QSAR Analysis of Natural Lupeol Analogs as Antimalarial agents

David Mary Rajathei[1*]
[1]Department of Bioinformatics, Bishop Heber College, Tiruchirappalli, 620 017, India
*Correspondingauthor:maryrajathei.bi@bhc.edu.in

Abstract.Malaria is an infectious disease that affects millions of people caused by Plasmodium parasite. The growing Plasmodium resistance underscores existing antimalarial drugs and necessities novel antimalarial drugs with high efficiency. Lupeol is a naturally occurring pentacyclic triterpenoid is found in various medicinal plants that possesses antiprotozoal activity. The experimental testing of chemical compounds takes time and is very expensive. The computer-based quantitative structure-activity relationship (QSAR) can be used to accelerate the screening of compounds which correlate quantitatively the chemical compound properties to its biological activity against a specific target. The lupeol and its derivatives have been experimentally tested for inhibition of protozoa species. However, the physical and chemical properties of the compounds responsible for the activity are not reported yet which would be useful for screening. In this study, we developed a QSAR model for lupeol analogs using Least Absolute Shrinkage and Selection Operator (LASSO) method for selection of significant descriptors, and a Systematic search based-Multiple Linear Regression (MLR) for exploring antimalarial properties. The five descriptors model of ionization potential, electro and topological properties of the molecules is obtained for inhibition with significant statistics of R2= 0.9773. Then, the antimalarial activities of related compounds were predicted using the developed model.

Keywords: Malaria, *Plasmodium*, anti-malarial drugs, lupeol, QSAR, LASSO.

1 Introduction

Malaria is an infectious disease that is estimated to occur in millions of cases and thousands of deaths worldwide annually caused by Plasmodium parasite [1]. Among the six Plasmodium parasites, *Plasmodium falciparum* is a unicellular protozoan parasite transmitted to humans by the bites of female Anopheles mosquitoes and causes dangerous disease forms [2]. Drugs such as quinoline, primaquine, mefloquine, lumefantrine and amodiaquine have been widely used to inhibit the growth of *Plasmodium falciparum* [3]. However, the growing resistance of Plasmodium parasites necessitates more efficient antimalarial drugs. Also, people seek cures in medicinal plants because of fewer side effects.

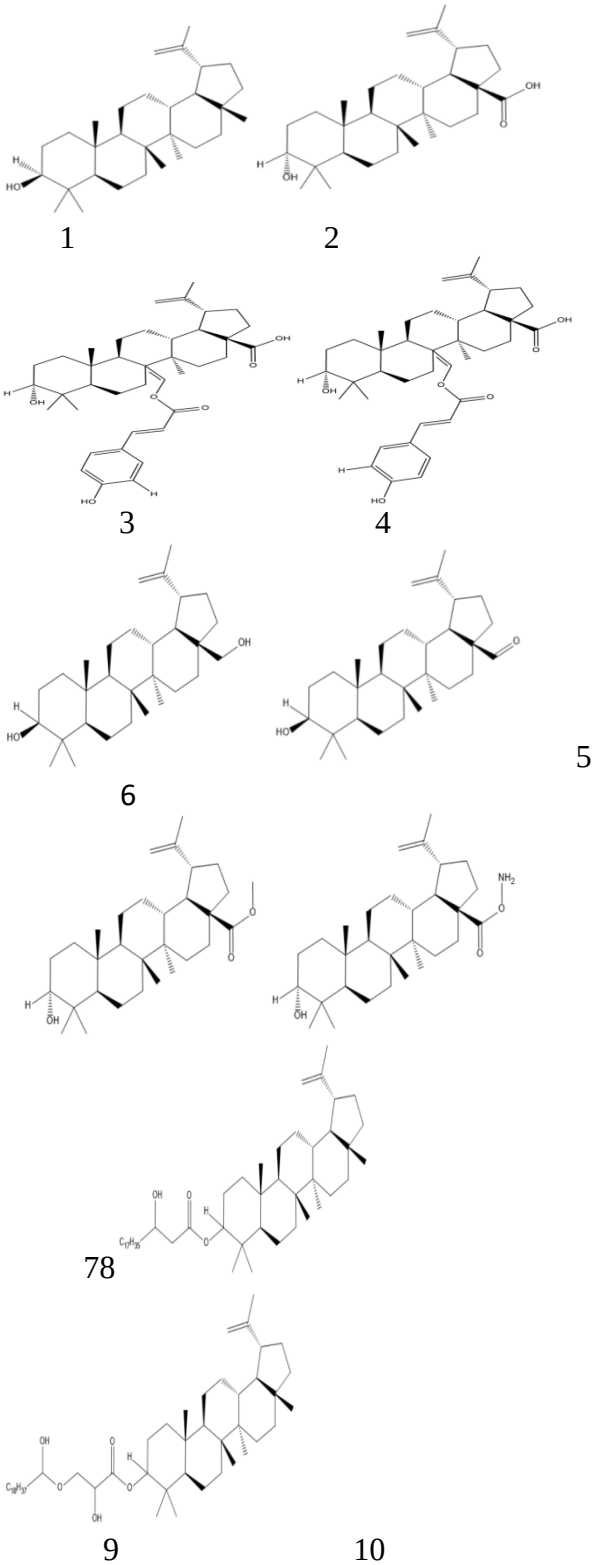
Lupeol, a naturally occurring pentacyclic triterpenoid, is found in various medicinal plants that possess immense biological and pharmacological activities against several diseases. The lupeol present in many plant species that exhibits antimalarial activity against several *P. falciparum* strains has been reported [4]. Further, it has been demonstrated that lupeol derivatives change inner layers of host cell membrane and prohibit the invasion and growth of parasites [5]. This suggests that lupeol can be considered in the discovery and development of antimalarial drugs.

The inhibitory activities of lupeol derivatives against *P. falciparum* strains are experimentally studied intensively. The *in vitro* biological testing for natural antimalarial agents revealed that lupeol inhibits the growth of *Plasmodium falciparum* [6]. Further, the *in vivo* efficacy of lupeol and its nanoemulsion preconcentrate (NPs) on antimalarial activity in *P. berghei* induced malaria in mice have been reported [7]. However, the properties of the compound that are responsible for the antimalarial activity are not reported yet, which would be useful for screening similar compounds. The quantitative structure-activity relationship (QSAR) establishes a mathematical relationship between the properties of the compounds and their biological activity against a specific target [8]. The present study is to develop QSAR model for lupeol derivatives based on 2D descriptors to explore the protozoa properties and to predict the antiprotozoal activity of related new compounds using the developed model. The result shows that the ionization potential, electro, and topological information of the atoms in the molecules are responsible for the protozoal inhibition activity, and the antiprotozoal activity of the related compounds in the PubChem database was predicted using the developed QSAR model.

2. Material and method

2.1 Collection of Data

A set of 18 derivatives of lupeol (Lup-20(29)-en-3β-ol) along with experimentally observed protozoa inhibition IC_{50} values was obtained from the literature [9,10]. The inhibition inhibition IC_{50} values were normalized into logarithmic scales of pIC_{50} [$-\log IC_{50}$]. The 2D structures of the compounds drawn using ChemDraw 8.0 developed by CambridgeSoft Corporation are shown in Fig. 1. The SMILES (Simplified Molecular Input Line Entry System) formats of the compounds along with their pIC_{50} are shown in Table 1.



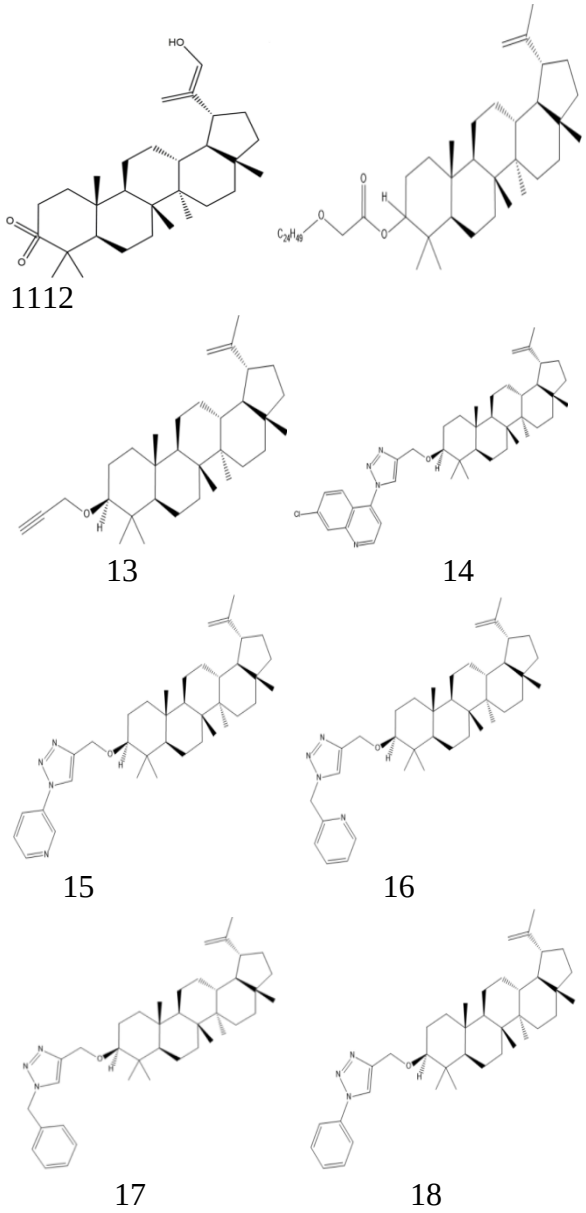


Fig.1. The 18 chemical structural derivatives of antiprotozoal compounds are 1-12 (Lupeol, betulinic acid, messagenic acid A, messagenic acid B, betulin, betulinaldehyde, methylbetulinate, betunic acid amide,3-O-(3'-hydroxyeicosanoyl)lupeol,3-O-[(2'-(tetracosyloxy)acetyl] lupeol,30-hyroxylupenone, 3-O-[(1''-hydroxyoctadecyloxy)-2'-hydroxypropanoyl] lupeol), 13 (propargyl ether of lupeol) ,14-18 (1,2,3-triazole lupeol derivatives)

2.2 Finding Molecular Descriptor

Molecular descriptors are the chemical structure in numerical values that refer to shape, complexity, branching, cyclicity, stereo electronics character, etc. The PaDEL , an open resource software was used to calculate the empirical, constitutional and topological based descriptors of the molecules [11]. The SMILES formats of the compounds were ported in to PaDEL for computing 1544 2D descriptors, including Constitutional (120), Topology (266), Basak (42), Connectivity (56), Kappa (3), Bcut (6), Estate (568), Autocorrelation of morgan and geary (346) Molecular property(15), Burden (96), Amino acid count (13), Quantum chemical (6) and other descriptors (17).

2.3 Selection of Descriptors

A set of 1055 descriptors was considered after eliminating descriptors with null and constant values. Tibshirani's proposed LASSO (Least Absolute Shrinkage and Selection Operator) method [12] was used to select significant descriptors (non-zero coefficient) by penalizing the irrelevant descriptor coefficient to zero. The selection of significant descriptors by the LASSO method was done by using the module GLMNET in R package. The input data of X (descriptors set) and Y (inhibition data) of the compounds was uploaded in the GLMNET for identifying descriptors related to protozoa inhibition.

2.4 Developing QSAR model

The BuildQSAR software developed by Anderson Coser Gaudio [13] was used to develop QSAR models with different combinations of significant descriptors of the compounds. A systematic search was performed to select descriptors based on user-given correlation criteria with respect to activity (pIC_{50}). The QSAR model using Multiple Linear Regression (MLR) method was built and the regression is of the follwing form:

$$Y = a*X1 + b*X2+.... + c$$

Where, Y is predicted value and X1, X2, X3 . . ., Xk are the variables. X1 is the first variable and a is the associated regression coefficient with it,

Table1. A set of 18 compound's SMILES format, experimentally observed pIC_{50} values and predicted pIC_{50} values using developed model in Eq. 1 against antiprotozoal activity.

SMILES Format	Experimental pIC_{50}	Predicted pIC_{50}
[H][C@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C	5.3	5.2
[H][C@@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C(O)=O)C(C)=C)C=CC1(C)C	4.7	5.0
[H]C1=CC(\C=C\C(=O)OC=[C]23CC[C@H]4C(C)(C)[C@]([H])(O)CC[C@]4(C)[C@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C(O)=O)C(C)=C)=CC=C1O	5.8	5.5
[H]C1=C(O)C=CC(\C=C\C(=O)OC=[C]23CC[C@H]4C(C)(C)[C@]([H])(O)CC[C@]4(C)[C@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C(O)=O)C(C)=C)=C1	5.4	5.5
[H][C@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(CO)CC[C@@]32C)C(C)=C)C1(C)C	5.4	5.4

[H][C@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C=O)C(C)=C)C1(C)C	5.2	5.4
[H][C@@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C(=O)OC)C(C)=C)C1(C)C	5.5	5.1
[H][C@@]1(O)CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@@]4(CC[C@@]32C)C(=O)ON)C(C)=C)C1(C)C	5.2	5.1
[H]C1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OC(=O)CC(O)CCCCCCCCCCCCCC	3.7	3.8
[H]C1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OC(=O)C(O)OCC(O)CCCCCCCCCCCCCC	4.2	4.5
C[C@]12CC[C@H]([C@@H]1[C@H]1CC[C@@H]3[C@@]4(C)CC[C](=O)(=O)C(C)(C)[C@@H]4CC[C@@]3(C)[C@]1(C)CC2)[C](=C)=CO	3.2	3.1
[H]C1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OC(=O)C(O)COC(O)CCCCCCCCCCCCCCCC	3.4	3.3
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC#C	4.2	4.3
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC1=CN(N=N1)C1=CC=NC2=C1C=CC(Cl)=C2	3.7	3.8
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC1=CN(N=N1)C1=CN=CC=C1	3.7	3.7
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC1=CN(CC2=CC=CC=N2)N=N1	3.7	3.7
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC1=CN(N=N1)C1=CC=CC=C1	3.7	3.7
[H][C@@]1(CC[C@@]2(C)[C@@H](CC[C@]3(C)[C@@H]2CC[C@@H]2[C@H]4[C@@H](CC[C@]4(C)CC[C@@]32C)C(C)=C)C1(C)C)OCC1=CN(CC2=CC=CC=C2)N=N1	3.7	3.6

X2 is the second variable and b is the associated regression coefficient with it and so on.

2.5 Validation of QSAR model

The QSAR model with high value of correlation coefficient (R), highe Fischer's value (F-Test), low standard error of estimate (s) and statistical significance (p) was selected. The internal fitness of the model was verified with a high cross-validated square of correlation coefficient (Q^2_{cv}). In that, the one compound is deleted from the data set and the model is created using remaining compounds. Then, the deleted compound is predicted using the developed model. The procedure is repeated until all the compounds are deleted once. The Q^2_{cv} value is calculated using the following formula:

$$Q^2_{cv} = 1- SPRESS/SDEP$$

Here SPRESS is the sum of the squared differences between the observed activity and the predicted activity. SDEP is standard deviation of error of prediction.

The performance of the obtained descriptors is also analysed with the receiver operating characteristic (ROC) curve by plotting sensitivity vs specificity. For that, the dataset was randomly split into 70% of the training set and 30% of the test set using the split function of R. Then, a logistic regression model for Y (observed) inhibition against the X (obtained descriptors) of the training data is developed using the lm function of R. Then, the accuracy of prediction in terms of sensitivity (True Positive Rate) and specificity (False Positive Rate) of the model in the test set was assessed using the prediction function in a ROCR package of R.

2.6 Finding similar chemical structures of the compounds used in QSAR model for prediction

The similar structures of the compounds used in model construction (Table 1) were found using a PubChem database search. The PubChem (https://pubchemncbi.nlm.nih.gov) is a database for chemical compounds and similar compounds to the compounds in the data set were obtained based on similarity search using the Tanimoto equation.

$$\text{Tanimoto} = \frac{AB}{A+B-AB}$$

Where, A and B are the binary fingerprints of respective compounds and AB is the common count bits. Compounds that are having Tanimoto scores >= 0.9 with the search compound were considered as related compounds for prediction.

3 Result And Discussion

The 11 descriptors of ATSC1v, ATSC4i, ATSC7i, AATSC5e, MATS2s, MATS4s, GATS8e, GATS7p, minHCsatu, MDEC-12, and MDEC-14 were identified using the LASSO method as significant descriptors for *Plasmodium falciparum* inhibition (Table 2). The QSAR model development for protozoa inhibition using the significant descriptors was discussed in the following sections.

Table. 2 The details of 11 significant descriptors obtained for antimalarial activity

S.No	Descriptor	Description	Description class
X1	ATSC1v	Centered Moreau-Broto autocorrelation of lag 1 weighted by vdw volume	Autocorrelation descriptors
X2	ATSC4i	Centered moreau-broto autocorrelation of lag 4 weighted by ionization potential	
X3	ATSC7i	Centered moreau-broto autocorrelation of lag 7 weighted by ionization potential	
X4	AATSC5e	Averaged Centered moreau-broto autocorrelation of lag 6 weighted by ionization potential	
X5	MATS2s	Moran coefficient of lag 2 weighted by intrinsic state	
X6	MATS4s	Moran coefficient of lag 4 weighted by intrinsic state	
X7	GATS8e	Geary autocorrelation–lag 6/weighted by Sanderson electro negativity	
X8	GATS7p	Geary autocorrelation of lag 7 weighted by polarizability	
X9	minHCsatu	Minimum atom-type H E-State: H on C sp3 bonded to unsaturated C	Electro topological state atom type indices
X10	MDEC-12	molecular distance edge between primary C and secondary C	Molecular Distance Edge
X11	MDEC-14	molecular distance edge between primary C and quaternary C	Molecular Distance Edge

3.1 QSAR Model for antiprotozoal activity

The 11 descriptors of the compounds (X-axis) and inhibition pIC_{50} values (Y-axis) are given as input into BuildQSAR. We considered one to six descriptors for QSAR model development because the correlation matrix among descriptors shows the descriptor ATSC1v is more correlated with the descriptors of ATSC4i, AATSC5e, and MDEC-12 of >0.7, and ATSC4i is correlated with MATS4s, MDEC-12, and MDEC-14 of >0.7. A systematic search for different combinations of descriptors was performed based on the criteria of the correlation coefficient R value (>0.6). The QSAR model that contained five variables of ATSC4i ATSC7i MATS2s GATS8e and minHCsatu produced the optimum model with squared correlation coefficient (R^2) of 0.9595, adjusted squared correlation coefficient (R^2adj) of 0.9427, and leave-one-out (LOO) cross-validation coefficient (Q^2_{cv}) value of 0.655.

$$Y1 == - 0.0509 (\pm 0.0154) * ATSC4i + 0.0258 (\pm 0.0119) * ATSC7i - 2.9039 (\pm 2.0117) * MATS2s - 0.4341 (\pm 0.9817) * GATS8e + 5.0458 (\pm 1.3824) * minHCsatu + 1.5147 (\pm 1.9679) \quad \textbf{(Eq. 1)}$$

(n =18 ; R^2 = 0.9595; R^2adj =0.9427; F =56.925 ; s =0.208 ; p < 0.0001 ;Q^2_{cv} =0.655 ; SPress = 0.607 ; SDEP = 0.510)

Where n is the number of compounds used for regression. ATSC4i is centered Moreau–Broto autocorrelation of lag 4 weighted by ionization potential, ATSC7i is centered Moreau–Broto autocorrelation of lag 7 weighted by ionization potential, MATS2s is Moran autocorrelation descriptor of lag 2 / weighted by I-state and GATS8e is Geary autocorrelation of lag 8/ weighted by Sanderson electronegativity and minHCsatu is electrotopological state atom type descriptor of minimum atom-type hydrogen sp3 bonded to unsaturated carbon.

ATSC4i is Moreau–Broto autocorrelation of the first ionization potential of the atoms of length 4. The descriptor has a negative value and suggests the ionization potential of atoms at the distance of 4 in molecule structure is inversely related to the activity. ATSC7i is Moreau–Broto autocorrelation of the first ionization potential of atoms at the topological distance of 7. This descriptor has a positive correlation value that indicates the ionization potential of the atoms connected within a topological distance of 7 is related to activity. The Moran autocorrelation descriptor MATS2s describes the intrinsic electron topological state between atoms separated by 2. This descriptor indicates a negative sign which suggests that the local electronic environment of atoms that are two bonds apart in the molecule's structure is inversely related to activity. The Geary autocorrelation descriptor GATS8e describes the electronegativity of atoms in paths of length 8 within the molecule's connectivity. The negative sign of the descriptor indicates that it is inversely related to information about electronegativity of atoms in the path length 8. The electro topological atom type descriptor minHCsatu describes the minimum atom-type hydrogen atoms connected to unsaturated carbon. This descriptor displays a positive sign which indicates hydrogen atoms directly attached to carbon atoms correlated to activity.

The QSAR model indicated that ionization potential of atoms in lag 4 and lag 7 (ATSC4i, ATSC7i), intrinsic electronic state of an atom (MATS2s), electronegativity (GATS8e), and atom type of H connected to C (minHCsatu) play a role in inhibition against P. falciparum strain. This implies that inhibition of lupeol is negatively influenced by ATSC4i, MATS2s and GATS8e and positively influenced by ATSC7i and minHCsatu. Through the analysis, it was observed that ionization potential, electron and topological environment of atoms in the molecules play a role in inhibition.

3.2 Effectiveness of prediction of antiprotozoal activity using the obtained descriptor model

The predicted pIC_{50} values for all compounds using the obtained five descriptors model (Eqn 1) are shown in Table 1. The scatter plot of predicted and experimental

pIC50 values shows a correlation of R^2 = 0.959 (Standard Error of 0.18024) with a fitted regression line of y = 0.959x + 0.178 (Fig. 2) which shows predicted pIC_{50} values are in good alignment with experimental values. While comparing the predicted inhibition values using the model with the experimental values, the residual variances in the range of 0.003 – 0.337 were observed. This suggests that the obtained descriptor model also calculates the protozoa inhibition of the compounds. Further, the predictive ability of the model was validated internally using the cross-validated Q^2_{CV} value of 0.655 by the LOO method.

Further, the data was divided into training (70%) and test (30%) of compounds for external validation. The linear regression model for the training set was obtained with regression of R^2_{train} = 0.977 and the test set was used to evaluate the predictive ability of the obtained model. We observed that most of the observed and predicted activity values of the test set are within the < ±1 range of differences with regression R^2_{test}=0.8774712.

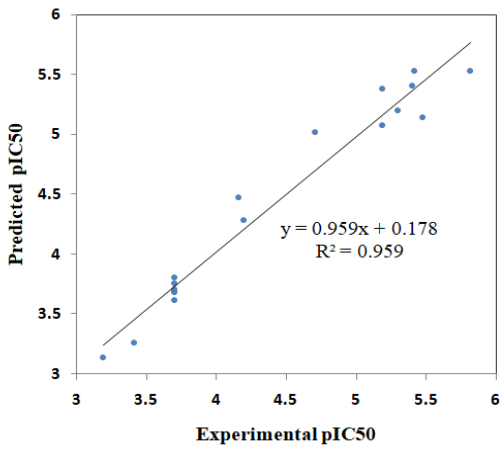


Fig. 2 . The scatter plot of pIC_{50} experimental values vs. pIC_{50} predicted values by five descriptors model in Equation 1 for protozoal inhibition

The obtained ROC curve and area under ROC (AUROC) curve of 0.80 for the test set are shown in Fig. 3.

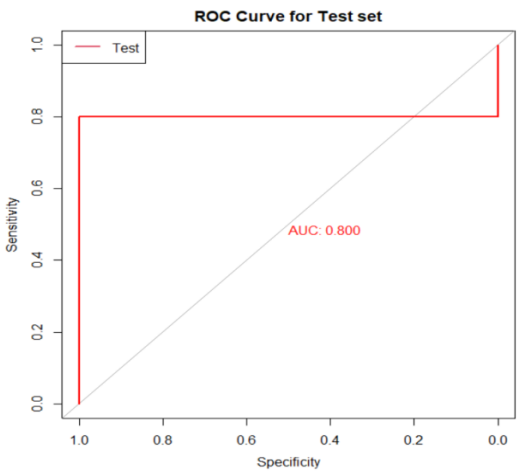


Fig. 3. The ROC/AUROC curve for test data using the developed model with five descriptors

Further, a scatter plot of test data superimposed over the training data shows the overlapping of the test set with that of the training set (Fig.4). This shows pIC50 values of the test set could be validated by using the training set.

3.3 Predicting the antiprotozoal activity of similar Compounds

The 785 PubChem compounds were obtained by similarity search with the compounds in the set (Table 1). Among these, the set of 21 compounds satisfying the drug like properties are chosen for prediction. Using calculated physical properties values (ATSC4i ATSC7i MATS2s, GATS8e and minHCsatu) of these compounds in Eqn 1, the inhibition activities of the compounds were predicted (Table 3). The predicted inhibition values of these compounds are in the range of 2.8-6.5, which are within the observed threshold (Table 1). The result shows that the developed model can predict the antimalarial activity of the similar compounds and suggests that initial prediction of compound activity could be performed using *in silico* computational methods.

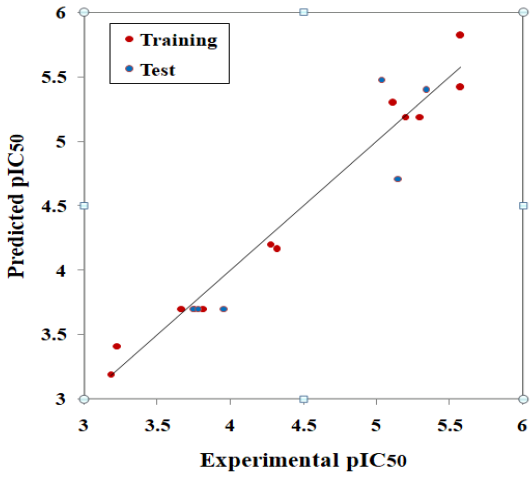


Fig. 4. A scatter plot of test data superimposed over that of the training data to analyse the obtained regression model

Table 3. The PubChem Compound ID, five physical properties of ATSC4i,ATSC7i, MATS2s,GATS8e and minHCsatu and Predicted pIC_{50}using developed modelfor 21 new similar compounds

S. No	PubChem Compound ID	ATSC4i	ATSC7i	MATS2s	GATS8e	minHCsatu	Predicted pIC50
1	44317358	-64.0	-36.8	0.1	1.2	0.5	5.4
2	101624720	-69.2	-18.3	0.1	0.8	0.4	6.0
3	71591584	-64.9	-22.4	0.0	0.9	0.4	5.8
4	44317312	-65.8	-33.4	0.1	1.1	0.5	5.5
5	155809468	-75.2	-19.6	0.1	1.2	0.4	6.2

6	162958297	-53.5	-27.2	0.1	1.1	0.0	2.8
7	21672691	-69.3	-12.9	0.0	1.0	0.4	6.4
8	122394979	-68.8	-15.7	0.1	1.1	0.5	6.5
9	100913693	-68.4	-23.1	0.1	1.3	0.6	6.5
10	16112782	-50.0	-33.8	0.1	1.1	0.4	4.4
11	137634579	-59.3	-12.1	0.0	1.1	0.4	5.8
12	101914881	-66.6	-26.0	0.1	1.1	0.5	5.9
13	137643624	-74.9	-24.2	0.1	1.4	0.4	6.0
14	122394980	-69.6	-16.8	0.1	1.0	0.4	5.8
`15	165360383	-40.7	-23.3	0.1	1.2	0.4	4.3
16	90941119	-46.6	-30.5	0.0	1.0	0.4	4.5
17	73352305	-73.4	-17.4	0.1	1.2	0.4	6.1
18	9982971	-63.8	-25.2	0.1	1.0	0.4	5.3
19	189260	-60.4	-32.9	0.0	1.2	0.3	5.1
20	132577482	-75.8	-20.4	0.1	1.2	0.4	5.9
21	53362469	-57.9	-17.3	0.0	1.0	0.3	5.3

4 Conclusions

In this study, the QSAR model for lupeol analogs was obtained to explore the physicochemical properties involved in the inhibition of *Plasmodium falciparum*. It was observed that the ionization potential of atoms at the distance of 4 & 7 (ATSC4i, ATSC7i), intrinsic electron state between two atoms (MATS2s), electronegativity of atoms in the paths of length 8 (GATS8e), and interaction of hydrogen atoms bonded to unsaturated carbons (minHCsatu) of the molecules play a critical role in protozoa inhibition. The descriptors ATSC4i, MATS2s and GATS8e were observed to make negative contributions to activity and on the other hand, the descriptors ATSC7i and minHCsatu render positive effect to activity. This information may help in exploring more potential analogues for the inhibition of *Plasmodium falciparum*. The statistical estimate of goodness of fit and prediction (R^2 and Q^2_{cv}) suggested that the acceptable reliability and predictive ability of developed model. Further, the related compound's antiprotozoal inhibition activities were predicted using the obtained QSAR model. The present study suggests that the developed QSAR model will contribute to the screening of similar properties containing compounds as well as the design, which would be useful for the researchers in the development of antimalarial drugs with improved activity.

References

1. A. F. Cowman, J. Healer, D. Marapana , K. Marsh, Malaria: Biology and Disease. Cell. **167**(3), 610-624(2016). https://doi.org/10.1016/j.cell.2016.07.055.
2. S. Sato, Plasmodium-a brief introduction to the parasites causing human malaria and their basic biology. J Physiol Anthropol. **40**(1), 1 (2021). https://doi.org/10.1186/s40101-020-00251-9.
3. E. Sevene, R. González, C. Menéndez, Current knowledge and challenges of antimalarial drugs for treatment and prevention in pregnancy. Expert Opin Pharmacother. **11**(8), 1277-1293 (2010). https://doi.org /10.1517/14656561003733599.
4. T. M. Alves, T. J. Nagem, L. H. de Carvalho , A. U. Krettli, C. L. Zani, Antiplasmodial triterpene from Vernonia brasiliana. Planta Med. **63**(6), 554-555 (1997). https://doi: 10.1055/s-2006-957764.
5. H. L. Ziegler, H. Franzyk, M. Sairafianpour, M. Tabatabai, M. D. Tehrani , K. Bagherzadeh , H. Hägerstrand, D. Staerk, J. W. Jaroszewski, Erythrocyte membrane modifying agents and the inhibition of Plasmodium falciparum growth: structure-activity relationships for betulinic acid analogues. Bioorg Med Chem. **12**(1), 119-127 (2004). doi: 10.1016/j.bmc.2003.10.010.
6. A. Parvez, M. A. Rahman, M. M. Rahman, A. I. Shimki, S. Ahmmed, F. A. Supti, M. H. Hasan, M. S. A. , Bristi, S. A. Ansari, M. T. Islam, Broad-Spectrum Therapeutic Potentials of the Multifaceted Triterpene Lupeol and Its Derivatives. Chem Biodivers. e202402286 (2025). https://doi.org/10.1002/cbdv.202402286.
7. A. C. G. Jyotshna, D. U. Bawankule, A. K. Verma, and K. Shanker, Nanoemulsion Preconcentrate of a Pentacyclic Triterpene for Improved Oral Efficacy: Formulation Design and In-Vivo Antimalarial Activity, Journal of Drug Delivery Science and Technology, **57**, 101734 (2020), https://doi.org/10.1016/J.JDDST.2020.101734.
8. J. Li, T. Zhao, Q. Yang, S. Du, L. Xu, A review of quantitative structure-activity relationship: the development and current status of data sets, molecular descriptors, and mathematical models. Chemometrics and Intelligent Laboratory Systems. **256**, 105278 (2025). https://doi.org/10.1016/j.chemolab.2024.105278.
9. M. B. C. Gallo, M. J . Sarachine, Biological activities of lupeol. In: International Journal of Biomedical and Pharmaceutical Sciences. 46–66 (2009).
10. T. F. Borgati, G. R. Pereira, G. Brand, J. O. Santos, D. A. Fernandes, R. C. Paula, M. F. A. Nascimento, L. F. Soares, J. C. Lopes, J. D. S. Filho, A. Oliveira, Synthesis of triazol derivatives of lupeol with potential antimalarial activity. Orbital: Electron. J. Chem. **4**, 21–22 (2012). https://doi.org/10.21577/0103-5053.20170013.
11. C. W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. **32**(7), 1466-1474 (2011). https://doi.org/10.1002/jcc.21707.
12. R. Tibshirani, Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological). **58**(1), 267-288(1996). https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
13. D. B. De Oliveira, A. C. Gaudio, BuildQSAR: a new computer program for QSAR analysis. Quant. Struct. Act. Relat. **19**, 599–601 (2000). https://doi.org/10.1002/1521-3838(200012).