Gutiérrez RA, Delle MF, *et al*. Isolation of antifungal Saponins from *Phytolacca tetramera,* an Argentinean species in critic risk. J Ethnopharmacol. 2002;82(1):29-34.
22. Hua S, Weishuo F, Weishuo CH. Structure activity relationships of Oleanane and ursane type triterpenoids. Bot Stud. 2006;47(4):339-68.
23. Ekabo OA, Farnsworth NR, Henderson TO, Mao G, Mukherjee R. Antifungal and Molluscicidal saponins from *Serjania salzmanniana*. J Nat Prod. 1996;59(4):431-5.
24. Bedir E, Kirmizipekmez H, Sticher O, Caliş I. Triterpene saponins from the fruits of *Hedera helix*. Phytochemistry. 2000;53(8):905-9.
25. Croux C, Joossens K. Influence of observations on the misclassification probability in quadratic discriminant analysis. J Multivar Anal. 2005;96(2):348-403.
26. Kubyani H. Variable selection in QSAR studies: An evolutionary algorithm. Quant Struct Act Relat. 1994;13(3):285-94.
27. Zheng W, Tropsha A. Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci. 2000;40(1):185-94.
28. Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H, Tropsha A. Quantitative structure–activity relationship analysis of functionalized amino acid anticonvulsant agents using k-nearest neighbor and simulated annealing PLS methods. J Med Chem. 2002;45(13):2811-23.
29. Zheng W, Tropsham A. Novel variable selection quantitative structure–property relationship approach based on the k-nearestneighbor principle. J Chem Inf Comput Sci. 2000;40(1):185-94.
30. Holland J. Adaptation in natural and artificial systems. University of Michigan Press. 1976.

Article History: Submission Date : 20-12-2019 ; Revised Date : 27-12-2019 ; Acceptance Date : 28-01-2020
Cite this article: Baviskar BA, Deore SL, Jadhav AI. 2D and 3D QSAR Studies of Saponin Analogues as Antifungal Agents against *Candida albicans*. J Young Pharm. 2020;12(1):48-54.