IJPSR (2014), Volume 5, Issue 10 **(Research Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 26 March 2014; received in revised form, 15 May 2014; accepted, 12 July 2014; published 01 October 2014

METAPHORICAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (2D & 3D-QSAR) ANALYSIS OF TYLOPHORINE DERIVATIVES AS EFFICACIOUS IN ANTIMALARIAL DRUG DESIGN

S. K. Patel [1], L. George [1], V. M. Khedkar [3], M. Y. Lone [4], P. C. Jha [* 4], Y. T. Jasrai [2], H. A. Pandya [2] and H. N. Highland [1]

Department of Zoology, Biomedical Technology and Human Genetics [1], Department of Bioinformatics, Applied Botany Center (ABC) [2], University School of Sciences, Gujarat University, Ahmedabad - 380009, Gujarat, India.
Combi. Chem. Bio. Resource Center, Organic Chemistry Division [3], National Chemical Laboratory, Pashan Road, Pune - 411008, Maharashtra, India.
School of Chemical Sciences [4], Central University of Gujarat, Sector - 30, Gandhinagar - 382030, Gujarat, India.

Keywords:

Plasmodium kinases, Quantitative Structure Activity Relationship (QSAR), Multiple Linear Regression (MLR), Partial Least Squares Regression (PLS), CoMFA, CoMSIA, *Tylophora indicia*

Correspondence to Author:
Dr. Prakash Chandra Jha

Assistant Professor,
School of Chemical Sciences,
Central University of Gujarat, Sector - 30, Gandhinagar - 382030, Gujarat, India.

E-mail: prakash.jha@cug.ac.in

ABSTRACT: Quantitative Structure-Activity Relationship (QSAR) for tylophorine derivatives acting as Plasmodium kinases inhibitors has been developed. The proposed 2D-QSAR model was found to be statistically significant concerning training, cross-validation, and external validation. The contribution of carbon chains with aromatic and electronegative features are found to be the most important descriptors in predicting Plasmodium kinases inhibitory activity. Furthermore, a 3D-QSAR study using CoMFA and CoMSIA was carried out to investigate the molecular property determinants. The CoMFA model suggested that the favorable substitution of steric groups with electronegative properties at ring A. Furthermore, CoMSIA model indicated the possible role of H bond donor groups at 5 and 6 positions of ring A whereas H bond acceptors at 7 and 8 positions of ring A in enhancing the biological activities. The developed QSAR models can be used to discover new effective antimalarial leads for further development as antimalarial drugs.

INTRODUCTION: The extended focus of the malaria society on the eradication of this disease has arisen both in responses to the moral imperative to save lives and eliminate the devastating disease burden and in recognition that a pure case-management approach is unsustainable [1].

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.5(10).4325-38

This article can be accessed online on www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.5(10).4325-38

Thus renewed efforts are being made to develop insecticides that overcome known resistance pathways and kill all mosquitos, to deliver vaccines that protect infants and children and to discover, develop and deliver medicines that not only clear the asexual blood stage parasites but also the asymptomatic forms (exo-erythrocytic stages) that secure the lifecycle of Plasmodium [2].

Malaria is an infectious disease caused by Plasmodium species, transmitted through Anopheles mosquitoes, a severe disease that still caused about 225 million cases and 781,000 human deaths. The plasmodium affects the human red

blood cells and damages them over time. Infection is usually accompanied by some symptoms, including fever, anemia, and splenomegaly [3]. About 50% of the world population lives in endemic areas, situated mostly in Africa, Asia, and South-America. Resistance has been developed over time to all of the five major classes of antimalarial drugs. Till the date no approved vaccine is available, so new effective and affordable drugs are urgently needed [4].

Tylophorine and its derivatives are phenanthroindolizidine alkaloids, also referred to as tylophora alkaloids isolated from *Tylophora indicia*. The leaves of this plant have been used for the treatment of asthma, bronchitis, rheumatism, and dysentery in India. Several key metabolic enzymes, including thymidylate synthase and dihydrofolate reductase, have been reported as biological targets of tylophorine alkaloids. Tylophorine derivatives also inhibit activator protein-1 mediated, CRE-mediated, and nuclear factor kappaB (NF-κB)-mediated transcription. Tylophorine arrests the cells at G1 phase in HepG2, HONE-1, and NUGC-3 carcinoma cells and down-regulates cyclin A2 expression [5].

Though tylophorine derivatives have shown significant antiviral, anti-inflammatory, antitumor, antiameobicidal and anticancer inhibitory activities, the antimalarial activity of these derivatives is yet unexplored. As part of our interest in plant-derived antimalarial agents, we choose to search for new tylophorine derivatives displaying significant antiplasmodial properties for the future development of new antimalarial drug candidates.

One could not, however, confirm that the compounds designed would always possess good inhibitory activity to Plasmodium kinases while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis. Quantitative structure-activity relationship (QSAR) is an area of computational research that builds virtual models to predict quantities such as the binding affinity or the toxic potential of existing or hypothetical molecules. QSAR attempts to predict the biological activity of molecules using the known structural properties.

Although, a wealth of experimental data emphasizes the active role of the target protein in the binding process, QSAR studies are frequently restricted to the properties of the small molecule ligand [6]. QSAR searches information relating chemical structure to biological and other activities by developing a predictive statistical model.

Several molecular descriptors are used to quantify the structural feature of the lead molecule. The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment [7].

The purpose of using QSAR-descriptors is to calculate the properties of molecules that serve as numerical descriptions or characterization of molecules in other calculations such as diversity analysis or combinatorial library design. Using such an approach, one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be synthesized and tested [8].

Developing sound QSAR models using various physicochemical parameters is considered to be an important task in lead optimization. While the classical 2D-QSAR methods are much simpler, faster and more amenable to automation with clearly-defined physiochemical descriptors; 3D-QSAR approaches are useful in investigating the specific substitutional requirements for the favorable receptor-drug interaction providing useful information in the characterization and differentiation of their binding sites.

Moreover, it is now a known fact that the 3D structural information (confirmation) is also responsible for the observed variations in biological activity and the accumulation of 3D structural information about drug molecules has led to the development of 3D-QSAR methods [9]. Ever, since its first inception, Comparative molecular field analysis (CoMFA) [10, 11] and Comparative Molecular Similarity Indices Analysis (CoMSIA) [12-14] have been well established for ligand-based 3D-QSAR. These are useful techniques in understanding the pharmacological profile of studied molecules because not only the 3D-QSAR

models are vivid and robust but also the ensuing steric and electrostatic maps (contours) further aid to understand the nature of the interaction of the ligand with the active site of the enzyme.

Therefore, to gain insights into the structure-activity relationships for tylophorine derivatives as efficacious in antimalarial activity and to understand the mechanism of their substitutional specificity, we report herein some statistically significant QSAR models between antimalarial activity and structural descriptors using 2D and 3D-QSAR approaches. We expect that the theoretical results obtained herein can offer some useful references for understanding the interaction mechanism and directing the design and synthesis of more potent tylophorine analogs for the treatment of malaria.

MATERIALS AND METHODS:

2D-QSAR Analysis: All the computational work required for building 2D-QSAR models was performed on a workstation (Intel Corei5 8-core processor) using VlifeMDS 4.3.2 QSAR plus software developed by VLife Sciences Technologies Pvt., Ltd., Pune, India[15] running on the Windows-7 platform.

Ligand Dataset Workout: A dataset of plant-derived alkaloids was taken from review article[16] to investigate the role of different substitution on these chemical structures in the observed antimalarial activity. These alkaloids have been reported to show significant inhibitory activity against Plasmodium kinases. The biological activity in terms of IC_{50} (also known as half maximal inhibitory concentration, defined as the concentration of an inhibitor required for 50% inhibition of its target) was used in the study.

The respective chemical structures (16 molecules) were drawn using ChemAxon Suite - MarvinSketch software [17] and saved in Tripos Mol2 (.mol2) file format. While the chemical structures of remaining (4 molecules) tyolophorine derivatives were retrieved from PubChem [18] in Structure Data File (SDF) format and converted to Tripos Mol2 (.mol2) file format using chemical file conversion tool OpenBabel [19].

Furthermore, an independent external test set of 6 molecules with known antimalarial activity was retrieved from PubChem [18] in Structure Data File (SDF) format and converted to Tripos Mol2 (.mol2) file format using chemical file conversion tool OpenBabel [19] to validate the QSAR models for their predictive ability.

Molecular Field Descriptor Calculation: It is necessary to calculate descriptors for the set of the molecule to build QSAR models. A descriptor is a quantitative property that depends on the structure of the molecule. Good descriptors are characterized into molecular properties which are important for molecular interactions. Energy-minimized geometry was used for calculation of descriptors, a total of 104 2D descriptors were calculated which encoded different aspects of molecular structure and consists of electronic, thermodynamic, spatial and structural descriptors, *e.g.*, retention index, atomic valence connectivity index, path count, chain path count, cluster, path cluster, element count, estate number, semi-empirical, molecular weight, molecular refractivity, logP and topological index.

In all these descriptors, the neglected descriptors are Dipole Moment, Electrostatic, Distance-Based Topological Indices, Semi-Empirical and Hydrophobicity base log P descriptors (as these are 3D descriptors).

Biological activity: The experimental information associated with biological activity, i.e. pIC_{50} is used as dependent variables to build the QSAR model. The experimental IC_{50} values were evaluated by Joanne Bero *et al*., review article of plant-derived antimalarial compounds [16]. Reported experimental IC_{50} values are converted into a suitable unit μM using online IC_{50} Conversions tool [20]. The negative logarithm of the measured IC_{50} (μM) against Plasmodium kinases is referred to as pIC_{50} [$pIC_{50} = -\log(IC_{50})$]. The term '$pIC_{50}$' is a scale for expressing IC_{50} value exponentially which normalizes the actual activity using the negative logarithmic function, which is considered as a prediction.

Since, some compounds exhibited insignificant activity/no inhibition, such compounds were excluded from the present study. The pIC_{50} values of the molecules under study spanned a wide range from 4 to 8 units [21].

Selection of Training and Test Datasets to Evaluate QSAR Model: The dataset of 16 molecules was divided into a training set (11 Molecules) and test set (5 Molecules) by Sphere Exclusion (SE) method for multiple linear regression (MLR) and partial least squares (PLS) model with dissimilarity value of 1.0 using pIC_{50} activity field as dependent variable and various 2D descriptors as independent variables.

QSAR by Multiple Linear Regression (MLR) Analysis: Multiple regressions is the standard method for multivariate data analysis which is used to model the linear relationship between a dependent variable Y (pIC_{50}) and one or more independent variables X (2D descriptors). The dependent variable is sometimes also called the predict and and the independent variables the predictors. MLR is based on least squares: the model is fit such that the sum-of-squares of differences of observed and predicted values is minimized. MLR measures the values of regression coefficients (r^2) by employing the least squares curve fitting method.

The model generates a relationship in the form of a straight line (linear) that best approximates all theindividual data points. In regression analysis, conditional mean of dependent variable (pIC_{50}) Y depends on (2D descriptors) X. MLR analysis extends this idea to include more than one independent variable. This leads to the following "multiple regression" mean function:

$$Y = B_1 * X_1 + B_2 * X_2 + B_3 * X_3 + C$$

Where Y is the dependent variable, 'B's are regression coefficients for corresponding 'X's (independent variable), 'C' is a regression constant or intercept.

QSAR by Partial Least Square regression (PLS) method: PLS analysis is a popular regression technique which can be used to relate one or more dependent variable (Y) to several independent (*X*) variables. PLS is useful in situations where the number of independent variables exceeds the number of observation, when X data contain colinearities or when N is less than 5 M, where N is a number of the compound and M is a number of the dependent variable. PLS creates orthogonal components using existing correlations between independent variables and corresponding outputs while also keeping most of the variance of independent variables. The main aim of PLS regression is to predict the activity (*Y*) from *X* and to describe their common structure. PLS is probably the least restrictive of various multivariate extensions of the MLR model.

Validation of QSAR Models: The models were rigorously validated internally using correlation coefficient (r^2) and Leave-One-Out (LOO) (q^2) cross-validation. For external validation, these models were used to predict the biological activity of compounds not included in the training set, i.e. test set. For this, the molecular descriptors were calculated for remaining tylophorine derivatives (5 molecules) which formed the test set.

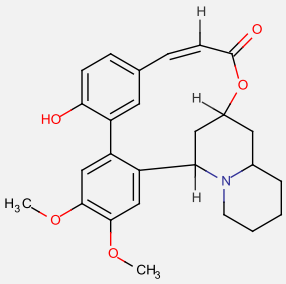
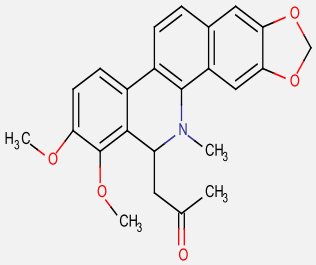
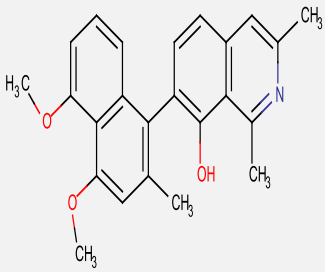
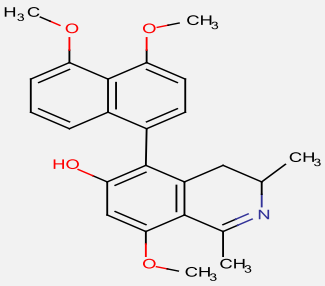
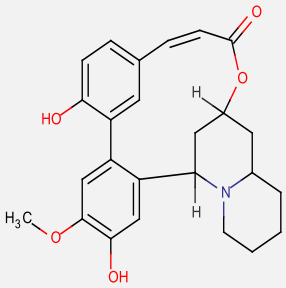
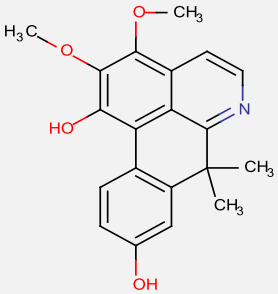
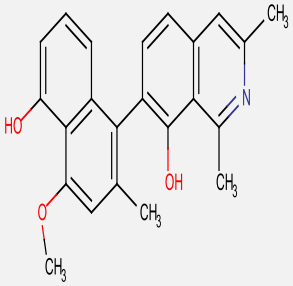
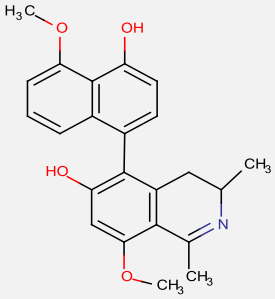
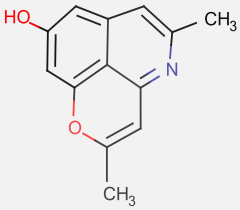
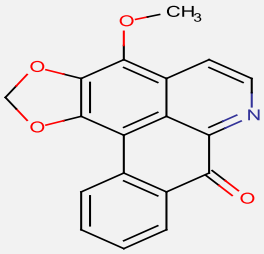
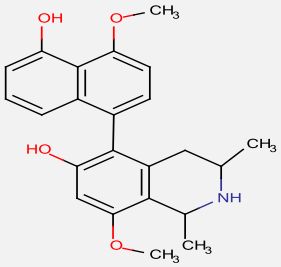
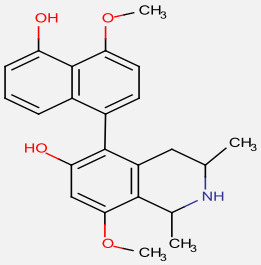
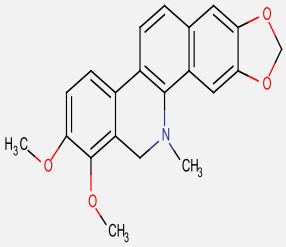
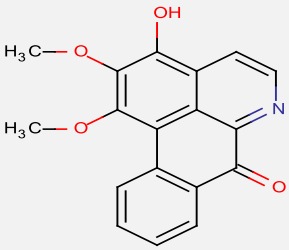
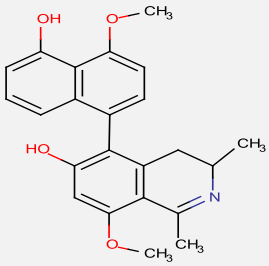
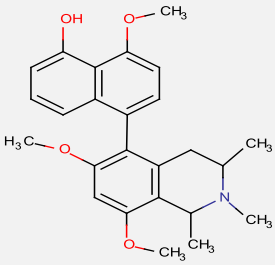
3D-QSAR Analysis (CoMFA and CoMSIA): The 3D-QSAR studies (CoMFA and CoMSIA) were performed with the QSAR module integrated into *Sybyl 7.1* molecular modeling software package from Tripos, Inc., USA[22]. The objective was to find features associated with activity within the molecular scaffold. The 3D-QSAR models were derived using the same distribution of training (11 molecules) and test set (5 molecules) as defined for the 2D-QSAR analysis.

The selection of the training and test sets was done such that the test set compounds had structural diversity and a range of biological activities similar to that of the training set. Partial least squares (PLS) regression was used to derive the 3D-QSAR models wherein the CoMFA, and CoMSIA field descriptors were used as independent variables while the biological activity data (pIC_{50}) served as the dependent variable.

RESULTS AND DISCUSSION:

2D-QSAR Analysis: A 2D-QSAR study of the tylophorine derivatives was performed by using – log of biological activity and various physiochemical descriptors as dependent and independent variable respectively. The correlations were established using Multiple Linear Regression and Partial Least Square analysis. The statistical parameters for assessing the distribution of activity in the training and test sets have been listed in **Table 1** given below:

TABLE 1: STRUCTURES AND STATISTICAL PARAMETERS FOR ASSESSING THE DISTRIBUTION OF ACTIVITY IN THE TRAINING AND TEST SETS

			
(10.9, 4.963)	(5.7, 5.244)	(10.5, 4.979)	(4.4, 5.356)
			
(6.7, 5.174)	(12.2, 4.914)	(8.6, 5.066)	(5.4, 5.268)
			
(0.023, 7.629)	(10.6, 4.975)	(2.5, 5.602)	(4.4, 5.356)
			
(10.6, 4.975)	(10.9, 5.066)	(7.2, 5.143)	(6.2, 5.208)

*IC_{50} and pIC_{50} values are provided in bracket respectively

Among the models generated, the following model was selected based on its statistical significance for further study.

Model-1 for MLR:

BA = 11.548 - 0.215 * 5 Chain Count - 0.432 * SaasCcount - 13.157 * chi6chain = 0.592

The QSAR model-1 for MLR shows that all the molecules are selective for predicting antimalarial activity (See **Fig. 1**).

The significant equation consists of three descriptors, *i.e.* 5 ChainCount, SaasCcount, and chi6chain **Fig. 2**. The model expressed the overall significance level is better, as the calculated F value (102.872) exceeds F value. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. The high value of the F-test indicates that the model is statistically significant. The squared correlation coefficient r^2 (0.97) is a relative measure of the quality of fit by the regression equation.

Correspondingly, it represents 97% variance in PDE-5 inhibitory activity exhibited by tylophorine derivatives.

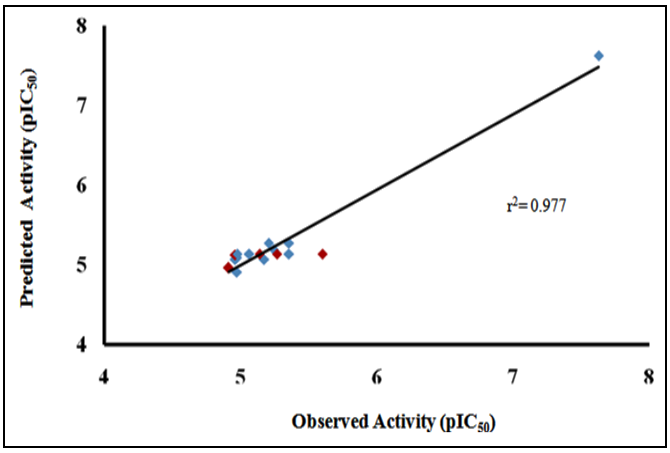


FIG. 1: GRAPHICAL PLOT BETWEEN OBSERVED VERSUS PREDICTED ACTIVITY VALUES FOR TRAINING AND TEST SET COMPOUNDS FOR ANTIMALARIAL INHIBITION

The value of r^2 is close to 1.0, which represents the better fit to the regression line. Standard deviation is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. This standard deviation is an absolute measure of the quality of fit and should have a low value for the regression to be significant. The leave-one-out procedure was used for internal validation of the model. In this procedure, high cross-validated r^2 (q^2 = 0.87), reflects the very good internal predictive power of the model. Another parameter for predictivity of test set compound is the predicted r^2 = 0.50, which shows the good external predictive power of the model. Observed activity versus the predicted activity of both test set and training set are depicted in **Table 2**.

The descriptors selected in the model mentioned above, three descriptors *viz.* 5 ChainCount, SaasCcount, and chi6chain are topological descriptors. These descriptors describe the overall topology of the molecules. The 5 ChainCount is a descriptor signifying the total number of five-membered rings in a compound.

The SaasCcount is a descriptor that defines the total number of carbon connected with one single bond along with two aromatic bonds. The chi6chain descriptor signifies a retention index for six-membered rings.

The descriptor 5 ChainCount is associated with a negative coefficient in the equation, which means reduce the number of 5 membered rings in a molecule to improve the activity. The negative contribution of SaasCcount descriptor revealed the increase of antimalarial activity of tylophorine derivatives with the presence of more number of carbons connected with single bond along with two aromatic bonds. The negative coefficient of SaasCcount (-54%) showed that a decrease in the values of this descriptor is beneficial for activity. Also, the negative coefficient of chi6chain descriptor helps in understanding the effect of the substituent at different position of molecules to increase the activity.

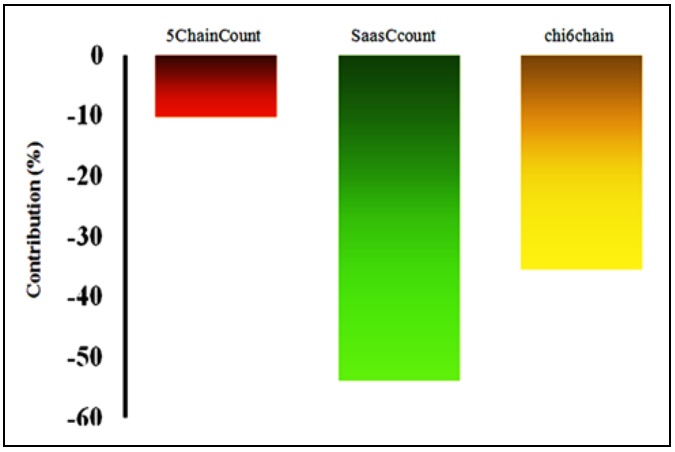


FIG. 2: CONTRIBUTION CHART OF THE SELECTED DESCRIPTORS IN THE MLR MODEL-1

Model-1 for PLS:

BA = 11.583 - 0.468 * SaasCcount - 0.913 * DeltaPsiA - 11.774 * chi6chain

The significant equation consists of three descriptors, *i.e.* SaasCcount, Delta PsiA, and chi6chain **Fig. 3**. The QSAR model-1 for PLS shows that some tylophorine derivatives are selective towards antimalarial inhibition **Fig. 4**. The model presented here explains variance for antimalarial inhibitory activity represented by tylophorine analogs.

The leave-one-out procedure was used for internal validation of the model. In this procedure, high cross-validated r^2 (q^2 = 0.76) and low q^2_se = 0.41 value, reflects the very good internal predictive power of the model and is reasonable check for overfitting of the data.

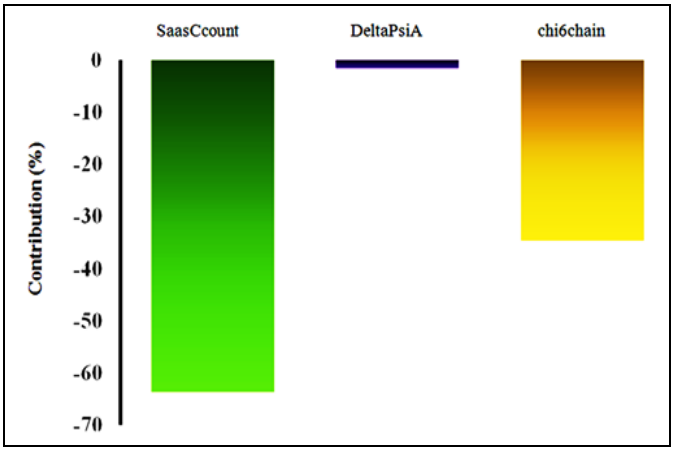


FIG. 3: CONTRIBUTION CHART OF THE SELECTED DESCRIPTORS IN THE PLS MODEL-1

The model demonstrates that the inhibition can be sufficiently explained by the three descriptors like SaasCcount, Delta PsiA, and chi6chain, contribution chart of the selected descriptors in the PLS model-1 is presented in **Fig. 4**.

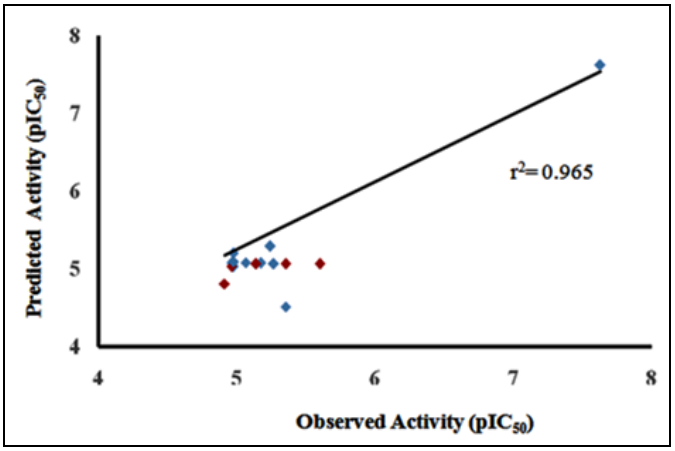


FIG. 4: GRAPHICAL PLOT BETWEEN OBSERVED VERSUS PREDICTED ACTIVITY VALUES FOR TRAINING AND TEST SET COMPOUNDS FOR ANTIMALARIAL INHIBITION

The SaasCcount is a descriptor that signifies the total number of carbon connected with one single bond along with two aromatic bonds. The chi6chain descriptor signifies a retention index for six-membered rings.

Model-2 for MLR:

BA= 12.806 - 4.083* ElectronegativityCount - 0.296 * SaaNcount - 10.403 * chiV6chain -0.538 * SaasCcount

The QSAR Model-2 for MLR developed to consist of a correlation coefficient (r^2) of 0.98, as the coefficient of determination (r^2) was considered. Model-1 can explain 98.47% of the variance in the observed activity values **Fig. 5**. The low standard error of r^2_se = 0.12 demonstrates the accuracy of the model. It shows an internal predictive power (q^2 = 0.87) of 87% and a predictivity for the external test set (pred_r^2 = 0.50) of about 50%. The F-test values of 92.617 show the overall statistical significance level to be of the model, which means that the probability of failure for the model is 1 in 10,000.

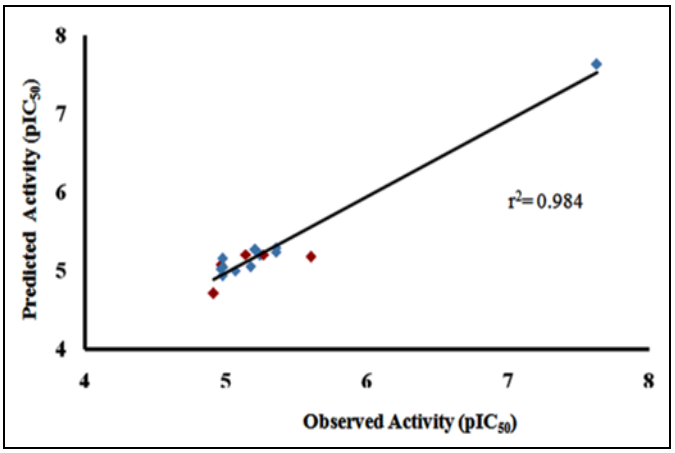


FIG. 5: GRAPHICAL PLOT BETWEEN OBSERVED VERSUS PREDICTED ACTIVITY VALUES FOR TRAINING AND TEST SET COMPOUNDS FOR ANTIMALARIAL INHIBITION

The statistically significant model shows a negative correlation with the ElectronegativityCount, chiV6chain, SaaNcount, and SaasCcount **Fig. 6**. The chiV6chain descriptor signifies atomic valence connectivity index for six-membered rings. The SaaNcount descriptor defines the total number of nitrogen connected with two aromatic bonds. The SaasCcount descriptor defines the total number of carbon connected with one single bond along with two aromatic bonds.

Model-2 for PLS:

BA = 12.147 - 0.826* Electronegativity CountEH - 0.466* SaasCcount - 11.911 * chi6chain

The QSAR Model-2 for PLS generated using the partial least squares analysis method, with 0.96 as the coefficient of determination (r^2) was considered using the same molecules in the test and training sets. The model can explain 96.55% of the variance in the observed activity values **Fig. 7**. The model shows an internal predictive power (q^2 = 0.81) of 81% and predictivity for the external test set

(pred_r^2 = 0.24) of about 24%. The F-test value of 111.925 shows the overall statistical significance level for 99.99% of the model. Out of three descriptors selected in the above-mentioned model, the descriptor Electronegativity CountEH, SaasCcount, and chi6chain are physiological descriptors **Fig. 8**. These descriptors describe the overall topology of the molecules. The SaasCcount defines the total number of carbon connected with one single bond along with two aromatic bonds. The chi6chain descriptor signifies a retention index for six-membered rings [23].

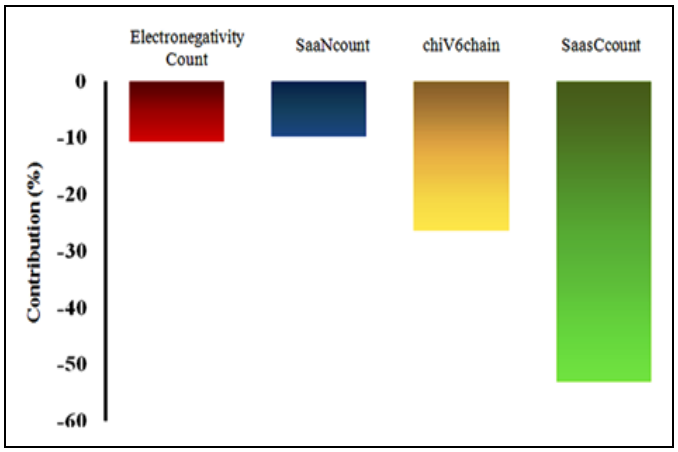


FIG. 6: CONTRIBUTION CHART OF THE SELECTED DESCRIPTORS IN THE MLR MODEL-2

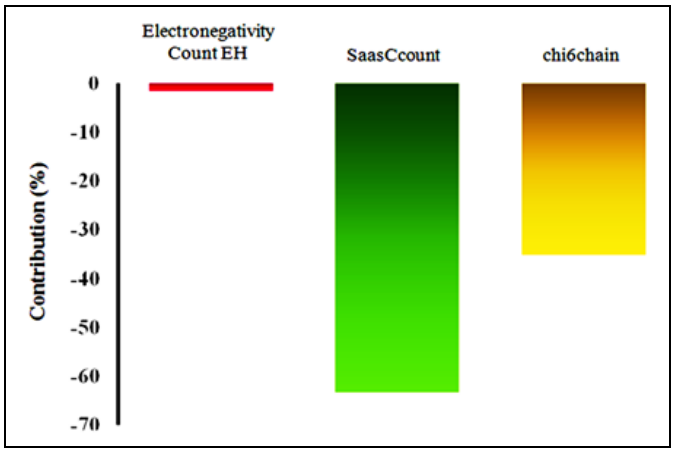


FIG. 8: CONTRIBUTION CHART OF THE SELECTED DESCRIPTORS IN THE PLS MODEL-2

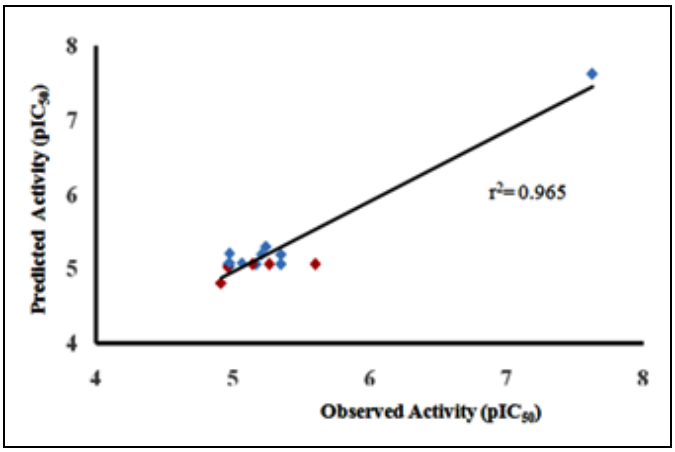


FIG. 7: GRAPHICAL PLOT BETWEEN OBSERVED VERSUS PREDICTED ACTIVITY VALUES FOR TRAINING AND TEST SET COMPOUNDS FOR ANTIMALARIAL INHIBITION

TABLE 2: ACTUAL AND PREDICTED ACTIVITY OF TRAINING AND TEST SET FOR ALL THE MODELS

Compound no.	Observed Activity (pIC_{50})	Predicted Activity (pIC_{50})			
		MLR Model-I	PLS Model-I	MLR Model-II	PLS Model-II
1	5.174	5.071	5.078	5.084	5.071
2	7.629	5.071	5.074	5.053	5.068
3	4.975	7.628	7.621	7.629	7.62
4	5.244	5.087	5.202	5.158	5.206
5	4.914	5.196	5.294	5.203	5.298
6	4.975	4.961	4.81	4.714	4.817
7	4.963	4.91	5.029	4.942	5.031
8	4.979	5.126	5.029	5.027	5.031
9	5.066	5.138	5.088	5.04	5.087
10	5.602	5.138	5.085	5.004	5.084
11	5.143	5.138	5.063	5.18	5.065
12	5.356	5.138	5.063	5.205	5.065
13	5.268	5.138	5.067	5.24	5.068
14	5.356	5.138	5.063	5.205	5.065
15	5.208	5.273	-41.514	5.298	5.196
16	5.174	5.273	10.178	5.272	5.193

TABLE 3: STATISTICAL RESULTS FOR ALL THE MODELS

Compound no.	Statistical parameter	2D QSAR Results			
		MLR Model-I	PLS Model-I	MLR Model-II	PLS Model-II
1	r^2	0.977	0.965	0.984	0.965
2	q^2	0.949	0.767	0.871	0.812
3	pred_r^2	0.440	0.241	0.507	0.248
4	r^2se	0.136	0.159	0.125	0.159
5	q^2se	0.206	0.414	0.356	0.372
6	pred_r^2se	0.255	0.297	0.239	0.295
7	F test	102.872	112.297	92.617	111.925
8	$n_{(training)}$	11	11	11	11
9	Zscore	0.936	0.461	0.519	0.389
10	Best rand_q^2	0.942	0.358	-0.472	0.106
11	Alpha rand q^2	99.000	99.000	99.000	99.000
12	DF	7	8	6	8

Among these 2D QSAR models, MLR model II found to be more statistically significant than the other models **Table 3**. This developed MLR-QSAR model showed considerable values with respect to training (r^2 = 0.98), cross-validation (q^2 = 0.87), and external validation (pred_r^2 = 0.51). This MLR model suggests that the contribution of carbon chains with aromatic and electronegative features are the most important descriptors in predicting plasmodium kinases inhibitory activity **Table 4**.

3D-QSAR Analysis (CoMFA and CoMSIA): The statistical results of CoMFA[10, 11] and CoMSIA[12-14] models are given in Table IV, which shows that all the statistical indices are in an accepted domain.

TABLE 4: PLASMODIUM KINASES INHIBITORY ACTIVITY

	CoMFA	CoMSIA (SHD)
N	16	16
q^2	0.76	0.66
r^2	0.99	0.99
r^2_{pred}	0.61	0.62
r^2_{bs}	0.98	0.97
r^2 *y-scrambling*	0.11	0.18
F	147.71	178.70
Standard error of estimate	0.02	0.01
Field contribution		
Steric	0.440	0.247
Electrostatic	0.560	0.289
Hydrophobic		0.180
H-bond Donor		0.120
H-bond Acceptor		0.164

The CoMFA model generated with a training set of 11 compounds has a cross-validated (leave-one-out) correlation coefficient (r^2_{cv}) of 0.76 and a standard error of 0.02.

The non-cross-validated PLS regression produced a model with correlation coefficient (r^2) of 0.99 with F-test value of 147.71, indicating a good linear correlation between the observed and predicted activities for the molecules in the training set. An r^2 value of 0.98 obtained for 100 runs of bootstrap analysis further advocates the robustness and the statistical validity of the derived CoMFA model while a significantly low r^2 of 0.11 obtained for y-scrambling eliminated the possibility of chance correlation.

The same structural alignment and distribution of training/test sets as defined in the CoMFA was used to derive the CoMSIA models. Several models were generated considering steric (S), electrostatic (E), hydrogen bonding (donor (D) and acceptor (A)) and hydrophobic (H) fields separately or in various combinations. However, the model with all the fields taken together produced the highest cross-validated coefficient (r^2_{cv}) equal to 0.66 having a non-cross-validated r^2 of 0.99, standard error of estimate 0.01 and F-test value 178.70.

The bootstrap r^2 of 0.97 obtained for 100 runs further supports the robustness and the statistical validity of the derived CoMSIA model. Also, a low r^2 of 0.18 obtained for y-scrambling indicate that the model is stable, and the results are not based on chance correlation.

The statistical significance of these models is further supported by the residual between observed *vs.* predicted activity of compounds **Table 5**.

The predictive ability of the 3D-QSAR models was evaluated from the r^2_{pred} calculated for each of these models. This external validation with the test set yielded a predictive r^2 of 0.61 and 0.62 for the CoMFA and CoMSIA models respectively, signifying the ability of the model to predict the activity of untested compounds.

TABLE 5: STATISTICAL SIGNIFICANCE OF RESIDUAL OBSERVED VS PREDICTED ACTIVITY OF COMPOUNDS

Compound ID	Experimental	Predicted
1	4.693	5.179
2	5.174	5.174
3	7.629	6.167
4	4.975	5.240
5	5.244	5.243
6	4.914	4.912
7	4.975	4.974
8	4.963	4.964
9	4.979	4.983
10	5.066	5.065
11	5.602	5.259
12	5.143	5.157
13	5.356	5.341
14	5.268	5.278
15	5.356	5.372
16	5.208	5.193

The results of the 3D-QSAR studies were visualized as 3D 'coefficient contour maps' derived from interpolation of the pairwise products between the PLS coefficients and the standard deviations (stdev*coeff) associated with the corresponding CoMFA or CoMSIA descriptor values.

These maps signify those areas in 3D space where alteration of steric, electrostatic, hydrophobic and hydrogen bonding features in the molecular scaffold correlate strongly with corresponding changes in biological activity.

These contour maps are important tools in drug design, as they show regions in 3D space where modifications of particular molecular fields strongly correlate with concomitant changes in biological activity.

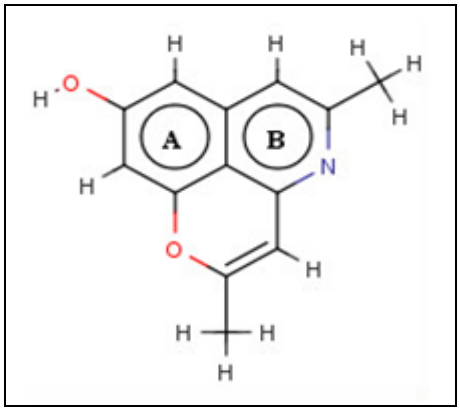


FIG. 9: MOLECULE 3

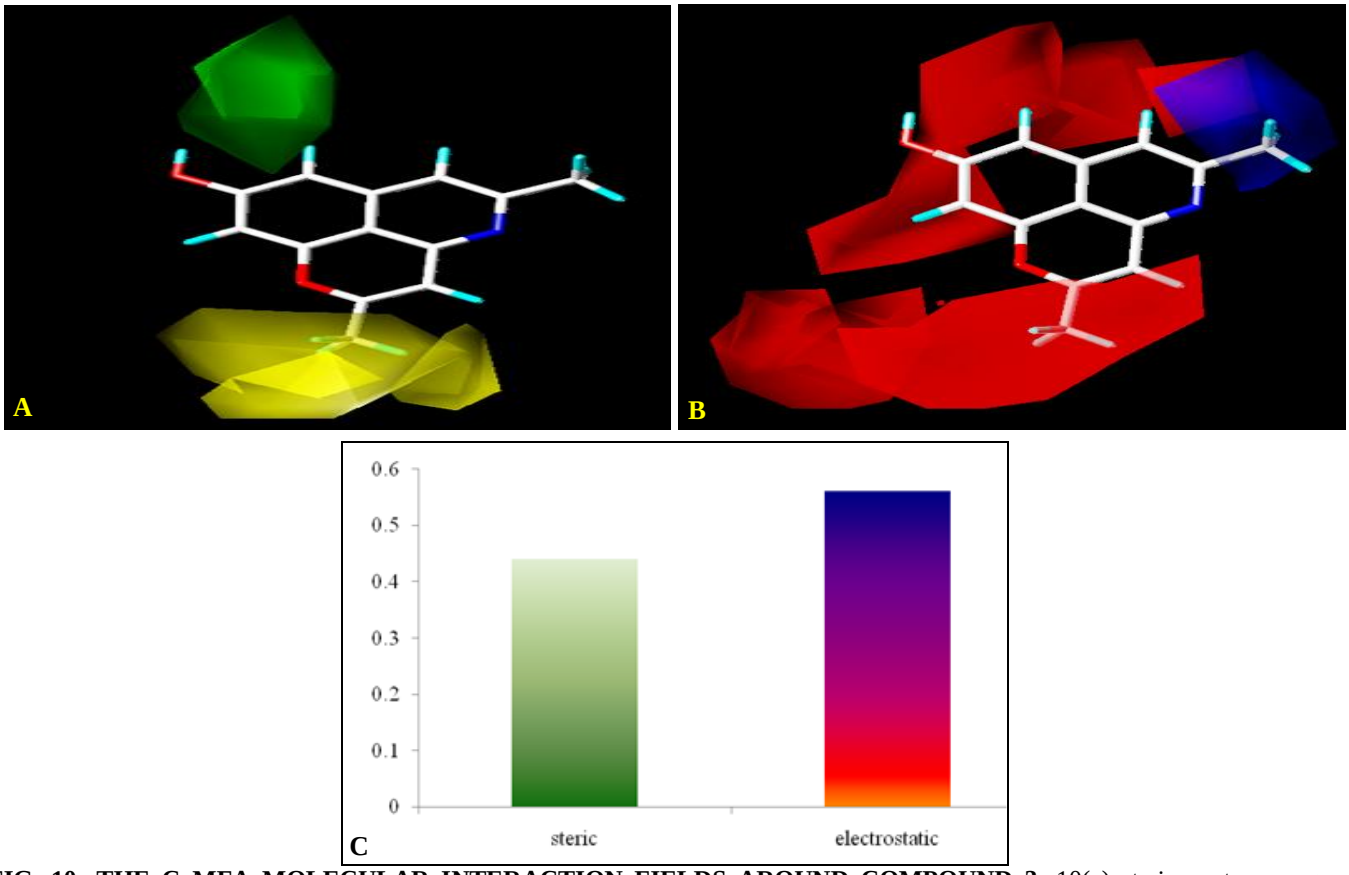


FIG. 10: THE CoMFA MOLECULAR INTERACTION FIELDS AROUND COMPOUND 3. 10(a) steric contours - green contours signify regions where bulky groups increase activity, whereas yellow contours signify regions where bulky groups decrease activity; 10(b) electrostatic contours - red contour signify regions where negative groups increase activity, whereas blue contours indicate regions where negative charge decreases activity; 10(c) CoMFA contribution map.

The steric and electrostatic contour maps obtained from the CoMFA analysis are shown in **Fig. 10(a)** and **10(b)** respectively, associated with reference molecule, 3. Green contours signify regions where steric bulk is favorable and can be exploited to enhance the activity while yellow contours indicate regions where steric bulk is detrimental to the activity. Large green contours are found to be localized around 5 and 6 positions of the ring (A) indicating that increase in the steric bulk over these positions is favorable for the activity while yellow contours disfavoring steric bulk were observed around 7 and 8 positions.

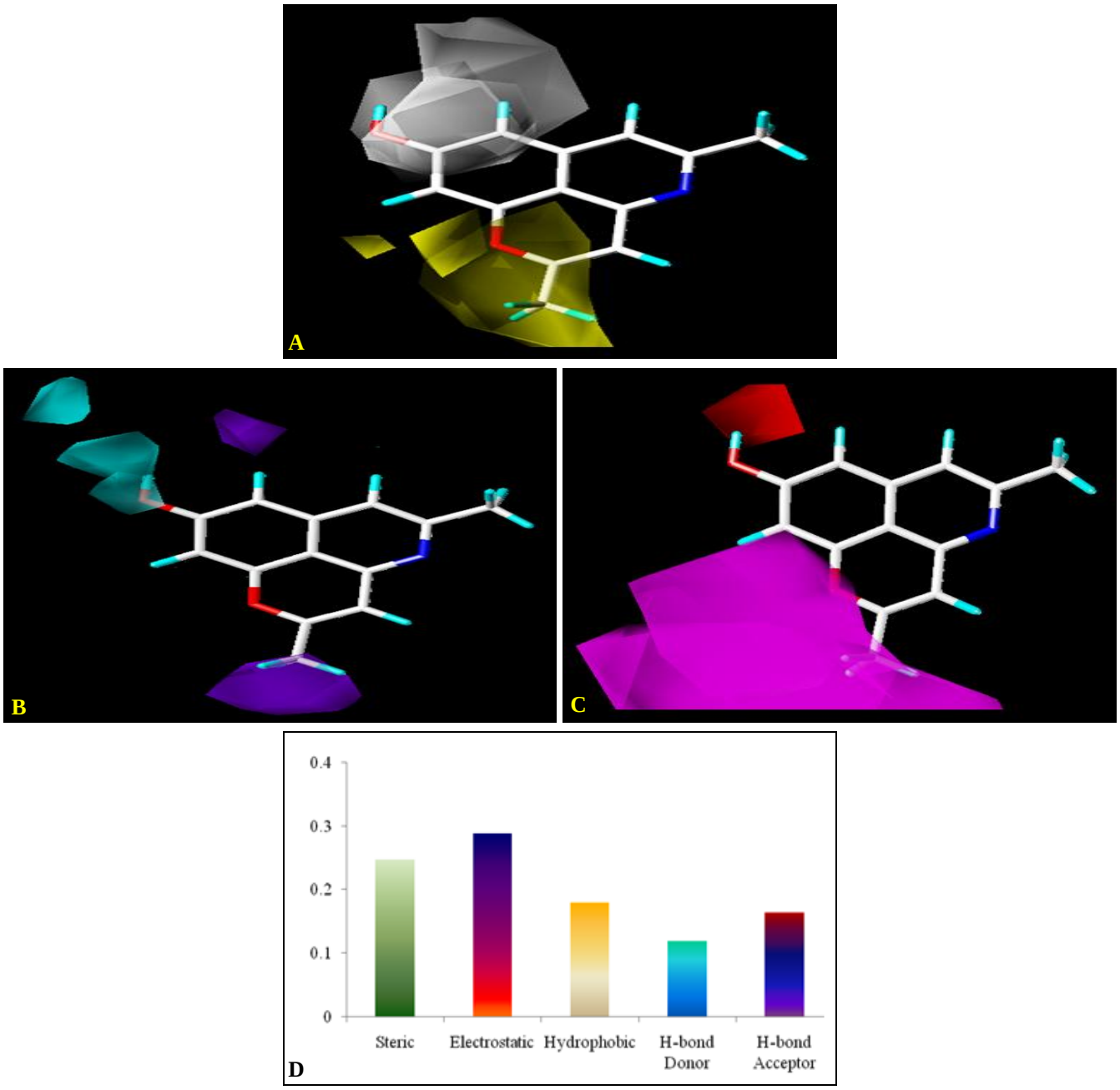


FIG. 11: THE COMSIA MOLECULAR INTERACTION FIELDS AROUND COMPOUND 3; 11(a) Hydrophobic contour plots-white contours indicate regions where hydrophobic groups increases activity, whereas yellow contours indicate regions where hydrophobic group decreases activity; 11(b) hydrogen bond donor contours – cyan contours indicate regions where H-bond donor group increases activity, whereas purple contours indicate regions where H-bond donor group decreases activity; 11(c) hydrogen bond acceptor contours – magenta contours indicate regions where H-bond acceptor group increases activity, whereas red contours indicate regions where H-bond acceptor group decreases activity (however red contour was observed in the present analysis); 11(d) CoMSIA contribution map.

This is in agreement with the experimental results where compounds 11, 12, 13, 14, 15 and 16 having steric substitution around 5 and 6 positions and smaller functional groups around 7 and 8 positions are more found to be active. As regards electrostatic contours, the electronegative contours (red) are found to influence the CoMFA model more than the electropositive counterpart (blue). No significant contribution is observed for electropositive substitution, but both CoMFA, as well as CoMSIA models, displayed large red contours around the ring (A) suggesting that an overall electronegative substitution over this ring is beneficial for the activity.

This is substantiated from the fact that molecules 5, 10, 11, 12, 13, 14, 15 and 16 having more electronegative substitutions over electropositive functionalities showed better activity. Only a small blue contour favoring electropositive substitution is observed around the ring (B).

The CoMSIA model provides additional structural insights into the ligand-receptor interactions by incorporating fields related to hydrophobic and hydrogen bond donor-acceptor interactions in addition to steric and electrostatic properties.

The most active molecule 3 is displayed surrounded by the CoMSIA fields, as an illustration. The CoMSIA steric and electrostatic contours maps were found to be in harmony with that of the CoMFA model and therefore for the sake of brevity have not been discussed again.

The CoMSIA model provides additional structural insights into the ligand-receptor interactions by incorporating fields related to hydrophobic and hydrogen bond donor-acceptor interactions in addition to steric and electrostatic properties. The most active molecule 3 is displayed surrounded by the CoMSIA fields, as an illustration. The CoMSIA steric and electrostatic contours maps were found to be in harmony with that of the CoMFA model and therefore for the sake of brevity have not been discussed again.

The hydrophobic CoMSIA fields are represented by white and yellow contours where white contours signify favored hydrophobicity while yellow contour indicates disfavored hydrophobicity. There is a great correspondence observed between the steric and the hydrophobic contour maps. Large white contours were observed around 5 and 6 positions of the ring (B) indicating that the substituents these positions need to be hydrophobic to enhance the activity while yellow contours disfavoring steric bulk were observed around 7 and 8 positions of the ring (B).

Contour maps representing hydrogen bond donor and acceptor fields were found to be complementary to each other. Hydrogen bond donor fields are represented by cyan and purple contours where cyan contours signify regions favoring hydrogen bond donor group while regions disfavoring such functionalities are represented by purple contours.

Also for hydrogen bond acceptor property, magenta contours represent regions favoring hydrogen bond acceptor groups while red contours signify that presence of H-bond acceptor group is disfavorable. It was observed that cyan contours favoring hydrogen bond donor group cloaked around the 6 positions of the ring (B). Similarly, red contours disfavoring H-bond acceptor groups were observed around this position.

Furthermore, purple contours disfavoring H-bond donor groups were observed around 7 and 8 positions of the ring (B) while magenta contours favoring H-bond acceptor groups were observed around these positions. These findings are in agreement with the experimental results obtained for all the molecules considered in this study.

TABLE 6: PREDICTED ACTIVITY RESULTS

Compound no.	Compound name	Predicted Activity (pIC_{50})					
		MLR Model-I	PLS Model-I	MLR Model-II	PLS Model-II	CoMFA	CoMSIA
e1	Tylophorine	6.056	6.246	6.314	6.243	6.110	6.133
e2	Tinosporine	8.387	8.677	8.343	8.663	8.459	8.652
e3	Tylophorinidine	6.462	6.682	6.819	6.679	6.667	6.701
e4	Tylophorinine	6.462	6.687	6.851	6.682	6.760	6.891
e5	Vinblastine	4.611	5.317	5.680	5.306	4.983	5.001
e6	Vincristine	4.611	5.309	5.627	5.298	4.880	4.987

External Validation: As stated earlier, an external and independent set of 6 molecules with known antimalarial activity were used to further evaluate the predictive power of the resulting QSAR models. The predicted activity for this validation set was found to be in good agreement with the experimental activity, which defends the predictive power of these models **Table 6**.

CONCLUSION: 2D QSAR studies revealed that the carbon chains, along with aromatic features substituted with electronegative groups are influential for biological activities. Ligand-based CoMFA and CoMSIA interaction fields obtained from the 3D-QSAR analysis will help improve the antimalarial activity in this series of molecules due to the absence of information on the binding mode of tylophorine derivatives. While the CoMFA model could point to the significance of the steric and electrostatic fields, the CoMSIA model could explain in addition to the CoMFA fields the contribution of hydrophobic and hydrogen bonding (donor and acceptor) fields as well. This work further shows that a combined 2D and 3D-QSAR study can establish more pertinent and reliable QSAR model.

ACKNOWLEDGEMENT: SKP, LBG, and HNH acknowledge Gujarat State Biotechnology Mission (GSBTM), Gandhinagar for financial support. One of the authors, VMK, would like to thank Prof. Evans Coutinho, Bombay College of Pharmacy, Bombay, India, for computational help and useful discussion. Another author, PCJ, would like to thanks the Central University of Gujarat, Gandhinagar for providing computational resources and University Grants Commission (UGC), New Delhi, for providing start-up grants.

CONFLICT OF INTEREST: Nil

REFERENCES:

1. Lucet IS, Tobin A, Drewry D, Wilks AF and Doerig C: *Plasmodium kinases* as targets for new generation antimalarials. Future Medicinal Chemistry 2012; 4(18): 2295-10.
2. Burrows JN: Antimalarial drug discovery. Future Medicinal Chemistry 2012; 4(18): 2233-35.
3. Huthmacher C, Hoppe A, Bulik S and Holzhütter HG: Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. BMC Systems Biology 2010; 4: 120.
4. Anderson T: Mapping the Spread of Malaria Drug Resistance. PLoS Med 2009; 6(4): 1000054.
5. Saraswati S, Kanaujia PK, Kumar S, Kumar R and Alhaider AA: *Tylophorine*, a phenanthroindolizidine alkaloid isolated from *Tylophora indica* exerts antiangiogenic and antitumor activity by targeting vascular endothelial growth factor receptor 2–mediated angiogenesis. Molecular Cancer 2013; 12: 82.
6. Lill MA: Multi-dimensional QSAR in drug discovery. Drug Discovery Today 2007; 12(23-24): 1013-17.
7. Vyas VK, Ghate M and Katariya H: 2D and 3D-QSAR study on 4-anilinoquinozaline derivatives as potent apoptosis inducer and efficacious anticancer agent. Organic and Medicinal Chemistry Letters 2011; 1: 13.
8. Sharma MC and Kohli DV: Comprehensive structure-activity relationship analysis of isoxazolinyl and isoxazolidinyl substituted quinazolinone derivatives as angiotensin II receptor antagonists. Journal of Saudi Chemical Society 2012.
9. Allen FH, Davies JE, Galloy JJ, Johnson O, Kennard O, Macrae CF, Mitchell EM, Mitchell GF, Smith JM and Watson DG: J Chem Inf Comput Sci 1991; 31: 187.
10. Cramer RDIII, Patterson DE and Bunce JD: Comparative molecular field analysis (CoMFA). I. effect of shape on the binding of steroids to carrier proteins. J Am Chem Soc 1988; 110: 5959-67.
11. Cramer RDIII, Patterson DE and Bunce JD: Recent advances in comparative molecular field analysis (CoMFA). Prog Clin Biol Res 1989; 291: 161-65.
12. Klebe G, Abraham U and Mietzner T: Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 1994; (37): 4130-4146.
13. Klebe G: Comparative molecular similarity indices analysis: CoMSIA. Perspect Drug Dis Desi 1998; 12: 87-104.
14. Klebe G and Abraham U: Comparative molecular similarity index analysis (CoMSIA) to study hydrogen bonding properties and to score combinatorial libraries. J Comput Aided Mol Des 1999; 13: 1-10.
15. 4.3, Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India 2012.
16. Beroa J, Fre´de´richb M and Joe¨lle Quetin-Leclercqa: Antimalarial compounds isolated from plants used intraditional medicine. J of Pha and Ph 2009; 61: 1401-33.
17. Marvin Sketch 5.12.2, ChemAxon Kft., Budapest, Hungary 2013.
18. Bolton E, Wang Y, Thiessen PA and Bryant SH: PubChem: Integrated Platform of Small Molecules and Biological Activities. IN: Annual Reports in Computational Chemistry, American Chemical Society Washington DC 2008; 4: 217-41.
19. O'Boyle N, Banck M, James C, Morley C, Vandermeersch T and Hutchison G: Open Babel: An open chemical toolbox. Journal of Cheminformatics 2011; 3: 33.
20. Selvaraj C, Tripathi SK, Reddy KK and Singh SK: Tool development for Prediction of pIC_{50} values from the IC_{50} values-A pIC_{50} value calculator. Current Trends in Biotechnology and Pharmacy 2011; 5(2): 1104-09.
21. Patel SK, Kumar SP, Pandya HA, Jasrai YT and Patni MI: 2D-QSAR analysis of Dihydrofolate Reductase (DHFR) inhibitors with activity in *Toxoplasma gondii* and *Lactobacillus casei*. Journal of Advanced Bioinformatics Applications and Research 2011; 2(2): 161-66.
22. Sybyl, version 7.1; Tripos Associates Inc: 1699S Hanley Rd, St. Louis, MO 631444, USA 2005.
23. Prasad RK, Narsinghani T and Sharma R: QSAR analysis of novel N-alkyl substituted isatins derivatives as anticancer agents. Journal of Chemical and Pharmaceutical Research 2009; 1(1): 199-06.

How to cite this article:
Patel SK, George L, Khedkar VM, Lone MY, Jha PC, Jasrai YT, Pandya HA and Highland HN: Metaphorical quantitative structure-activity relationship (2D & 3D-QSAR) analysis of tylophorine derivatives as efficacious in antimalarial drug design. Int J Pharm Sci & Res 2014; 5(10): 4325-38. doi: 10.13040/IJPSR.0975-8232.5(10).4325-38.

All © 2013 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **ANDROID OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)