Asian J. Exp. Sci., Vol. 28, No. 2, 2014; 11-17

QSAR Analysis of Glutathione Reductase Inhibitors as Potent Anti-Malarials

Ankita Agarwal *, Ruchi Mishra
Department of Pharmacy, Banasthali University, Rajasthan-304022, India
*E-mail: paliwalsarvesh@yahoo.com

Abstract : Robust QSAR models were developed using MLR, PLS and FFNN for a set of 29, 4-aminoquinoline derivatives as *Pf* (K1 strain) glutathione reductase (GR) inhibitors. The statistical values of all the developed models were analyzed and compared. The results obtained from MLR, PLS models were comparable to the FFNN model. The results obtained from this study indicate that Dipole Moment X Component, VAMP LUMO and First Atom E-State index of 4-aminoquinoline derivatives plays an important role in determining the anti-malarial activity of the said compounds.

Keywords: Quantitative structure activity relationship, Multiple linear regression, Partial least square, Feed forward neural network.

INTRODUCTION

Malaria has been the biggest health problem accounting for 1.5 to 2.7 million deaths each year. This situation is rapidly worsening, mainly due to unavailability & effective drugs and development of drug resistance to the existing first line drugs (Smith, 1987 and Nosten, 1982 and Smith 1987) . It has been observed anti-malarial 4-aminoquinolines, like chloroquine, targets the acidic food vacuole of the parasites and inhibits heme bio-mineralization. Since safe, effective and affordable orally active therapies capable of addressing the problem of resistance are stringently needed, identification of new anti-malarial drug candidates is an urgent priority (Singh, 2006). There is an urgent need to evaluate the binding requirements of anti-malarial by employing computational approach. One of the most promising techniques to get insight into the structural requirements is QSAR, which is a mathematical relationship linking chemical structure and pharmacological activity in a quantitative manner for any given series of compounds. (Neaz, 2008). In view of this we decided to develop predictive models using MLR, PLS and FFNN. The main objective of the investigation has been to determine structural elements of 4-aminoquinoline derivatives responsible for anti-malarial activity, which can help in the design of novel GR inhibitors.

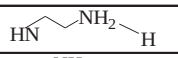
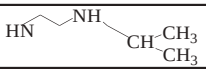
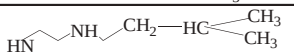
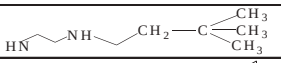
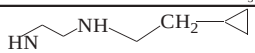
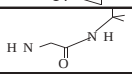
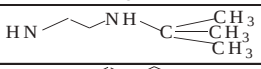
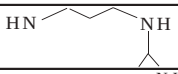
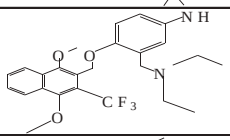
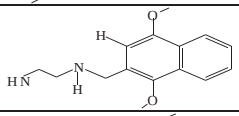
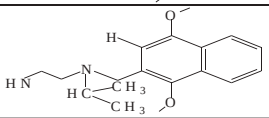
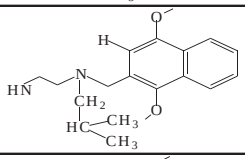
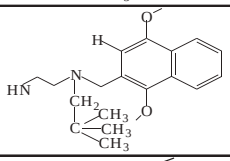
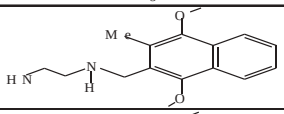
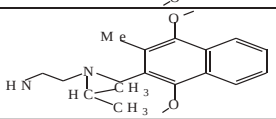
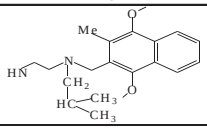
MATERIAL AND METHODS

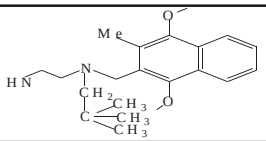
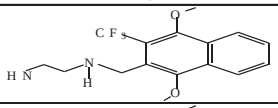
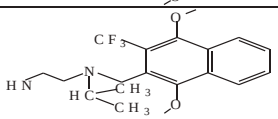
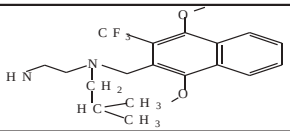
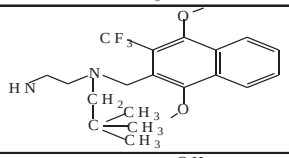
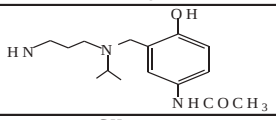
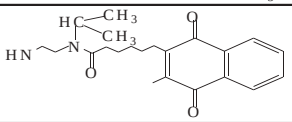
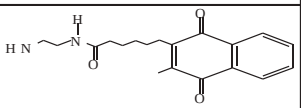
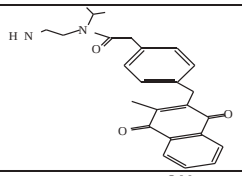
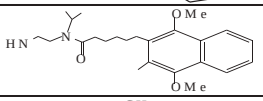
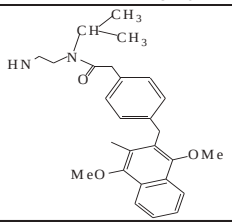
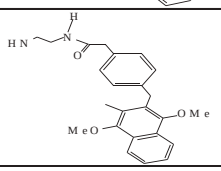
QSAR study was performed on a series of 29 4-aminoquinoline derivatives. Structures and biological activities of all compounds are given in table 1 (Friebolin and Jannak 2008).

Preparation of input for QSAR studies

The molecular structures were drawn and their geometries were cleaned using standalone module of Discovery Studio software and were subjected to energy minimization. All the structures were loaded to the worksheet of TSAR. Chemical encoding scheme was used to define substituent attached to the scaffold template by a single bond. All the structures and their defined substituent were converted into high quality 3D structures using Corina-make 3D option [Kramer, 2009]. Charges were calculated using charge-2 package available with TSAR.

Table I: Structure and biological activity data of 4-aminoquinoline derivatives

Comp. Name	R_1	IC_{50} (nM)
1		1800
2		96.7
3		72.0
4		6.8
5		145.0
7		39235
8		1,439
9		127.8
31		144.0
14		32.0
19		28.4
22		344.8
25		188.3
28		154.1
20		22.9
23		104.6
26		40.6

29		134.2
21		16.0
24		298.9
27		198.3
30		115.7
32		317.5
34		9.4
35		367.3
36		5.3
39		302.4
40		268.3
41		505.3

Dataset preparation and descriptor calculation

The IC_{50} values of all the 29 compounds used in the present study were converted into negative logarithm of IC_{50}. The data set was randomly divided into the training and test set of 21 and 8 (namely 1, 2, 5, 8, 9, 23, 28 and 31) compounds, respectively. The training set compounds were used to develop the QSAR model while the test set compounds were used to validate the developed model. The quality of QSAR model always depends on the accurate definition and appropriate use of molecular descriptors. In present study molecular descriptors were calculated for all the compounds under consideration using TSAR 3.3 software. Descriptors were obtained for all the structures as well as for substituents defined. Descriptors with the same values for all compounds were discarded. Pair-wise correlation analysis of remaining descriptors was performed (Agarwal *et al*., 2012). For each pair of descriptors, the correlation coefficient was calculated. If for two descriptors the correlation coefficient was higher than 0.6, regression for each of them was checked with biological activity. Descriptors having high regression value were retained and other one was discarded (Mishra *et al*., 2012).

Linear regression analysis

A linear mathematical function that relates descriptor values to the activity may be created using MLR and PLS.

The standard way of evaluating a model in QSAR is to analyze r^2 between the activity and a set of independent descriptors. However, a good QSAR model is considered to be one with a high value of r^2, r^2_{cv} and F. r^2 represents the goodness of fit of the model, r^2cv is the cross-validated r^2 (a measure of the quality of the QSAR model). The modeling is taken to be optimal when *F* reached a maximum together with minimum standard error (Konovalov, 2008). Another reliability check used was value of standard deviation (S), which should be low. As an approach to check the robustness and the predictive ability of the models generated using MLR analysis, PLS analysis was performed on the same training set of compounds. Similar to the cross-validation method used in MLR, model generated during PLS analysis was also validated using leave out one row (Paliwal *et al*., 2009 and Cramer, 1993).

During the course of MLR and PLS analysis some outliers were identified to maximize the predictability of the model (Kramer, 2009) hence removed from the training set various models were developed and checked for statistical fitness. The models which satisfied the entire statistical requirement were chosen and analyzed for descriptor contribution towards anti-malarial activity.

Feed forward neural network analysis

To display the dependency of each molecular descriptor (in a qualitative manner), a constant value was fed into all input nodes, except for the molecular descriptor in question, which was varied over a range of 0.1-1.0. An initial weighting value of 1.0 was applied to all connections. Starting weights in the range of -0.03 to +0.03 and -1 to +1 for the initial node biases were selected. The FFNN architecture was set to 3-1-1. The results were visualized on a 2D plot of output node against input.

RESULT AND DISCUSSION

Linear multivariate analysis

MLR and PLS were used to derive the QSAR equations. After data reduction three descriptors were identified which were independent to each other and were used to develop regression model.

Outliers in QSAR can be very important and interesting, especially when the observed biological activity is higher than the predicted one by the developed model (higher residual value). Four outliers (4, 19, 20 and 26) were detected with the help of regression line of equation. After excluding these compounds from the training set, the statistical characteristics were improved significantly. Therefore, these compounds were eventually excluded from the training set. Best MLR model obtained, with excellent r^2_{cv} and r-values for the training set are represented by equation-1.

Equation-1 (MLR)

$IC_{50} = -0.26 \times X1 + 2.47 \times X2 - 1.35 \times X3 - 11.63$

$r = 0.91$, $r^2 = 0.83$, $r^2_{cv} = 0.76$, f value = 37.39 and s value = 0.37

Results obtained from conventional MLR were checked with PLS analysis using same data set. PLS model (equation-2) complemented the MLR model in terms of (r^2) and predictability (r^2_{cv}).

Equation-2 (PLS)

$IC_{50} = -0.26 \times X1 + 2.47 \times X2 - 1.35 \times X3 - 11.63$

In equation [1] and [2], X1 is Dipole Moment X Component (Subst. 1), X2 is First Atom E-State index (Subst. 1), X3 is VAMP LUMO (Whole Molecule)

Statistical significance = 0.96, $r^2 = 0.83$, $r^2_{cv} = 0.77$

Since for a well defined problem, both MLR and PLS should generate comparable results, the r^2 and r^2_{cv} values of MLR and the PLS models were evaluated and it was found that both the models have comparable value of $r^2 = 0.8370$ for MLR and $r^2 = 0.8370$ for PLS, which is a relative measure of fit by the regression equation. The high cross-validated squared correlation coefficient value of $r^2_{cv} = 0.76$ for MLR and r^2_{cv} value of = 0.77 for PLS also shows good internal productivity of the model (Hawkins, 2003).

Test set prediction

The external predictive capability of QSAR model was also checked using test sets of compounds that were excluded during model development. All the compounds in the test set were treated in a manner analogous to the compounds in the training set. The r^2 value of MLR= 0.76

and PLS= 0.76 derived for the test set illustrate the high predictive ability of the model.

Feed forward neural network analysis

The results were visualized on a 2D plot of output node against input. In present study a three layered neural network has been used. The input descriptors were the same as used for multivariate regression (MLR and PLS). A close correlation coefficients for training set were given by the trained neural network architecture ($r^2_{training}$=0.841). The predictivity of MLR was compared to FFNN using same external test set and value of r^2 was found to be comparative (0.808). The dependency plots obtained in FFNN are given in fig. 4-6. The analysis of all the plots reveals that the relationship between biological activity and three descriptors is linear and analogous to MLR and PLS analysis.

The actual and predicted activity obtained from MLR, PLS and FFNN analysis for the training and test set of compounds are shown in fig. 1-3.

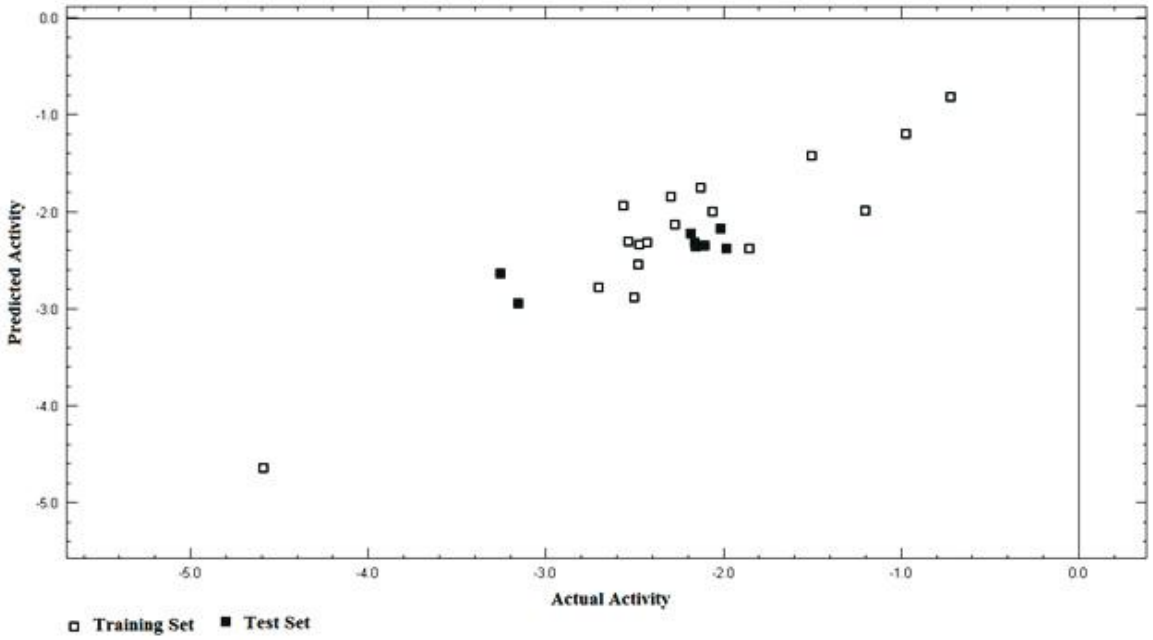


Fig. 1: Plot between actual and predicted value using MLR analysis

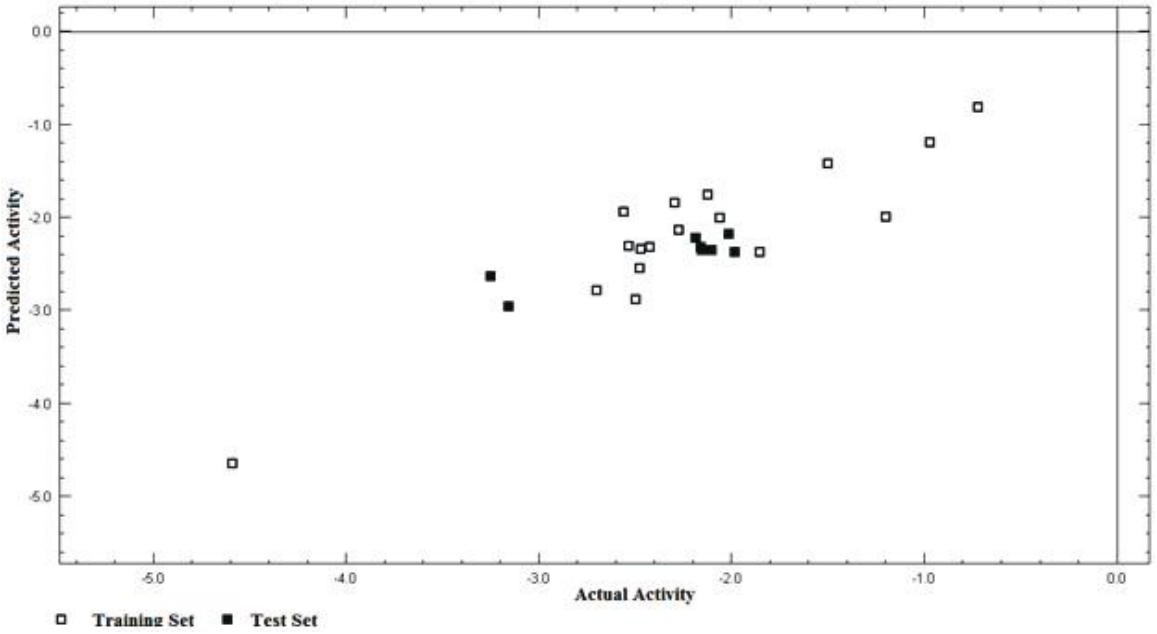


Fig. 2: Plot between actual and predicted value using PLS analysis

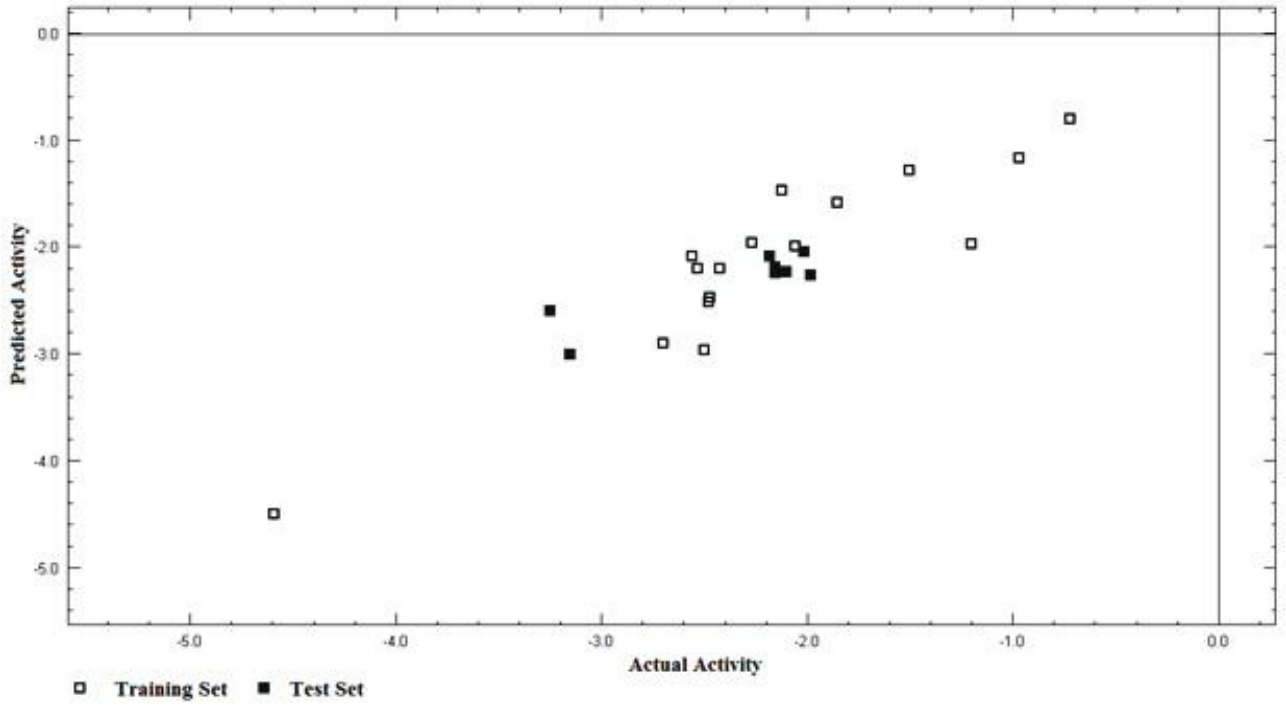


Fig. 3: Plot between actual and predicted value using FFNN analysis

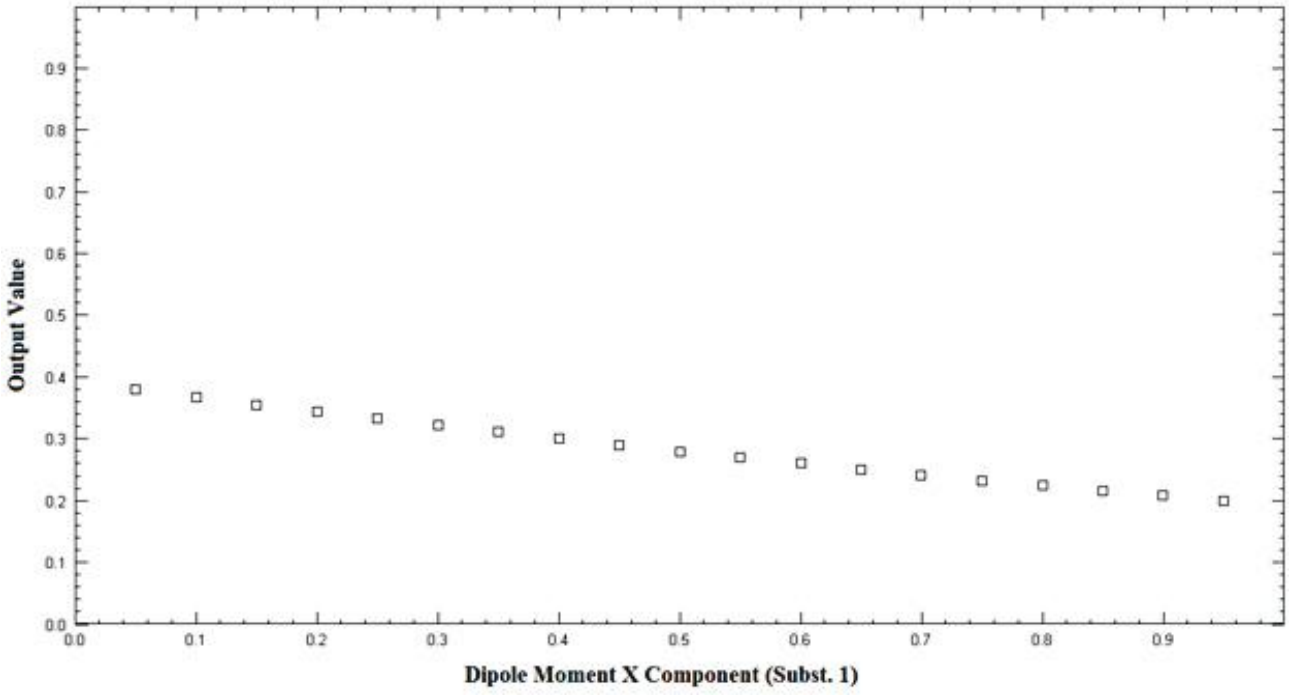


Fig. 4: Plot between biological activity and dipole moment (subst. 1)

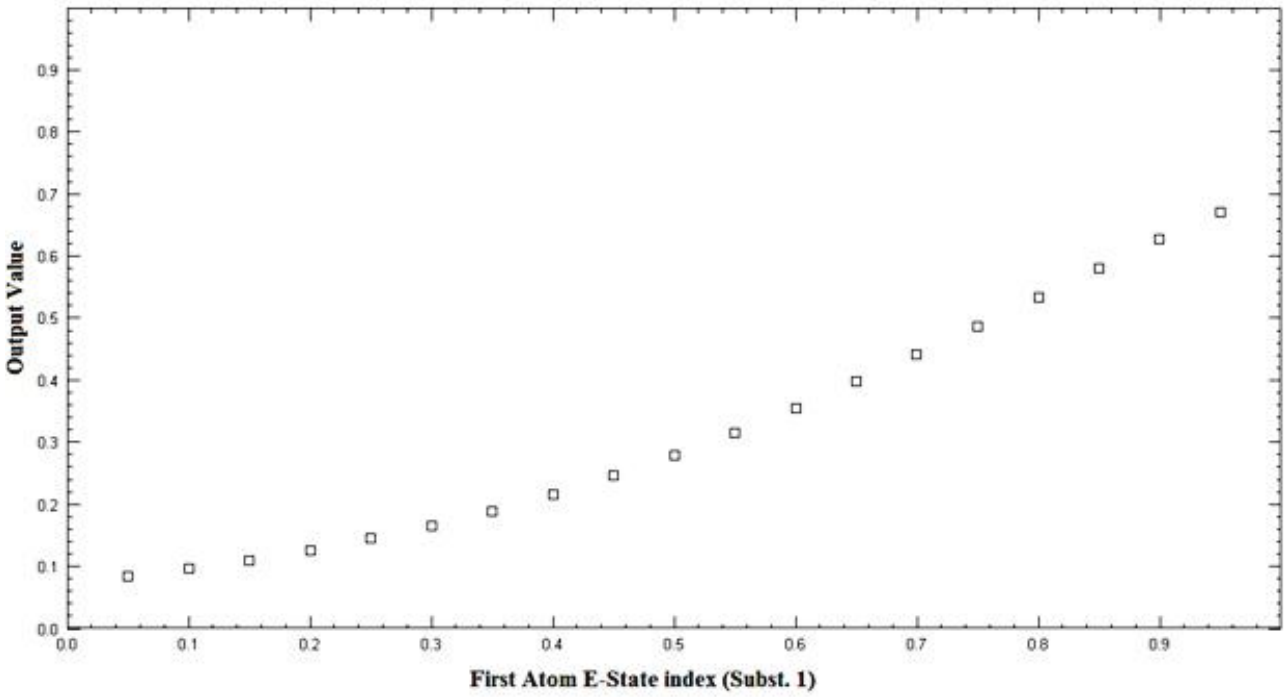


Fig. 5: Plot between biological activity and first atom E-state index (subst.1)

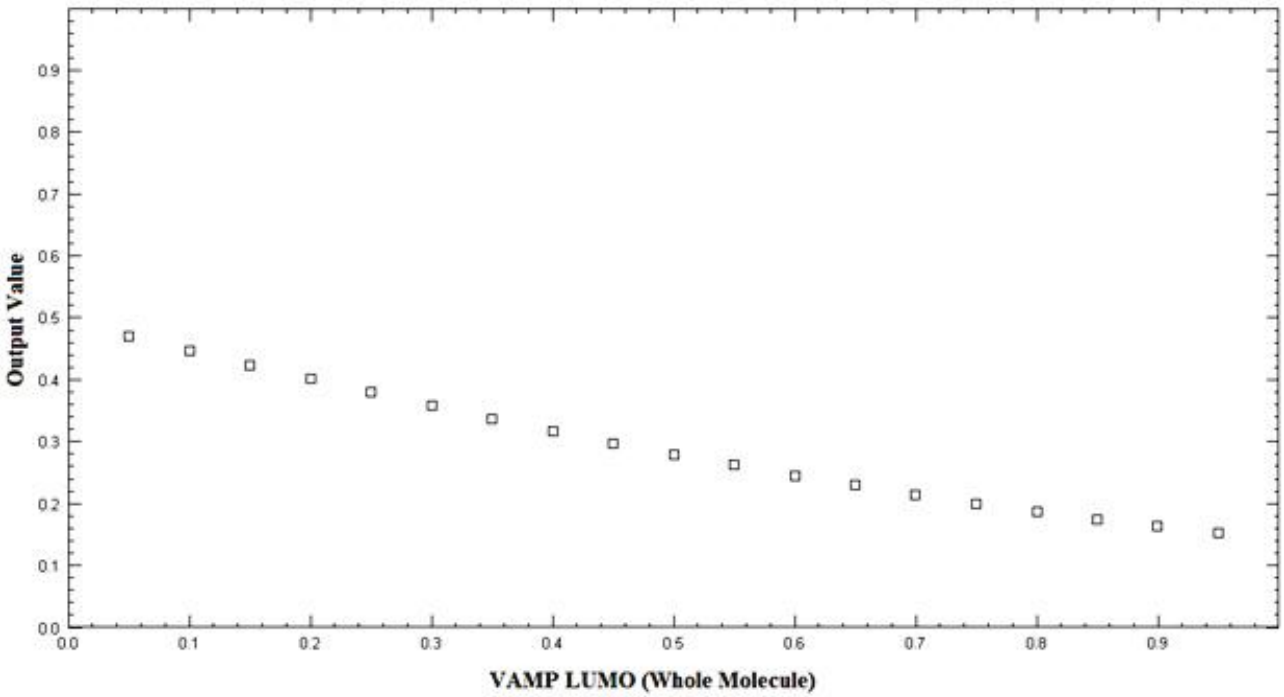


Fig. 6: Plot between biological activity and VAMP LUMO (whole molecule)

Interpretation of descriptors entered

Dipole moment X component (subst.1)

The dipole descriptor is a 3D electronic descriptor that indicates the strength and orientation behavior of a molecule in an electrostatic field. Dipole moment X component (Subst. 1) is negatively correlated with the biological activity as evident from linear regression analysis. Dependency plot of neural network analysis also shows negative correlation (fig. 4). So with decrease in dipole moment of molecule there will be increase in biological activity.

First atom E-state index (subst.1)

Another important descriptor entering the model was first atom E-state index (subst.1). It is a type of electro topological descriptor. The E-State has been established as a composite index encoding both electronic and steric properties. First atom E-state index is positively correlated with biological activity, so at subst. 1 of whole molecule with groups that increase its value will lead to an increase in biological activity of lead molecule. Dependency plot of FFNN analysis also shows positive correlation (fig. 5). So with an increase in value of first atom E-state index in molecule there will be increase in biological activity.

VAMP LUMO (whole molecule)

The LUMO (lowest unoccupied molecular orbital) descriptor adds the energy (in electron volts) of the LUMO for each model. It is important in governing molecular reactivity and properties. The LUMO descriptor is a measure of electrophilicity of a molecule. Negative correlation of Vamp LUMO (whole molecule) with biological activity clearly explains that the groups with low LUMO are required for good biological activity. In the FFNN dependence graphs of the triparametric model (Fig. 6), log IC_{50} decreases with the increase in values of the Vamp LUMO. The deceasing trend is consistent with MLR and PLS models.

CONCLUSION

The MLR, PLS and FFNN were employed to study the anti-malarial activity of 4-aminoquinoline derivatives. Highly predictive QSAR models were obtained using the MLR, PLS and FFNN. All the models were validated using external test set of eight compounds. All three different statistical approaches (MLR, PLS and FFNN) generated comparable results. The findings of present study will certainly aid in the design of more potent anti-malarial agents with improved activity and reduced mechanism based side effects of traditional anti-malarial agents.

ACKNOWLEDGEMENT

We gratefully acknowledge support for this research from ICMR, New Delhi and the Vice Chancellor of Banasthali University for extending all the necessary facilities.

REFERENCES

Agarwal A., Mishra R. and Paliwal S. (2012): A QSAR Study Investigating the Potential Anti-Leishmanial Activity of Cationic 2-Phenylbenzofurans. In: Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Springer-Verlag Berlin Heidelberg, New York, pp 137-141.

Cramer R.D. (1993): Partial Least Squares (PLS): its strengths and limitations. *Perspect Drug Discov Des*. **1**, 269–278.

Friebolin W. and Jannack B. (2008): Antimalarial Dual Drugs Based on Potent Inhibitors of Glutathione Reductase from *Plasmodium falciparum.* J. Med. Chem. **51**, 1260–1277.

Hawkins D.M. and Mills D. (2003): Assessing model fit by cross validation. *J. Chem. Inf. Comp. Sci*. **43**, 579–586.

Konovalov A.D. (2008): Robust Cross-Validation of Linear Regression QSAR Models. *J. Chem. Inf. Mod*. **48**, 2081–2094.

Kramer C. (2009): Sharpening the toolbox of computational chemistry: A new approximation of critical F-values for multiple linear regression. *J. Chem. Inf. Mod.* **49**, 28–34.

Mishra R., Agarwal A. and Paliwal S. (2012): QSAR Analysis of Anti-Toxoplasma Agents. In: Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Springer-Verlag Berlin Heidelberg, New York, pp 131-135.

Neaz, M.M. (2008): 2D-QSAR of non benzodiazepines to benzodiazepines receptor (BZR). *Med. Chem. Res*. **18**, 98-111.

Nosten F. (1982). Mefloquine- resistant *falciparum* malaria on the Thai-Burmes border. *Lancet*, **337**, 1140-1143.

Paliwal S., Narayan A. and Paliwal S. (2009): Quantitative structure activity relationship analysis of dicationic diphenylisoxazole as potent anti-trypanosomal agents. *QSAR Comb Sci.* **28**, 1367–1375.

Singh N. (2006). Drug resistance mechanisms in clinical isolates of *Leishmania donovani*. *Indian J. Med. Res.* **123**, 411- 422.

Smith N. (1987): ABC of AIDS. Tumours. *British Medical Journal.* **294**, 1274-1277.