Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2012, 4(1):465-473

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

2D QSAR of novel 2, N^6-disubstituted 1,2-dihydro-1,3,5-triazine-4,6-diamines as potential antimalarials

Pritam N. Dube*, Arun Kashid, Pranali Alkutkar, Prasanna Datar and Prashant Gawade

Department of Pharmaceutical Chemistry, Sinhgad Institute of Pharmacy, Narhe, Pune

---

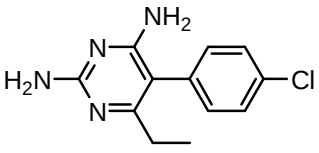
ABSTRACT

The quantitative structure–activity relationship (QSAR) analyses were carried out for a series 2, N^6-disubstituted 1,2-dihydro-1,3,5-trizine-4,6-diamines. The nature of the substituent(s) on C-2; the nature of the substituent(s) on the distal aryl ring; as well as the nature and length of the flexible tether between the rings, to find out the structural requirements of their antimalarial activities against cycloguanil resistant (FCR-3) Plasmodium falciparum strain and sensitive to pyrimethamine. The statistically significant best 2D QSAR models for FCR-3, having correlation coefficient (r^2) = 0.9821 and cross validated squared correlation coefficient (q^2) = 0.6471 were developed by multiple linear regression stepwise (SW–MLR) forward algorithm. The results of the present study may be useful on the designing of more potent analogues as antimalarial agents.

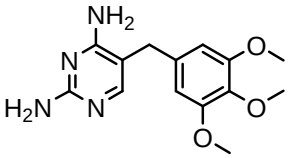
Keywords: QSAR; 2, N^6-disubstituted 1,2-dihydro-1,3,5-trizine-4,6-diamines; Antimalarials; SW-MLR; plasmodium falciparum strain.

INTRODUCTION

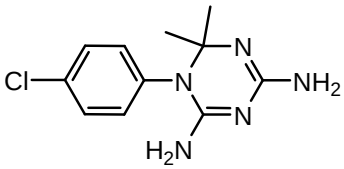
Malaria is one of the most widespread diseases in the world. According to WHO estimates 40% of the world's populations presently live under malarial threat [1]. Around 300 and 500 million cases of malaria occur annually, leading to 1-3 million deaths [1]. Its control is globally a high priority task. 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, undergoes the urgency and continuous need for the discovery of new therapeutics. 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 [2].



1.Pyrimethamine



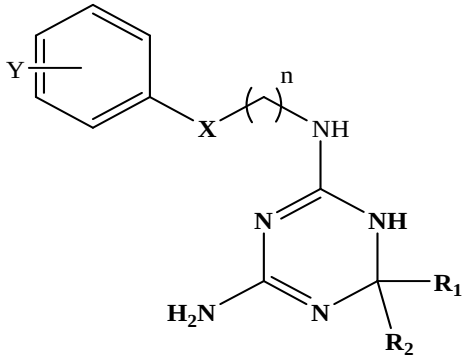
2.Trimethoprim



3.Cycloguanil

Pyrimethamine, trimethoprim and cycloguanil inhibit malarial dihydrofolate reductase (DHFR), one of the few well-defined, validated targets in malarial chemotherapy [3]. These antimalarials inhibit DHFR by competing with the natural substrate dihydrofolic acid. Unfortunately, point mutations at certain amino acid residues surrounding the P.

falciparum DHFR active site have resulted in resistance, compromising the clinical effectiveness of pyrimethamine and cycloguanil [4,5]. Despite this, the folate pathway remains a good target for malarial chemotherapy because the enzyme is limited in its mutational capability, owing to loss in enzyme function [6-8].



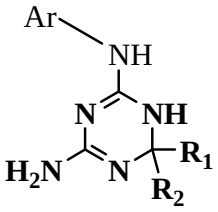
A series of cycloguanil-like compounds that possess a flexible tether interpolated between the 1,2-dihydro- 1,3,5-triazine-4,6-diamine heterocycle and the substituted phenyl ring [9]. So, rather than identifying new molecules for efficacy, modified 1,2-dihydro- 1,3,5-triazine-4,6-diamine heterocycle having many advantages and efficiency are now in priority for antimalarial chemotherapy.

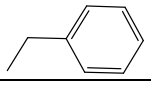
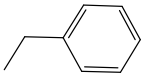
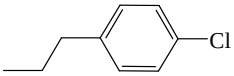
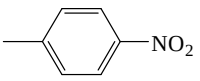
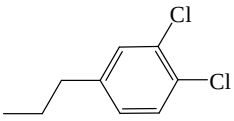
EXPERIMENTAL SECTION

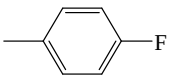
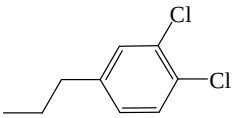
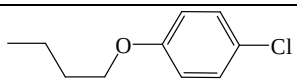
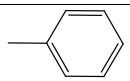
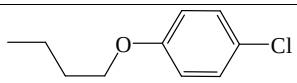
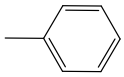
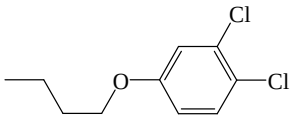
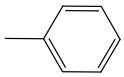
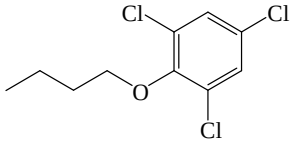
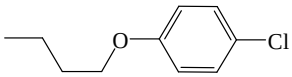
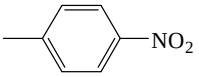
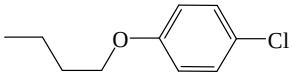
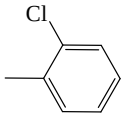
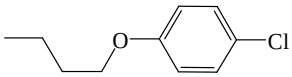
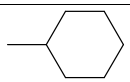
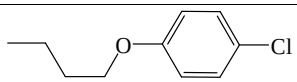
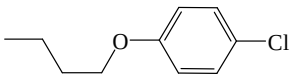
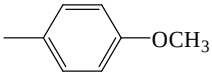
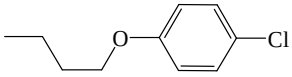
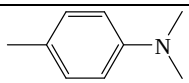
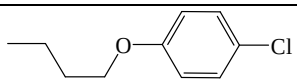
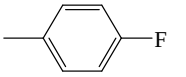
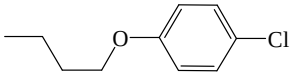
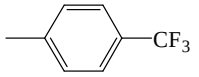
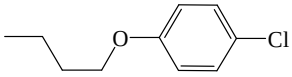
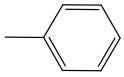
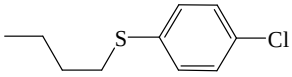
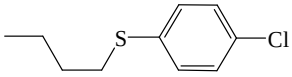
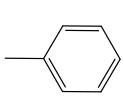
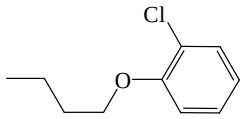
2.1. Data set

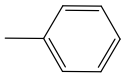
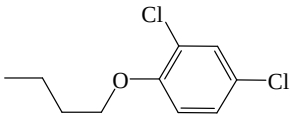
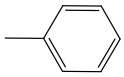
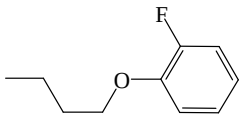
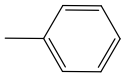
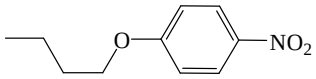
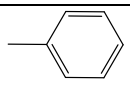
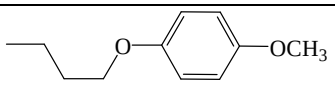
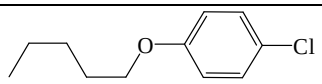
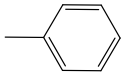
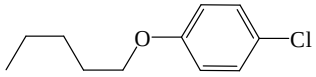
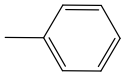
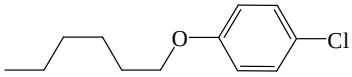
A data set of 28 compounds of side chain modified 1,2-dihydro- 1,3,5-triazine-4,6-diamine heterocycle for antimalarial activities against pyrimethamine and cycloguanil sensitive and resistant (FCR-3) P. falciparum strains was used for the present 2D QSAR study [10]. There is high structural diversity and a sufficient range of the biological activity in the selected series of these derivatives (Table 1). It insists as to select these series of compounds for our QSAR studies. The biological activity values [IC_{50} (nM)] reported in literature were converted to their molar units and then further to negative logarithmic scale (pIC_{50}) and subsequently used as the dependent variable for the QSAR analysis.

Table No.1 Structures and antimalarial activity of 2,N^6-disubstituted 1,2-dihydro-1,3,5-triazine-4,6-diamines



Comp. No.	R_1	R_2	Ar	IC_{50}
1	$-CH_3$	$-CH_3$		49.75
2				9.89
3	$-CH_3$	$-CH_3$		16.55
4	-H			11.12

5	-H			8.41
6	$-CH_3$	$-CH_3$		11.33
7	-H			7.11
8	-H			6.73
9	-H			4.32
10	-H			0.99
11	-H			7.39
12	-H			6.66
13	-H			4.59
14				7.94
15	-H			6.66
16	-H			44.75
17	-H			6.21
18	-H			5.54
19	-H			1.30
20	$-CH_3$	$-CH_3$		7.84
21	-H			4.53

22	-H			4.52
23	-H			10.85
24	-H			3.87
25	-H			12.31
26	$-CH_3$	$-CH_3$		7.20
27	-H			2.71
28	-H			1.36

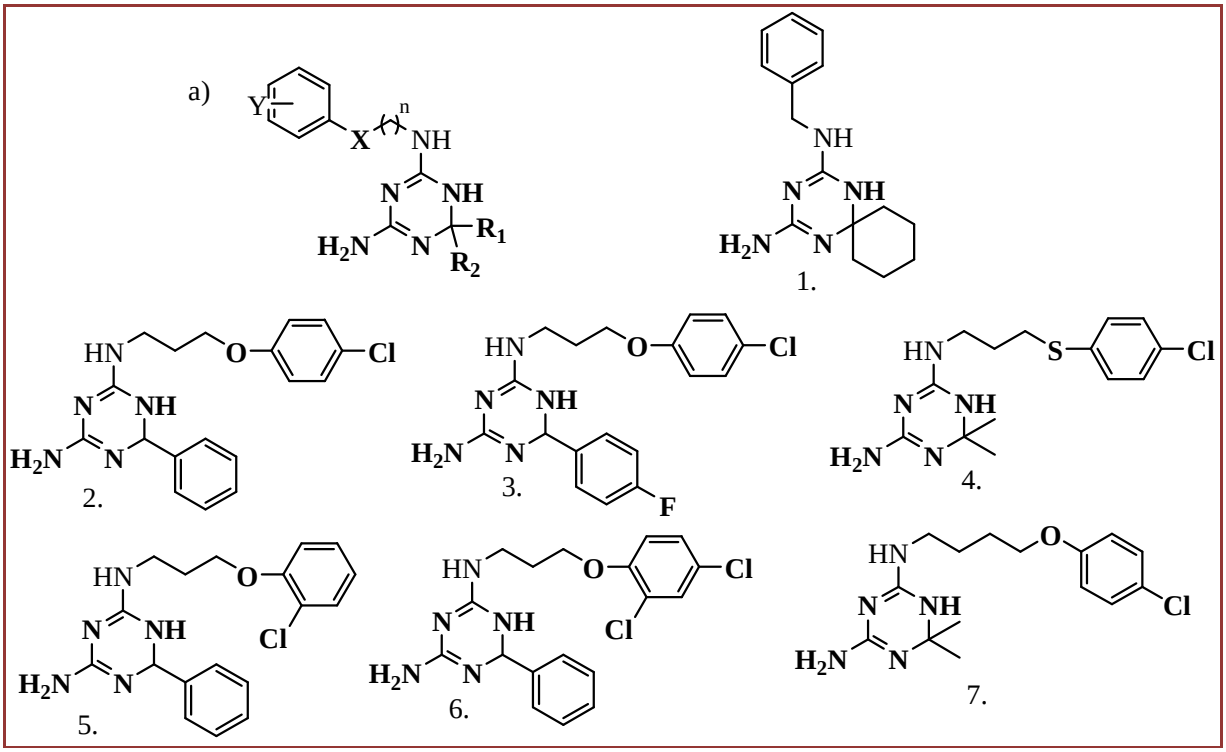
All 28 compounds were built on workspace of molecular modeling software VLife MDS 3.5 (Vlife Sciences Technologies Pvt. Ltd. Pune, India) and then the structure was converted to three-dimensional space for further analysis. All molecules were batch optimized for the minimization of energies using Universal force field (UFF) followed by considering distance-dependent dielectric constant of 1.0, convergence criterion or root-mean-square (RMS) gradient at 0.01 kcal/mol A˚ and the interaction limit to 10,000 [11]. 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-impirical, Hydrophillic–hydrophobic 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 289 descriptors to be used for QSAR analysis.

Structures of Test set compounds

The manual data selection method [12-15] was adopted for division of training and test data set comprising of 21 and 7 molecules, respectively. Seven compounds, namely 1.1-1.7 were used as test set while the remaining molecules were used as the training set by considering chemical variation. The unicolumn statistics of the training and test sets is reported in Table 2.

2.2. 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. Among several search algorithms, stepwise (SW) forward–backward variable selection method, genetic algorithms (GA) and simulated annealing (SA) based feature selection procedures are most popular for building QSAR models and can explain the situation more effectively [14-16].



In the selected 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 auto scaling in which the number of random iterations was set at 100. In SW forward–backward variable selection algorithm, 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). In GA method, population and number of generations were set as 10 and 1000, respectively and speed of 9999.

2.3. Model quality and validation

The developed QSAR models are evaluated using the following statistical measures: n, (the number of compounds in regression); k, (number of variables); DF, (degree of freedom); optimum component, (number of optimum PLS components in the model); r^2, (the squared correlation coefficient); r^2se, (standard error of squared correlation coefficient); F test, (Fischer's value) for statistical significance; q^2, (cross-validated correlation coefficient); q^2_se, (standard error of cross-validated square correlation co-efficient); pred_r^2, (r^2 for external test set); pred_r^2se, (standard error of predicted squared regression); Z score, (Z score calculated by the randomization test); best_ran_q^2, (highest q^2 value in the randomization test); best_ran_r^2, (highest r^2 value in the randomization test). 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 [13]. 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^2se) shows absolute quality of fitness of the model.

Internal validation was carried out using 'leave-one-out' (q^2, LOO) method [17]. The cross-validated coefficient, q^2, was calculated using the following equation:

$$q2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2}$$

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 pIC50 value of the nine 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.

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

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

RESULTS AND DISCUSSION

The QSAR study of 28 new side chain modified 1,2-dihydro- 1,3,5-triazine-4,6-diamine heterocycle derivatives for antimalarial activities (Table 1) through MLR methodology, based on various feature selection methods viz. SW using VLife MDS 3.5 software resulted in the following statistically significant models (Table 2), considering the term selection criterion as r^2, q^2 and pred_r^2. The training and test sets were selected by manual selection 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. A Uni-column statistics for training set and test set was generated to check correctness of selection criteria for trainings and test set molecules.

Table 2: Developed 2D-QSAR Model

Size	Equation	r^2	q^2	F test	r^2 se	q^2 se
Training Set Size = 20 Test Set Size = 8	Activity = + 0.2708(± 0.0122) T_N_O_5 + 0.5878(± 0.0400) SulfursCount - 0.0305(± 0.0032) SsCH3E-index - 0.0339(± 0.0087) T_C_N_4 - 0.0893(± 0.0400) T_O_F_3 + 0.4860	0.9821	0.6471	153.9562	0.0386	0.1714

Model 1 (SW–MLR)
Activity = + 0.2708(± 0.0122) T_N_O_5 + 0.5878(± 0.0400) SulfursCount - 0.0305(± 0.0032) SsCH3E-index - 0.0339(± 0.0087) T_C_N_4 - 0.0893(± 0.0400) T_O_F_3 + 0.4860.

The statistically best model (Model 1) for antimalarial activity against FCR-3 with a coefficient of determination (r^2) =0.9821 was considered.

Prediction of activity:
The generated equation was used to predict the activity of test set as compare to actual activity with respect to various physicochemical parameters. The MLR model shows better results as compare to two other models (PCR and PLR). Therefore the activity predicted by MLR is only considered in this study (Table 3).

Table 3: Predicted Activity of Compounds (Test set) by MLR

Test Compound No.	Antimalarial Activity	Predicted	Difference
02	0.1011	0.079829	0.02127
11	0.1353	0.147532	-0.0122
17	0.1610	0.181383	-0.0203
21	0.2207	0.181383	0.03931
22	0.2212	0.181383	0.0398
26	0.1388	0.333176	-0.1943

Multiple linear regression (MLR), Principal component regression (PCR) and Partial Least Squares regression (PLR) were carried out to find out the factors responsible for the biological activity (Table 3). Contribution chart, (% contributions of different descriptors in Model 1 / Equation 1) representing the contribution of descriptors in the 2D-QSAR model developed by MLR is shown in Fig 1.

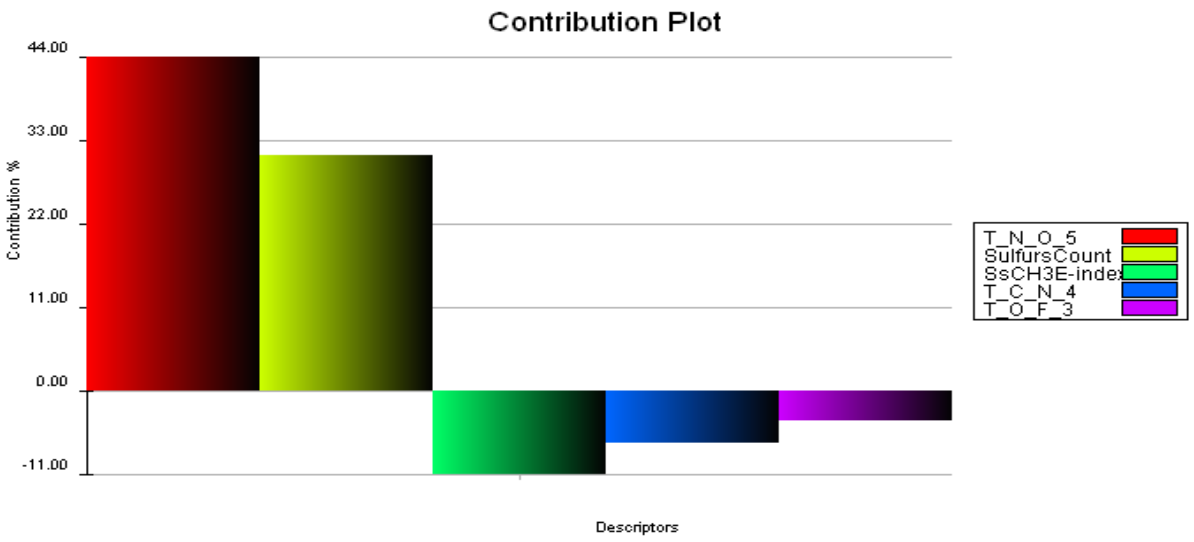


Figure1: Contribution plot of various descriptors.

Descriptors description:

From any other Oxygen atom (single double or triple bonded) by distance of 5 bonds in a T_N_O_5 is the count of number of Nitrogen atoms (single double or triple bonded) separated molecule.

T_C_N_4 is the count of number of Carbon atoms (single double or triple bonded) separated from any other Nitrogen atom (single double or triple bonded) by distance of 4 bonds in a molecule.

T_O_F_3 is the count of number of Oxygen atoms (single double or triple bonded) separated from any other Fluorine atom (single double or triple bonded) by distance of 3 bonds in a molecule.

SsCH3E-index: Electrotopological state indices for number of -CH3 group connected with one single bond.

Sulferscount is the number of sulfur atoms presents in a molecule.

T_N_O_5 and Sulfurs Count are directly proportional while T_O_F_3, T_C_N_4, SsCH3E-index are inversely proportional to the antimalarial activity. T_N_O_5 resembles optimal chain length at N^6 substitution. T_O_F_3 resembles the optimal position of Fluorine at aryl substitution. T_C_N_4 resembles triazine ring. SsCH3E-index resembles electrotopology of number of CH_3 groups attached at C-2 position. Sulfurs Count denotes the presence of sulfur at N^6 chain is beneficial for activity.

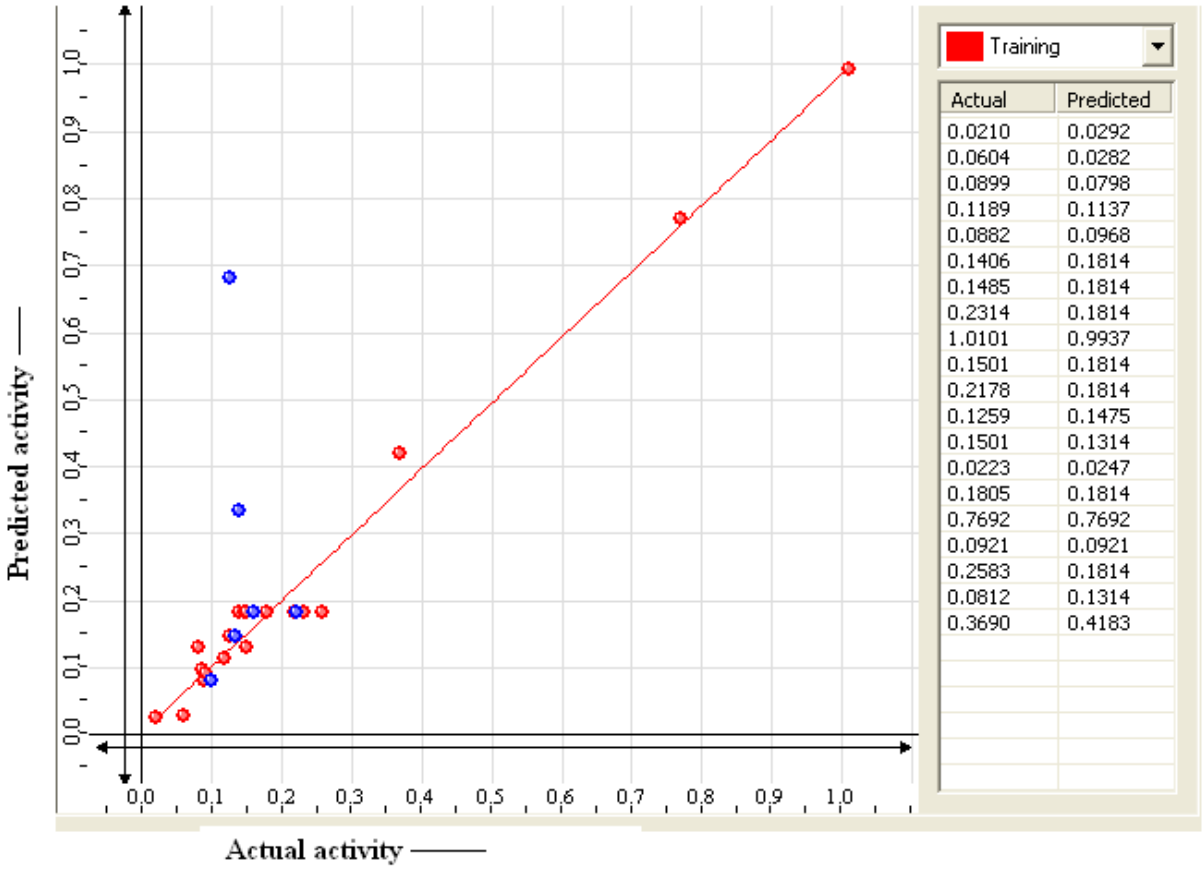


Figure-2: Correlation plots of observed and predicted activities of the training and test compounds for best QSAR Model.

Table 4: Percent Contribution of Descriptors.

Sr.No.	Parameter	Result
1.	T_N_O_5	44.98%
2.	SulfursCount	31.38%
3.	SsCH3E-index	-11.70%
4.	T_C_N_4	-7.17%
5.	T_O_F_3	-4.77%

CONCLUSION

The present work shows how a set of antimalarial activities of various 1,2-dihydro- 1,3,5-triazine-4,6-diamine heterocycle 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 awareness and understanding of the descriptors involved in antimalarial activity of these compounds could provide a great opportunity for the ligand structures design with appropriate features, and for the explanation of the way in which these features affect the biological data upon binding to the respective receptor target. 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 software.

REFERENCES

[1] WHO, World Malaria Report, **2011**, Chapter 1-4, 1-34.
[2] N.K. Sahu; S. Sahu; D.V. Kohli. *Chem. Biol. Drug Des*., **2008**, 71, 287–297.
[3] G. Rastelli; W. Sirawaraporn; P. Sompornpisut; T. Vilaivan; S. Kamchonwongpaisan; R. Quarrell; G. Lowe; Y. Thebtaranonth; Y. Yuthavong. *Bioorg. Med. Chem.*, **2000**, 8, 1117-1128.
[4] Y. Yuthavong. *Microbes Infect.* **2002**, 4, 175-182.
[5] A. Gregson; C.V. Plowe. *Pharmacol. Rev*., **2005**, 57, 117–140.
[6] S.R. Meshnick; M.J. Dobson. Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery, Humana Press, Totowa, NJ, **2001**, 15-25.
[7] R.S. Philips. *Clin. Microbiol. Rev*. **2001**, 14, 208–226.
[8] N.J. White, *J. Clin. Infect*. **2004**, 113, 1084–1092.
[9] D. Gravestock; A. Rousseau; L. Lourens; S. S. Moleele; R. L. van Zyl; P. A. Steenkamp. *Eur J Med Chem*., **2011**, 46, 2022-2030.
[10] K. Yearick; K. Ekoue-Kovi; D. P. Iwaniuk; J. K. Natarajan; J. Alumasa; A. C. de Dios; P.D. Roepe; C. Wolf. *J. Med. Chem*., **2008**, 51, 1995.
[11] T. A. Halgren. *J. Comput. Chem*., **1996**, 17, 553–586.
[12] B. D. Hudson; R. M. Hyde; E. Rahr; J. Wood. *Quant. Struct. Act. Rel*., **1996**, 15, 285–289.
[13] A. Golbraikh; A. Tropsha; *J. Mol. Graph. Model*., **2002**, 20, 269–276.
[14] W. Zheng; A. Tropsha. *J. Chem. Inf. Comput. Sci*., **2000**. 40, 185–194.
[15] A. Coi; F. L. Fiamingo; O. Livi. *Bioorg. Med. Chem*., **2009**, 17, 319– 325.
[16] K. Hasegawa; T. Kimura; K. Funatsu. *Quant. Struct.Act. Rel*., **1999**. 18, 262–272.
[17] R. D. Cramer; D. E. Patterson; J. D. Bunce. *J. Am. Chem. Soc.*, **1988**. 110, 5959–5967.
[18] VLife MDS 3.5, **2011**, Molecular design suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India.