QSAR analysis of 7-substituted 4-aminoquinolines as antimalarial agents

Nitendra K. Sahu

RKDF College of Pharmacy, S.R.K. University, Bhopal, India

* For Correspondance: Nitendra K. Sahu
Tel: 91-9981219293
E-mail: nksahupharma@gmail.com

ABSTRACT: The emergence and rapid spread of chloroquine resistant strains of *Plasmodium falciparum* has dramatically reduced the chemotherapeutic options. Towards this goal, the quantitative structure-activity relationship analysis of some synthesized 7-substituted 4-aminoquinolines were performed for their antiplasmodial activity against chloroquine-resistant parasites to find out the structural features responsible for the biological activity. The statistically significant best 2D QSAR model having correlation coefficient ($r^2$) = 0.8631 and cross validated squared correlation coefficient ($q^2$) = 0.8101 with external predictive ability (pred_r^2) = 0.6740 was developed by Partial Least Square Regression coupled with stepwise backward-forward method using Vlife MDS 3.5 software and showed that the parameters Average-vePotential, T_2_N_5, XcompDipole and QMDipoleX are highly correlated with antiplasmodial activity of 7-substituted 4-aminoquinolines. The developed models are interpretable, with good statistical and predictive significance, and can be used for guiding ligand modification for the development of potential new antimalarial agents.

KEYWORDS: QSAR; 7-substituted 4-aminoquinolines; antimalarial agents; MLR; PLSR
Introduction

Malaria is a devastating disease caused by four species of genus plasmodium which afflicts more than 40% of the world population, causing an estimated mortality of 1.5–2.7 million people annually\(^1\). During the past six decades, chloroquine (CQ) and other aminoquinolines have been the frontline antimalarial agents because of their therapeutic efficacy and lower cost. However, development of resistance has severely limited the choice of available antimalarial drugs\(^2\), which clearly highlights the urgent need of novel chemotherapeutic agents for the treatment of malaria.

A major initiative in this direction is to find enzyme targets that are critical to the disease process or essential for the survival of the parasite. Identification and design of novel chemical entities specifically affecting these targets could lead to better drugs for the treatment of malaria\(^3\). 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 bindings\(^4\). So, rather than identifying new molecules for efficacy, modified 7-chloroquinolines having many advantages and efficiency are now in priority for antimalarial chemotherapy.

Quantitative structure activity relationship (QSAR) is one of the major tools in drug discovery to explore ligand–receptor/enzyme interactions, especially when the structural details of the target are not known or if there are multiple targets.

The quantitative structure-activity relationship (QSAR) approach helps to correlate the specific biological activities or physical properties of a series of compounds with the measured or computed molecular properties of the compounds, in terms of descriptors\(^5\). QSAR methodologies save resources and expedite the process of the development of new molecules and drugs. There have been many QSAR researches related to design of anti-malarial drugs so far\(^6-10\) but a systematic QSAR study is yet to be carried out for series of 7-substituted, 4-aminoquinolines.

The present study aimed to elucidate the structural features of 7-substituted, 4-aminoquinolines required for antimalarial activity and to obtain predictive QSAR models, which may guide the rational synthesis of novel plasmodium inhibitors.
Experimental

All molecular modeling studies were performed using the Vlife Molecular Design Suite software package\textsuperscript{11}. Structures of all the 40 compounds were sketched using the 2D draw application and converted to 3D structures. Energy minimization and geometry optimization were conducted using the Merck molecular force field (MMFF) method with the root mean square (RMS) gradient set to 0.01 kcal/mol Å and the iteration limit to 10,000\textsuperscript{12}.

Data set

A number of 7-substituted, 4-aminoquinoline derivatives having antiplasmodial activity\textsuperscript{13} were considered in the present QSAR study (Table 1). The biological activity values [IC\textsubscript{50} (nM)] reported in literature were converted to their molar units and then further to negative logarithmic scale (pIC\textsubscript{50}) and subsequently used as the dependent variable for the QSAR analysis.

The energy-minimized geometry was used for the calculation of the various 2D descriptors (Individual, Chi, ChiV, Path count, ChiChain, ChiVChain, Chainpathcount, Cluster, Pathcluster, Kapa, Element Count, Estate number, Estate contribution, Semi-empirical, Hydophillic-hydophobic and Polar surface area). The various alignment-independent (AI) descriptors were also calculated. For calculation of alignment, the independent descriptor was assigned the utmost three attributes. The first attribute was T to characterize the topology of the molecule. The second attribute was the atom type, and the third attribute was assigned to atoms taking part in the double or triple bond. The preprocessing of the independent variables (i.e., 2D descriptors) was done by removing invariable (constant column), which resulted in total 280 descriptors to be used for QSAR analysis.

The sphere exclusion (SE) method\textsuperscript{14-15} was adopted for division of training and test data set comprising of 30 and 10 molecules, respectively, with dissimilarity value of 2.0 where the dissimilarity value gives the sphere exclusion radius. The spherical exclusion method employs the following algorithm: (i) select a point and include it in the training set; (ii) build a sphere with radius R with a center in this point; (iii) include all points within the sphere, except for the center, in the test set; (iv) discard all points in the sphere from the initial set; (v) if no points are left, stop, otherwise go to step (i). Ten compounds, namely, 24, 26, 36, 44, 47, 49, 53, 57, 60 and 66 were used as test set while the remaining molecules as the training set (Table 2). The unicolumn statistics of the training and test sets is reported in Table 3.
Statistical computation

All the calculated descriptors were considered as independent variable and biological activity as dependent variable. VLife Molecular Design Suite (VLifeMDS software was used to generate QSAR models by Multiple Linear Regression (MLR), Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) method analysis.

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 $r^2$. 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.

The developed QSAR models were evaluated using the following statistical measures: $r^2$ (the squared correlation coefficient), F test (Fischer’s value) for statistical significance, $q^2$ (cross-validated correlation coefficient); pred$_r$$^2$, $r^2$ for external test set. The regression coefficient $r^2$ is a relative measure of fit by the regression equation. It represents the part of the variation in the observed data that is explained by the regression. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: $r^2 > 0.6$, $q^2 > 0.6$ and pred$_r$$^2 > 0.5^{15}$. 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 $r^2$ ($r^2_{-se}$), $q^2$ ($q^2_{-se}$) and pred$_r$$^2$ (Pred$_r$$^2_{-se}$) shows absolute quality of fitness of the model.

Internal validation was carried out using ‘leave-one-out’ ($q^2$, LOO) method$^{16}$. The cross-validated coefficient, $q^2$, was calculated using the following equation:

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

where $y_i$, and $\hat{y}_i$ are the actual and predicted activity of the $i$th molecule in the training set, respectively, and $y_{mean}$ is the average activity of all molecules in the training set.

However, a high $q^2$ value does not necessarily give a suitable representation of the real predictive power of the model for antimalarial ligands. So, an external validation was also carried out in the present study. The external predictive power of the model was assessed by predicting pIC$_{50}$ value of the 10 test set molecules, which were not included in the QSAR model development. The predictive ability of the selected model was also confirmed by pred$_r$$^2$. 

102
where \( y_i \) and \( \hat{y}_i \) are the actual and predicted activity of the \( i^{\text{th}} \) molecule in the test set, respectively, and \( y_{\text{mean}} \) is the average activity of all molecules in the training set.

### Results and Discussion

The 2D QSAR study of 40 compounds (divided into 10 test and 30 training) for antiplasmodial activity through MLR, PLSR, and PCR analysis using VLife MDS resulted in the following statistically significant model, considering the term selection criterion as \( r^2 \), \( q^2 \) and \( \text{pred}_r^2 \). The training and test sets were selected by sphere exclusion method and the models were validated by both internal and external validation procedure. To ensure a fair comparison, the same training and test sets were used for each model’s development (Table 2). 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 3).

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

1. The maximum value of pIC\(_{50}\) of test set should be less than or equal to maximum value of pIC\(_{50}\) of training set.
2. The minimum value of pIC\(_{50}\) of test set should be higher than or equal to minimum value of pIC\(_{50}\) 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 pIC\(_{50}\) 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.

The statistically significant model (Model 1) with \( r^2 = 0.8631 \) was considered, as the model showed an internal predictive power (\( q^2 = 0.8101 \)) of 81% and a predictivity for the external test set (\( \text{pred}_r^2 = 0.6740 \)) of about 67%. This model indicates the positive contribution of Average-vePotential, XcompDipole and QMDipoleX and increase in the values of these descriptors are beneficial for the antimalarial activity of 7-substituted 4-aminquinolines. The descriptors T_2_N_5 defines the count of number of double bounded atoms separated from Nitrogen atom.
by 5 bonds and make a positive contribution to activity (like in compounds 45, 37, 49). The statistical results of QSAR models are shown in Table 4.

**Model 1 (SW-PLSR)**
\[
p_{IC_{50}} = 180.0310 \text{ (Average-vePotential) } + 0.5004 \text{ (T\textsubscript{2-N-5}) } + 0.2723 \text{ (XcompDipole) } + 0.1412 \text{ (QMDipoleX) } + 10.4043
\]

The descriptors selected for this model are summarized in Table 2 and the correlation matrix between the physico-chemical parameters and the biological activity for the models 1 is presented in Table 5.

The same data set subjected to the MLR method resulted in a coefficient of correlation of 0.8645 and an internal predictive power of 81%, with external predictivity of 66%.

**Model 2 (SW-MLR)**
\[
p_{IC_{50}} = 178.4470 \text{ (Average-vePotential) } + 0.5219 \text{ (T\textsubscript{2-N-5}) } + 0.2444 \text{ (XcompDipole) } + 0.1497 \text{ (QMDipoleX) } + 10.3724
\]

To improve the external predictivity of the model, PCR analysis with the same data set was performed, which resulted in \( r^2 \) of 0.6027 and an internal predictive power of 52%, with the good external predictivity of 67.6%. Average-vePotential contributes in the same manner as above. \( T\textsubscript{C-N-7} \) defines the count of number of Carbon atoms (single double or triple bonded) separated from any Nitrogen atom (single or double bonded) by 7 bond distance in a molecule and makes a negative contribution to activity.

**Model 3 (SW-PCR)**
\[
p_{IC_{50}} = 95.3804 \text{ (Average-vePotential) } + 20.0116 \text{ (Most-ve Potential) } - 0.1705 \text{ (T\textsubscript{C-N-7}) } + 12.1643
\]

The plots of calculated vs. observed values of \( p_{IC_{50}} \) and contribution chart for best Models 1 are shown in Fig. 1. The predicted (LOO) activities of the compounds by the above best models are shown in Table 6.
Conclusion

The present work shows how a set of antimalarial activities of various 4-quinolylhydrazones may be treated statistically to uncover the molecular characteristics which are essential for high activity. The generated models were analyzed and validated for their statistical significance and external prediction power. The results derived may be useful in further designing more novel antimalarial agents in series.

Acknowledgments

The author thanks Vlife Science Technologies Pvt. Ltd for providing the demo software for the study.

References

F.M. Mapping antimalarial pharmacophores as a useful tool for the rapid discovery of
drugs effective in vivo, design; construction; characterization; and pharmacology of
Fara, D.C.; Karelson, M. Antimalarial activity, a QSAR modeling using CODESSA PRO
10. Deshpande, S.; Solomon, V.R.; Katti, S.B.; Prabhakar, Y.S. Topological descriptors in
modelling antimalarial activity, N(1)-(7-chloro-4-quinolyl)-1;4-bis(3-
11. Halgren, T.A. Merck molecular force field. III. Molecular geometries and vibrational
antiplasmodial activity among 7-substituted 4-aminoquinolines. J. Med. Chem. 1998, 41,
4918-4926.
14. Golbraikh, A.; Tropsha, A. QSAR modeling using chirality descriptors derived from
16. Cramer, R.D.; Patterson, D.E.; Bunce, J.D. Comparative molecular field analysis
(CoMFA), Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc.
1988, 110, 5959-5967.