Current Research in Pharmaceutical Sciences
Available online at www.crpsonline.com

ISSN 2250-2688

Received: 13/02/2013
Revised: 24/02/2013
Accepted: 29/03/2013

QSAR Studies of Thiourea, Thiazolidinedione and Thioparabanic Acid Derivatives of 4-Aminoquinoline as Antimalarial Agents

Nitendra K Sahu, Kamlendra S Bhadoriya, Mukesh C Sharma and D V Kohli

Nitendra K Sahu
School of Pharmacy, ITM University, Gwalior (MP) 474001, India

Kamlendra S Bhadoriya
R. C. Patel Institute of Pharmaceutical Education and Research, Shirpur, (MS), 425405 India

Mukesh C Sharma, D V Kohli
Drug Research Laboratory, Department of Pharmaceutical Sciences, Dr. H. S. Gour University, Sagar (MP),470003 India

Correspondence
Nitendra K Sahu
School of Pharmacy, ITM University, Gwalior (MP) 474001, India
E mail: nitendrasemail@yahoo.com

ABSTRACT

The QSAR studies were performed on a series of thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline to find out the structural requirements of their antimalarial activities. The 2D-QSAR studies were performed using three methods: MLR, giving r^2 = 0.7808, q^2 = 0.6492 and pred_r^2 = 0.9044; the PLS, with r^2 = 0.7785, q^2 = 0.6420 and pred_r^2 = 0.7181; and PCR, giving r^2 = 0.7393, q^2 = 0.6222 and pred_r^2 = 0.9135. The QSAR models showed that the LUMOEnergy, SdsCHE-index and Quadrupole1 played an important role in determining biological activities. The 3D-QSAR studies were performed using the stepwise variable selection k-nearest neighbor molecular field analysis (kNN-MFA) approach; q^2 of 0.7107and pred_r^2 of 0.8521 were obtained. The results of the present study may be useful for the designing of more potent analogues as antimalarial agents.

Keywords: QSAR, 4-aminoquinoline, antimalarial agents, PLSR, kNN-MFA

1. INTRODUCTION

Malaria is estimated to kill more than 1 million people annually and possibly as many as 3 million, with most of the deaths among children under age six living in sub-Saharan Africa. According to the WHO, between 300 million and 500 million clinical cases of malaria occur every year [1]. Its control is globally a high priority task. Although continued attempts to develop a vaccine for malaria are ongoing, drugs continue to be the only treatment option[2]. Although effective antimalarial agents have been known for a long time, the alarming spread of drug resistant strains of *Plasmodium falciparum*, which is the most lethal parasite species, underscores the urgency and continuous need for the discovery of new therapeutics. 7-Chloro-4-aminoquinoline derivatives including chloroquine (CQ), sontoquine, and amodiaquine are among the most potent antimalarial drugs reported to date[3,4], and new agents with improved activity against CQ resistant (CQR) strains have been introduced via synthetic modifications of the CQ side chain[5,6]. The search for novel drug candidates against specific parasitic targets is an important goal for antimalarial drug discovery[7, 8]. Among old and new drug targets of malaria, host heme molecule remains one of the most attractive target and 7-chloroquinoline compounds are very selective towards heme binding[9]. So, rather than identifying the new molecules for efficacy, 7-chloroquinolines having many advantages and efficiency are now in priority for antimalarial chemotherapy. Based on this observation, the antimalarial activities of some thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline have generated great interest[10].

Understanding the effect of structural features on the activity would help the researchers to design new molecules that may come up as potential new lead molecules. Quantitative structure activity relationship (QSAR) represents an attempt to correlate structural or property descriptors of compounds quantitatively with biological activities. This quantitative technology can be utilized to improve the structure of the inhibitor molecule and to interpret the improved structure in terms of favorable biological interactions[11, 12]. QSAR models lead to assessment of the specific effects of various types of substituents reducing trial experiments and investments. QSAR studies have provided valuable insight in the design and development of antimalarial agents[13-16].

The present study aimed to elucidate the structural features of some thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline required for antimalarial activity and to obtain predictive two- and three-dimensional (2D and 3D) QSAR models, which may guide the rational synthesis of more active antimalarial agents. This is accomplished by combining one of the stochastic search methods, such as multiple regression (MLR), partial least squares regression (PLSR), or principle component regression (PCR) analysis[17-19]. The results obtained may contribute to further designing of novel antimalarial agents.

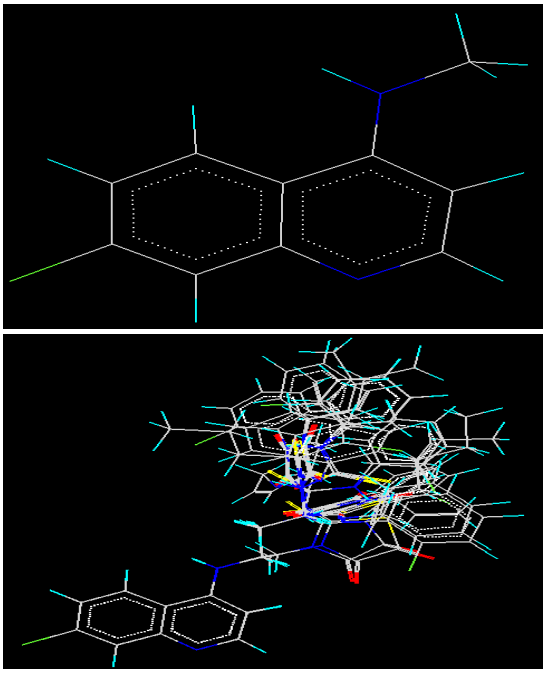


Figure 1. Superposition of compounds in the training and test sets using the template-based alignment method (a) 7-chloro-N-ethylquinolin-4-amine as common template (b) Stereo-view of aligned molecules in training set and test set.

2. MATERIALS AND METHODS

2.1 Data Set for analysis

A data set of 24 molecules was taken from the literature (Table 1) in which authors had reported anti-malarial activity of thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline against CQ sensitive strain 3D7 of P. falciparum [10]. All the values of biological data shown in IC50 (ng/mL) were converted into -log IC50 (M) values for the convenience of computational work.

2.2 2D QSAR Study

The molecular structure of all the 24 molecules were built using the 2D draw application of VLife MDS 3.5 software 20 and then the structures were converted to 3D structures for further analysis. All the compounds were batch optimized for the minimization of energies and geometry optimization using Merck molecular force field (MMFF) followed by considering distance-dependent dielectric constant of 1.0, convergence criterion or root-mean-square (RMS) gradient at 0.01 kcal/mol Å and the iteration limit to 10,000 21.

Most stable structure for each compound was generated and used for the calculation of the various physico-chemical descriptors (Individual, Chi, ChiV, Path count, ChiChain, ChiVChain, Chainpathcount, Cluster, Pathcluster, Kapa, Element Count, Estate number, Estate contribution, Semi-impirical and Polar surface area) (Table 2).

The invariable descriptors (descriptors that are constant for all the molecules) were removed, as they do not contribute to the QSAR, which resulted in total 157 descriptors was considered as independent variables in the present study. The values of the descriptors that are significant in the QSAR models are shown in Table 3.

The total set of compounds was divided into a training set (19 compounds) for generating 2D QSAR models and a test set (5 compounds) for validating the quality of the models[22]. Selection of the training set and test set molecules was done on the basis of structural diversity and a wide range of activity such that the test-set molecules represent a range of biological activity similar to that of the training set; thus, the test set is truly representative of the training set. Five compounds, namely, 5, 9, 14, 16, and 17 were used as test set while the remaining molecules as the training set. The unicolumn statistics of the training and test sets are reported in Table 4.

2.3 3D QSAR Study

Molecular alignment is a crucial step in 3D-QSAR study to obtain meaningful results. In the present study, the molecules of

the dataset are aligned by template based method18 in VLife MDS 3.5 software, where a template structure is defined and used as a basis for alignment of a set of molecules, and a reference molecule is chosen on which the other molecules of the dataset get aligned considering the chosen template. The template structure, i.e. 7-chloro-N-ethylquinolin-4-amine is used for alignment by considering the common elements of the series as shown in Figure 1(a). The reference molecule is chosen in such a way that it is the most active among the series of molecules considered. The compound 3 has very high antimalarial activity made it a valid lead molecule. Hence, compound 3 has been chosen as a reference molecule. The superimposition of all molecules based on minimizing root mean square deviation (RMSD) is shown in Figure 1(b).

For calculation of field descriptor values, using Tripos force field23, both electrostatic and steric field types, with cutoffs of 10.0 and 30.0 kcal/mol, respectively, were selected and charge type was selected as by Gasteiger and Marsili24. The dielectric constant was set to 1.0 considering the distance dependent dielectric function. Probe setting was carbon atom with charge 1.0. This resulted in calculation of 3366 field descriptors (1702 for electrostatic and 1664 for steric) for all the compounds in separate columns after removing descriptors having zero values or same values.

The sphere exclusion method22 was adopted for division of training and test data set comprising of 19 and 5 molecules, respectively, with dissimilarity value of 7.3. The test set consisted of seven compounds, namely, 2, 7, 9, 19 and 23. The remaining molecules were in the training set. The unicolumn statistics were the training and test sets, as reported in Table 4.

2.4 Feature selection and model development

Feature selection is a key step in QSAR analysis. An integral aspect of any model-building exercise is the selection of an appropriate set of features with low complexity and good predictive accuracy. This process forms the basis of a technique known as feature selection or variable selection[25].

In SW algorithm, the search procedure begins with developing a trial model step by step with a single independent variable and to each step; independent variables are added one at a time, examining the fit of the model. Thus, the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case F=4 for inclusion; F=3.99 for exclusion for the forward-backward selection method). The method continues until there is no more significant variable remaining outside the model[20].

In the selected 2D QSAR equations, the cross-correlation limit was set at 0.5, the number of variables at 10, and the term selection criteria at r2. An F value was specified to evaluate the significance of a variable. The variance cutoff was set at 0, with autoscaling in which the number of random iterations was set at 10.

MLR is the standard method for multivariate data analysis26. It estimates the values of the regression coefficients by applying least squares curve fitting method. For getting reliable results, dataset having typically 5 times as many data points (molecules) as independent variables (descriptors) is required.

PLSR is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables. It gives a reduced solution, which is statistically more robust than MLR. The linear PLSR model finds "new variables" (latent variables or X scores) which are linear combinations of the original variables. To avoid over-fitting, a strict test for the significance of each consecutive PLSR component is necessary and then stopping when the components are non-significant. Cross-validation is a practical and reliable method for testing this significance[27]. PLSR is normally used in combination with cross-validation to obtain the optimum number of components.

PCR analysis selects a new set of axes for the data. These are selected in decreasing order of variance within the data. They are also perpendicular to each other. The problem noted with MLR was that correlated variables cause instability. So, how about calculating principal components, throwing away the ones which only appear to contribute noise (or constants), and using MLR on these? This process gives the modeling method known as Principal Components Regression. Rather than forming a single model, as with MLR, a model can be formed using 1, 2, ... components and a decision can be made as to how many components are optimal.

k Nearest Neighbor Molecular Field Analysis (kNN MFA), methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN MFA method[28] was implemented simply as follows: (i) The distances between an unknown object (u) and all other objects in the training set were calculated. (ii) The k objects were selected from the training set most similar to object u, according to the calculated distances; and (iii) The object u was classified with the group to which the majority of the k objects belong. An optimal k value is selected by optimization through the classification of a test set of samples or by Leave-One Out (LOO) cross-validation.

Since there was a large pool of descriptors available to build models, stepwise (SW) variable selection methods were used along with kNN to find optimal subset of descriptors. kNN-MFA models were developed using the SW forward–backward method with the cross-correlation limit set to 0.5 and the term selection criterion as q2. F-test 'in' was set to 4.0, and F-test 'out' to 3.99. As some additional parameters, variance cutoff was set at 2 kcal/mol Å, and scaling to autoscaling; additionally, kNN parameter setting was done within the range of 2–5 and the prediction method was selected as the distance-based weighted average[20].

2.5 Model evaluation and validation

This is done to test the internal stability and predictive ability of the QSAR models. Internal validation was carried out using 'leave-one-out' (q2, LOO) method. The cross-validated coefficient, q2, was calculated using the following equation:

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \quad (1)$$

where yi, and ŷi are the actual and predicted activity of the ith molecule in the training set, respectively, and ymean is the average activity of all molecules in the training set.

For external validation, activity of each molecule in the test set was predicted using the model generated from the training set. The predictive ability of the selected model was also confirmed by pred_r2.

$$\text{pred}\_r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \quad (2)$$

where yi, and ŷi are the actual and predicted activity of the ith molecule in the test set, respectively, and ymean is the average activity of all molecules in the training set.

2.6 Evaluation of the quantitative models

The developed QSAR models were evaluated using the following statistical measures: r2 (the squared correlation coefficient), F test (Fischer's value) for statistical significance, q2 (cross-validated correlation coefficient); pred_r2, r2 for external test set. The regression coefficient r2 is a relative measure of fit by the regression equation. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: r2 > 0.6, q2 > 0.6 and pred_r2 > 0.5[29, 30]. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. The low standard error of Pred_r2se, q2_se and r2_se shows absolute quality of fitness of the model. The r2, q2 and pred_r2 values were used as deciding factors in selecting the optimal models.

3. RESULTS AND DISCUSSION

All QSAR studies were performed in V-Life MDS software Version 3.520. A series of 24 thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline tested for their anti-malarial activity was selected for QSAR Studies (Table 1). Selection of molecules in the training set and test is a key and important feature of any QSAR model. Therefore the care was taken in such a way that biological activities of all compounds in test set lie within the maximum and minimum value range of biological activities of training set of compounds. A Uni-Column statistics for training set and test set were generated to check correctness of selection criteria for trainings and test set molecules (Table 4).

The maximum and minimum value in training and set were compared in a way that:

- The maximum value of pIC50 of test set should be less than or equal to maximum value of pIC50 of training set.
- The minimum value of pIC50 of test set should be higher than or equal to minimum value of pIC50 of training set.

This observation showed that test set was interpolative and derived within the minimum–maximum range of training set. The mean and standard deviation of pIC50 values of sets of training and test provide insights to relative difference of mean and point density distribution of two sets. The mean of the test sets were higher than the train sets which indicates the presence of relatively more active molecules as compared to the inactive ones. To ensure a fair comparison, the same training and test sets were used for each model's development. Activity distribution graph is shown in Figure 2.

QSAR investigations of the 4-aminoquinolines series resulted in several QSAR equations, considering the term selection criterion as r2, q2 and pred_r2. Some statistically significant 2D and 3D QSAR models were chosen for discussion. The 2D QSAR study of 24 compounds (divided into 5 test and 19 training) for malaria activity (Table 1) through MLR, PLSR, and PCR analysis coupled with SW variable selection resulted in the some statistical models. The inter-correlation among the selected descriptors was very less due to auto-scaling and cross correlation limit permitted was 0.5.

Model 1 (MLR)

pIC50 = 0.5325 (LUMOEnergy) - 0.2134 (SdsCHE-index) + 0.0053 (Quadrupole1) + 5.0510

Ntraning = 19, Ntest = 5, DF= 15, r2 = 0.7808, q2 = 0.6492, F test = 17.8114, r2_se = 0.2203, q2_se = 0.2787, pred_r2 = 0.6125, pred_r2se = 0.3038, ZScore Q^2 = 4.63378, Best Rand Q^2 = 0.35968.

In model 1, compound 16 is an outlier, as the residual of observed and the predicted activity is more than twice the standard error of the equation. After removing the outlier compound 16, we obtain another QSAR model 2.

Model 2 (MLR)

pIC50 = 0.5325 (LUMOEnergy) - 0.2134 (SdsCHE-index) + 0.0053 (Quadrupole1) + 5.0510

Ntraning = 19, Ntest = 5, DF= 15, r2 = 0.7808, q2 = 0.6492, F test = 17.8114, r2_se = 0.2203, q2 se = 0.2787, pred_r2 = 0.9044, pred_r2se = 0.1443, ZScore Q^2 = 7.65957, Best Rand Q^2 = -0.08872.

The statistically best significant model (Model 2) with a coefficient of determination (r2) = 0.7808 was considered, as the model showed an internal predictive power (q2 = 0.6492) of 64% and a predictivity for the external test set (pred_r2 = 0.9044) of about 90%. The low standard error of r2_se=0.2203 demonstrates accuracy of the model. In this QSAR equation, the positive contribution of LUMOEnergy and Quadrupole1 on the biological activity indicates that the increase in LUMOEnergy and Quadrupole1 of molecule lead to better antimalarial activity. The negative coefficient of SdsCHE-index (Electrotopological state indices for number of –CH group connected with one double and one single bond) shows that increase in SdsCHE-index is detrimental for the activity. It suggests that the presence of allyl group at the R position is detrimental to activity.

The descriptors selected for modeling inhibitory activity of the 4-aminoquinoline derivatives are summarized in Table 2 and the correlation matrix between the physico-chemical parameters and the biological activity is presented in Table 5. The plots of calculated vs. observed values of pIC50 are shown in Figure 3.

The contribution charts for all the significant models are presented in Figure 4, which gives the percentage contribution of the descriptors used in deriving the model. The predicted (LOO) activities of the compounds by the above model is shown in Table 6. The same data set subjected to the PLSR method resulted in r2 of 0.7785 and an internal predictive power of 64%, with an external predictivity of 72% (Model 3).

Model 3 (PLS)

pIC50 = 0.6501 (LUMOEnergy) -0.2304 (SdsCHE-index) -0.0077 (Quadrupole3) + 5.3044

Ntraning = 19, Ntest = 5, Optimum Components = 2, DF= 16, r2 = 0.7785, q2 = 0.6420, F test = 28.1128, r2_se = 0.2145, q2_se = 0.2726, pred_r2 = 0.5773, pred_r2se = 0.3173, ZScore Q^2 = 3.42538, Best Rand Q^2 = 0.22586.

In model 3, compound 16 is an outlier, as the residual of observed and the predicted activity is more than twice the standard error of the equation. After removing the outlier compound 16, we obtain another QSAR model 4 with a coefficient of determination (r2) = 0.7785 was considered, as the model showed an internal predictive power (q2 = 0.6420) of 64% and a predictivity for the external test set (pred_r2 = 0.7181) of about 71%.

Model 4 (PLSR)

pIC50 = 0.6501 (LUMOEnergy) -0.2304 (SdsCHE-index) -0.0077 (Quadrupole3) + 5.3044

Ntraning = 19, Ntest = 5, Optimum Components = 2, DF= 16, r2 = 0.7785, q2 = 0.6420, F test = 28.1128, r2_se = 0.2145, q2_se = 0.2726, pred_r2 = 0.7181, pred_r2se = 0.2477, ZScore Q^2 = 3.42538, Best Rand Q^2 = 0.22586.

It is apparent from the model 4 that the descriptor LUMOEnergy plays a pivotal role in determining activity. The descriptors LUMOEnergy contributed positively in the mathematical representation of the model and is favored to biologic activity in the aforementioned model. The descriptors SdsCHE-index and Quadrupole3 indicate a negative contribution to the antimalarial activity.

To improve the external predictivity of the model, PCR analysis with the same data set was performed, which resulted in r2 of 0.7393 and an internal predictive power of 62%, with the good external predictivity of 91%. The overall statistical significance level was found to exceed 99.9% (Model 5). The q2 was 0.6222, which shows the good internal prediction power of this model. Another parameter for predictivity of test set compound is high pred_r2=0.9135 and low pred_r2se=0.1435, which is showing good external predictive power of the model.

Model 5 (PCR)

pIC50 = 0.3252 (kappa3) - 0.1610 (SdsCHE-index) + 0.0051 (Quadrupole1) + 0.0985 (ZcompDipole) + 2.4672

Ntraning = 19, Ntest = 5, Optimum Components = 3, DF= 15, r2 = 0.7393, q2 = 0.6222, F test = 20.1821, r2_se = 0.2403, q2_se = 0.2893, pred_r2 = 0.9135, pred_r2se = 0.1435, ZScore Q^2 = 6.56807, Best Rand Q^2 = 0.09818.

The positive coefficient of kappa3 shows that increase in kappa3 is favored for the antimalarial activity. The next most important descriptors which influence the activity variation are ZcompDipole and Quadrupole1and are directly proportional to the activity.

The statistical significance of these models was further supported by the ‘fitness plot’ obtained for each model; this is a plot of observed vs. predicted activity of training- and test-set compounds and provides an idea about how well the model was trained and how well it predicts the activity of the external test set

(Figure 3). The closeness of observed to predicted activity reported in Table 6 also adds to this fact.

For a better understanding of the QSAR models of such 4-aminoquinoline compounds, an attempt to generate a 3D-QSAR model on MFA for the same set of compounds has also been made. For 3D QSAR, kNN–MFA of 4-aminoquinoline derivatives having inhibitory activities against 3D7 was performed. The SW variable selection method resulted in several statistically significant models, of which the corresponding best model 6 is reported herein (Table 6). The model selection criterion is the value of q2, the internal predictive ability of the model, and that of pred_r2, the ability of the model to predict the activity of external test set.

Model 6 (kNN MFA)

S_1398 (-0.0024 -0.0023), S_56 (-0.0024 -0.0024)

Ntraning = 19, Ntest = 5, q2 = 0.7107, q2_se = 0.2376, pred_r2 = 0.8521, pred_r2se = 0.1689

For antimalarial activity against 3D7, model 6 was found to be statistically most significant, especially with respect to the internal predictive ability (q2 = 0.7107) of the model. As the cross-validated correlation coefficient (q2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. The value of pred_r2 was obtained for the test set and gave better results, with a value of 0.8521, which means 85% predictive power for the external test set.

The kNN-MFA contour plots (Figure 5), which showed the relative position and ranges of the corresponding important electrostatic/steric fields in the model, provided guidelines for new molecule design. The range is based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbor set. The plot of observed versus predicted activities for the training and test sets of compounds are represented in Figure 3.

The steric descriptor with positive or negative coefficients shows a region where bulky substituent is favored or disfavored, respectively. Electrostatic field descriptors with positive coefficients represent regions where electropositive (electron-withdrawing) groups are favorable, whereas negative coefficient indicates that electronegative (electron-rich or electron-donating) groups are favorable in this region[31].

It is observed that electrostatic descriptors like S_1398 with negative coefficient indicates that less bulky groups are favorable on this site and presence of bulky groups decrease the antimalarial activity of these compounds. Most of the compounds (like 2-9 etc) with higher activity having less bulky substitution (thiourea group) at the side chain of 7-chloro, 4-aminoquinoline ring strongly support the above comment.

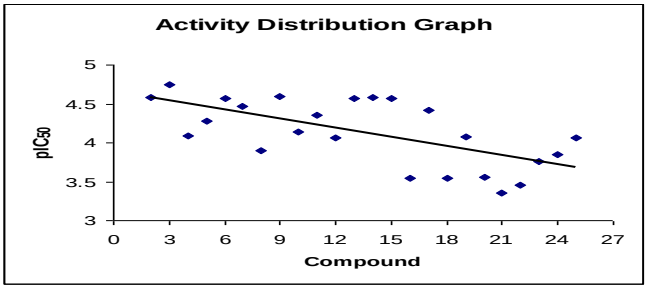
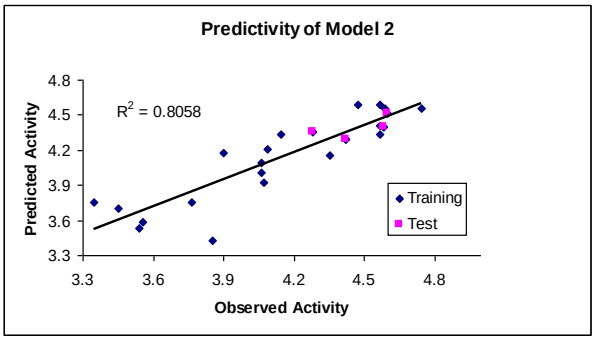
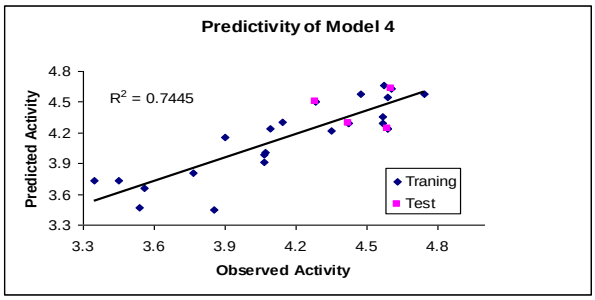


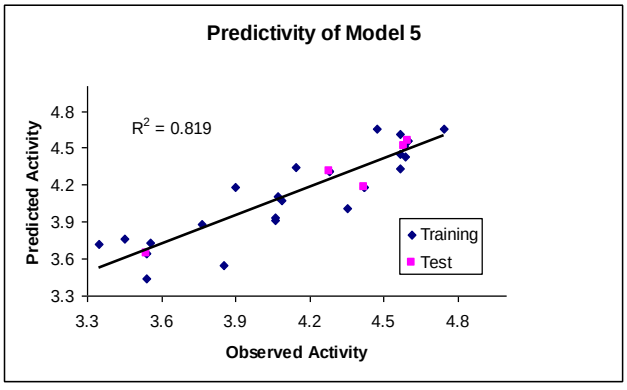
Figure 2. Activity distribution graph



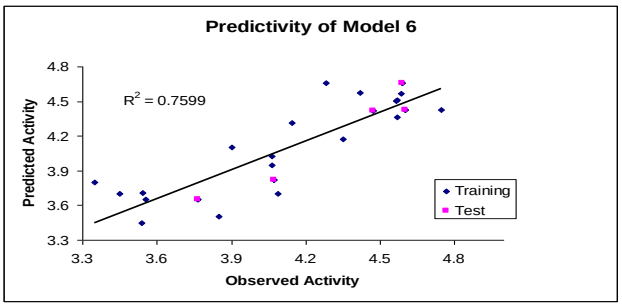
Model 2 (MLR)



Model 4 (PLSR)

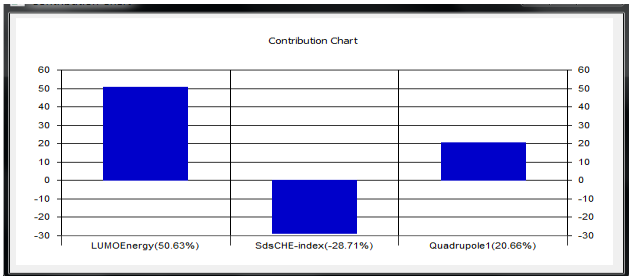


Model 5 (PCR)

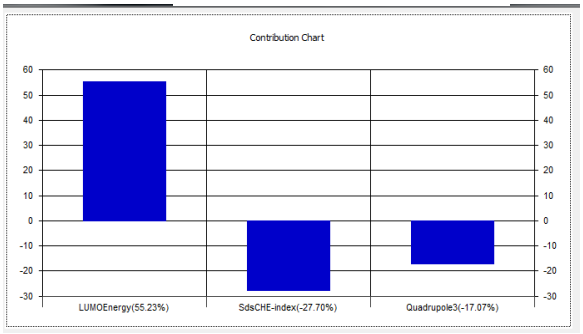


Model 6 (kNN-MFA)

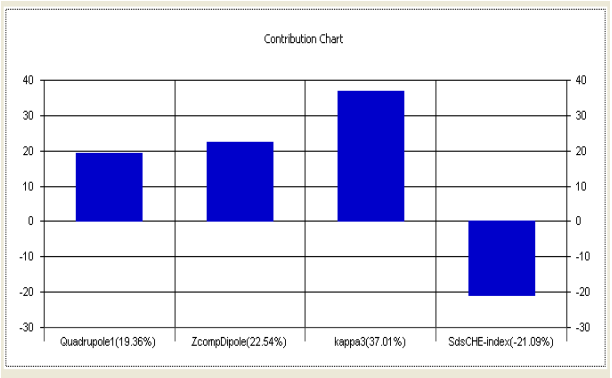
Figure 3. Graphs of observed vs. predicted activity of models 2, 4, 5 and 6.



Model 2 (MLR)



Model 4 (PLSR)



Model 5 (PCR)

Figure 4. Contribution charts of the statistically significant models obtained through 2D QSAR analysis

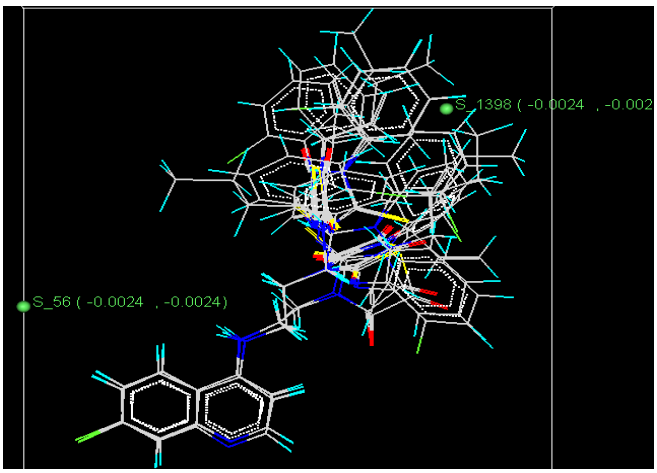


Figure 5. Contribution plot for steric and electrostatic interactions

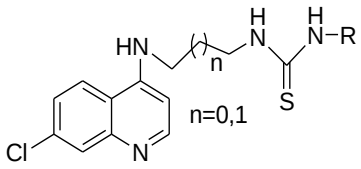
4. CONCLUSION

The present work reveals how the antimalarial activities of various side chain modified 7-chloro, 4-aminoquinolines may be treated statistically to uncover the molecular characteristics which are essential for high activity. 2D- and 3D-QSAR studies have been carried out on a series of 4-aminoquinoline derivatives with antimalarial activity against 3D7. The best 2D QSAR model 2 resulted in r2 = 0.7808, q2 = 0.6492 and pred_r2 = 0.9044 by SW-MLR confirms a positive contribution of the positive contribution of LUMOEnergy and Quadrupole1 to the antimalarial activity. Among various combinations, SW-based kNN method provides the best results (model 6) in 3D QSAR study with q2 = 0.7107 and pred_r2 = 0.8521. These results should prove to be an essential guide for the future work.

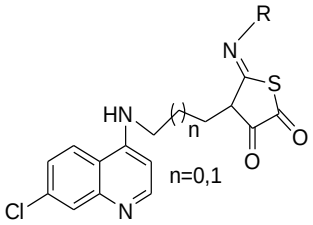
5. ACKNOWLEDGMENTS

The first author, Nitendra K Sahu is thankful to the Council of Scientific and Industrial Research, New Delhi, India for providing financial assistance.

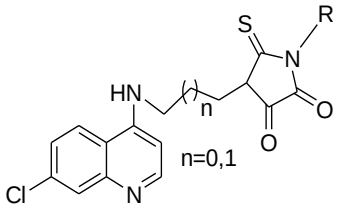
Table1. Structure and antimalarial activity of 4-aminoquinoline derivatives



(2-9)



(10-17)



(18-25)

Compound	n	R	IC_{50} (ng/mL)	pIC_{50} (M)
2[b]	0	Phenyl	9.22	4.588
3	0	Butyl	6.07	4.744
4	0	Allyl	26.11	4.089
5[a]	1	*o*-chlorophenyl	20.51	4.281
6	1	Phenyl	10.01	4.569
7[b]	1	Butyl	11.82	4.473
8	0	Allyl	42.02	3.901
9[a,b]	0	*o*-chlorophenyl	10.16	4.601
10	0	Phenyl	29.49	4.144
11	1	Butyl	17.44	4.351
12	1	Allyl	32.44	4.063
13	1	*o*-chlorophenyl	12.11	4.566
14[a]	0	Phenyl	11.05	4.585
15	0	Butyl	11.03	4.565
16[a]	0	Allyl	111.61	3.542
17[a]	1	*o*-chlorophenyl	17.48	4.42
18	1	Phenyl	119.10	3.538
19[b]	1	Butyl	33.08	4.072
20	0	Allyl	104.09	3.556
21	0	*o*-chlorophenyl	199.31	3.349
22	0	Phenyl	150.55	3.451
23[b]	1	Butyl	69.70	3.764
24	1	Allyl	54.85	3.851
25	1	*o*-chlorophenyl	39.71	4.063

[a]Indicates the compounds considered in the test set for 2D QSAR study.

[b]Indicates the compounds considered in the test set for 3D QSAR study.

Table 2. List of selected 2D descriptors used in the present QSAR study

Descriptor	Defination
LUMOEnergy	This descriptor signifies energy of highest unoccupied molecular orbital.
Quadrupole1	This descriptor signifies magnitude of first tensor of quadrupole moments.
SdsCHE-index	Electrotopological state indices for number of –CH group connected with one double and one single bond.
kappa3	This descriptor signifies third kappa shape index: (n-1) (n-3)2 / p32 for odd n, and (n-3) (n-2)2 / p32 for even n
ZcompDipole	This descriptor signifies the z component of the dipole moment (external coordinates).

Table 4. Unicolumn statistics of the training and test sets for antimalarial activity

Data Set	Average (Mean)	Max	Min	StdDev	Sum
2D					
Training	4.0893	4.7443	3.3492	0.4296	77.6966
Test	4.2856	4.6009	3.5421	0.4358	21.4282
3D					
Training	4.0856	4.7443	3.3492	0.4417	77.6268
Test	4.2996	4.6009	3.7641	0.3681	21.4979

Table 5. Correlation matrix between descriptors present in the 2D QSAR models 2

	pIC_{50}	**LUMO Energy**	**SdsCHE-index**	**Quadrupole1**
pIC_{50}	1			
LUMOEnergy	0.706	1		
SdsCHE-index	-0.4	0.091	1	
Quadrupole1	0.412	0.358	0.181	1

Table 3. Selected 2D and 3D descriptors of side chain modified 4-aminoquinoline derivatives

Compd	LUMOEnergy	SdsCHE-index	Quadrupole1	Quadrupole3	kappa3	ZcompDipole	S_56	S_1398
2	-1.095	0	15.95	5.816	5.762	0.092	-0.0024	-0.0023
3	-1.029	0	16.71	5.054	6.504	-0.12	-0.0023	-0.0022
4	-0.954	1.873	11.877	1.496	5.606	0.317	-0.0024	-0.002
5	-1.174	0	-13.107	5.497	5.736	0.434	-0.0024	-0.0024
6	-1.043	0	17.26	-5.068	6.282	0.135	-0.0024	-0.0023
7	-1.029	0	15.722	6.811	6.65	-0.56	-0.0024	-0.0021
8	-1.03	1.881	14.865	5.419	6.204	-0.73	-0.0024	-0.0016
9	-1.108	0	9.8883	-6.07	6.261	0.051	-0.0024	-0.0024
10	-1.533	0	19.056	1.207	5.202	0.874	-0.0025	-0.0016
11	-1.445	0	-23.424	19.33	5.299	-0.6	-0.0025	-0.0029
12	-1.488	1.678	20.715	5.255	4.838	0.628	-0.0024	-0.0015
13	-1.578	0	22.847	-2.54	5.258	1.618	-0.0025	-0.0016
14	-1.489	0	26.988	12.71	5.627	0.812	-0.0025	-0.0016
15	-1.422	0	22.456	3.469	5.758	-1.18	-0.0025	-0.0016
16	-1.441	1.686	32.383	-3.902	5.299	-4.13	-0.0028	-0.0036
17	-1.538	0	12.019	1.174	5.689	-2	-0.0025	-0.0017
18	-2.356	0	-49.728	38.68	4.542	-2.56	-0.0025	-0.0064
19	-2.282	0	17.358	-23.99	4.716	0.156	-0.0024	-0.0019
20	-2.27	1.644	17.858	-26.74	4.296	0.367	-0.0024	-0.0018
21	-2.333	0	-10.288	7.402	4.621	-1.98	-0.0025	-0.0043
22	-2.304	0	-22.475	8.898	4.924	-1.89	-0.0025	-0.0018
23	-2.24	0	-19.545	5.5	5.136	-1.6	-0.0025	-0.0019
24	-2.128	1.652	-10.435	3.256	4.716	-1.35	-0.0025	-0.0018
25	-2.255	0	45.263	-18.81	5.01	-4.22	-0.0027	-0.0033

Table 6. Comparative Observed and Predicted Activities (LOO) of side chain modified 4-aminoquinolines by 2D and 3D QSAR Models.

Comp	pIC_{50} (M)	2D QSAR						3D QSAR	
		MLR (2)	*Res.	PLSR (4)	*Res.	PCR (6)	*Res.	kNN-MFA (7)	*Res.
2	4.588	4.553	0.035	4.548	0.04	4.431	0.156	4.657	-0.069
3	4.744	4.557	0.188	4.581	0.164	4.656	0.213	4.425	0.32
4	4.089	4.206	-0.12	4.241	-0.15	4.08	0.009	3.704	0.386
5	4.281	4.356	-0.08	4.499	-0.22	4.308	-0.028	4.657	-0.376
6	4.569	4.592	-0.02	4.665	-0.1	4.617	-0.048	4.512	0.056
7	4.473	4.586	-0.11	4.583	-0.11	4.655	-0.227	4.417	0.056
8	3.901	4.18	-0.28	4.159	-0.26	4.186	-0.285	4.103	-0.202
9	4.601	4.514	0.087	4.631	-0.03	4.559	0.042	4.425	0.176
10	4.144	4.336	-0.19	4.299	-0.15	4.342	-0.198	4.314	-0.17
11	4.351	4.157	0.193	4.216	0.135	4.011	0.339	4.172	0.179
12	4.063	4.011	0.052	3.91	0.153	3.938	0.125	4.023	0.04
13	4.566	4.332	0.234	4.298	0.268	4.453	0.113	4.364	0.201
14	4.585	4.401	0.184	4.238	0.347	4.515	0.07	4.565	0.02
15	4.565	4.413	0.152	4.353	0.212	4.338	0.227	4.502	0.063
16*	3.542	-	-	-	-	3.642	-0.135	3.706	-0.164
17	4.42	4.296	0.124	4.296	0.124	4.182	0.238	4.575	-0.155
18	3.538	3.532	0.005	3.474	0.064	3.438	0.1	3.446	0.092
19	4.072	3.928	0.144	4.006	0.066	4.105	-0.032	3.823	0.25
20	3.556	3.586	-0.03	3.657	-0.1	3.727	-0.17	3.651	-0.094
21	3.349	3.754	-0.4	3.73	-0.38	3.722	-0.373	3.803	-0.453
22	3.451	3.705	-0.25	3.738	-0.29	3.768	-0.317	3.704	-0.253
23	3.764	3.754	0.01	3.805	-0.04	3.88	-0.116	3.651	0.114
24	3.851	3.425	0.426	3.45	0.401	3.549	0.333	3.504	0.347
25	4.063	4.09	-0.03	3.983	0.08	3.912	0.151	3.946	0.117

* Outlier

REFERENCES

1 Aregawi M, Cibulskis R, Otten M, Williams R, Dye C. World malaria report, World Health Organization, 2008.

2 Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. Nat Rev Drug Discov. 2006; 5: 941–955.

3 O'Neill PM, Bray PG, Hawley SR, Ward SA, Park BK. 4-aminoquinolines-past, present, and future; a chemical perspective. Pharmacol Ther. 1998; 77: 29.

4 Dominguez JN. Chemotherapeutic agents against malaria: what next after chloroquine? Curr Topics Med Chem. 2002; 2: 1173.

5 Madrid PB, Liou AP, DeRisi JL, Guy RK. Incorporation of an Intramolecular Hydrogen-bonding Motif in the Side-Chain of 4-Aminoquinolines Enhances Activity against Drug-Resistant P. falciparum. J Med Chem. 2006; 49: 4535.

6 Kaur K, Jain M, Reddy RP, Jain R. Quinolines and structurally related heterocycles as antimalarials. Eur J Med Chem. 2010; 45: 3245-3264.

7 Olliaro PL, Yuthavong Y. An overview of chemotherapeutic targets for antimalarial drug discovery. Pharmacol Ther. 1999; 81: 91–110.

8 Sahu NK, Sahu S, Kohli DV. Novel molecular targets for antimalarial drug development. Chem Biol Drug Des. 2008; 71: 287–297.

9 Vippagunta SR, Dorn A, Matile H, Bhattacharjee AK, Karle JM, Ellis WY, Ridely RG, Vennerstrom JL. Structural specificity of chloroquine-hematin binding related to inhibition of hematin polymerization and parasite growth. J Med Chem. 1999; 42: 4630.

10 Sunduru N, Srivastava K, Rajakumar S, Puri SK, Saxena JK, Chauhan PMS. Synthesis of novel thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline as potent antimalarials. Bioorg Med Chem Lett. 2009; 19: 2570–2573.

11 Hanch C, Leo A. Exploring QSAR, Fundamentals and Applications in Chemistry and Biology.: American Chemical Society, Washington, DC (1995) 90–100

12 Abraham DJ. Burger's, Medicinal Chemistry and Drug Discovery. 6th edition, vol.I- Drug Discovery, A John Wiley & Sons Inc. Publication, 3-27.

13 Agrawal VK, Srivastava R, Khadikar, PV. QSAR studies on some antimalarial sulfonamides. Bioorg Med Chem. 2001; 9: 287-293.

14 Xue CX, Cui SY, Liu MC, Hu ZD, Fan BT. 3D QSAR studies on antimalarial alkoxylated and hydroxylated chalcones by CoMFA and CoMSIA. Eur J Med Chem 2004; 39: 745-753.

15 Zahouily M, Lazar M, Elmakssoudi A, Rakik J, Elaychi S, Rayadh A. QSAR for anti-malarial activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives. J Mol Model. 2006; 12: 398-405.

16 Srivastava M, Singh H, Naik PK. Quantitative structure–activity relationship (QSAR) of artemisinin: the development of predictive in vivo antimalarial activity models. J Chemometrics. 2009; 23: 618–635.

17 Kubinyi H. Variable selection in QSAR studies. Quant Struct Act Relat. 1994; 13: 285–294.

18 Ajmani S, Jadhav K, Kulkarni SA. Three-Dimensional QSAR using the k-Nearest Neighbor method and its interpretation. J Chem Inf Model. 2006; 46: 24-31.

19 Ajmani S, Kulkarni SA. A dual-response partial least squares regression QSAR model and its application in design of dual activators of PPARα and PPARγ. QSAR Comb Sci. 2008; 27: 1291-1304.

20 VLife: molecular design suite. Version 3.5. VLife Technologies, Pune, India. Available at www.vlifesciences.com

21 Halgren TA. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem. 1996; 17: 553–586.

22 Golbraikh A, Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput-Aided Mol Des. 2002; 16: 357-369.

23 Clark M, Cramer RD III, Van ON. Validation of the general purpose Tripose 5.2 Force Field. J Comput Chem. 1989; 10:982–1012.

24 Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. Tetrahedron. 1980 ; 36: 3219–3228.

25 Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learning Res. 2003; 3: 1157-11820.

26 Seidel JK, Schaper KJ. Chemische. Struktur und biologische. Aktivität von Wirkstoffen, New York: Verlag Chemie, Weinheim, 1979. 115-252

27 Wold S. PLS for multivariate linear modeling. VCH: Weinheim, 1995. 195-218

28 Sharaf MA, Illman DL, Kowalski BR. Chemometrics. New York: Wiley, 1986.

29 Golbraikh A, Tropsha A. Beware of q^2! J Mol Graph Model. 2002; 20: 269-276.

30 Melagraki AG, Sarimveis H, Igglessi-Markopoulou O, Kolliasnovel GA. QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-aryl-[1,4] diazepane ureas. Eur J Med Chem. 2009; 44: 877-884.

31 Samee W, Nunthanavanit P, Ungwitayatorn J. 3D-QSAR investigation of synthetic antioxidant chromone derivatives by molecular field analysis. Int J Mol Sci. 2008; 9: 235-246.