International Standard Serial Number (ISSN): 2249-6807

International Journal of Institutional Pharmacy and Life Sciences 7(6): November-December 2017

INTERNATIONAL JOURNAL OF INSTITUTIONAL PHARMACY AND LIFE SCIENCES

Pharmaceutical Sciences

Research Article……!!!

Received: 09-08-2017; Revised: 30-11-2017; Accepted: 02-12-2017

IN-SILICO MODELING OF DHFR INHIBITORS AS POTENT ANTI-PNEUMOCYSTIS AGENTS

Ruchi Mishra[1*], Ankita Agarwal[2], Shalini Gupta[1]

[1]Department of Chemistry, Khandelwal College of Management Science and Technology, Bareilly

[2]Department of Chemistry, FVFOS, Swami Vivekanand Subharthi University, Meerut

Keywords:

QSAR, TSAR, MLR, PLS, FFNN

For Correspondence:

Ruchi Mishra

Department of Chemistry, Khandelwal College of Management Science and Technology, Bareilly

ABSTRACT

The QSAR study was conducted on 29 pyrimidine and quinazoline analogues with anti-pneumocystis activity using multiple linear regression (MLR), partial least squares (PLS) and feed forward neural network (FFNN) techniques methods. The statistical values from both the techniques were analyzed and compared to establish the good predictability of the models obtained. The MLR and PLS generated comparable models with good predictive ability and all other statistical values, r, r^2, r^2_{cv}, r^2 (test set), F and S values, were 0.94, 0.89, 0.86, 0.84 and 36.04, 0.41 respectively, for MLR and r^2, r^2_{cv}, r^2 (test set) and statistical significance value were 0.89, 0.86, 0.83 and 0.99 respectively, for PLS, were satisfactory. The model developed from feed forward neural network (FFNN) technique also showed good correlation value of r^2= 0.87. The analysis helped to ascertain the role of Molecular volume, Lipole Y component, Lipole Z component and Kappa 1 index for whole molecule in determining the *Pneumocystis carinii* DHFR inhibitory activity of the compounds.

INTRODUCTION

Pneumocystis carinii pneumonia is one of the premier causes of morbidity and mortality in patients with acquired immunodeficiency syndrome (AIDS) and also affects patients with other immune disorders. *Pneumocystis carinii* is a eukaryotic microorganism that is found worldwide. Its host range is wide and includes humans and other mammals. The metabolism of folate plays an important role in the biosynthesis of nucleic acid precursors. During the synthesis of purines and thymidylate, the cofactor tetrahydrofolate is oxidised to 7, 8-dihydrofolate and subsequently converted back to tetrahydrofolate by the enzyme dihydrofolate reductase (DHFR). The inhibition of DHFR causes the depletion of tetrahydrofolate and disrupts DNA synthesis, leading to cell death. For this reason, DHFR inhibitors such as methotrexate have been used as antitumor, anti-bacterial and antiprotozoan agents. Because MTX and other classical antifolates require an active transport mechanism for their uptake, they are not effective for the treatment of infections caused by *Pneumocystis carinii* that lack these mechanisms[1]. Trimetrexate (TMQ) and piritrexim (PTX) are potent lipophilic inhibitors of pcDHFR taken up by passive diffusion, but inhibit mammalian DHFR to a greater extent. This results in toxicity to mammalian tissue and requires that PTX or TMQ be co-administered with leucovorin which is taken up by active transport and protects the host tissue[2]. Treatment with leucovorin is costly and subject to serious side effects that may require interruption of treatment. As such, there is great interest in developing potent and selective inhibitors of *Pneumocystis carinii* DHFR. In the present paper we describe the QSAR studies of inhibition of DHFR in *P. carinii* by 2, 4-diamino-8-deazafolate analogues. The study was aimed to identify the physicochemical parameters associated with DHFR inhibitor which are more selective for *Pneumocystis carinii* than for mammalian DHFR.

MATERIAL AND METHODS

The structures of 29, pyrimidine and quinazoline analogues[3,4] were sketched using Chem Draw software and were imported on TSAR (Version 3.3; Accelrys Inc, oxford, England) software. The generated 3D models of all derivatives created were cleaned up and subjected to charge calculation and energy minimization. More than 300 molecular descriptors were calculated for all the compounds under consideration. TSAR affords the calculation of the following descriptors: atomic attributes (like molecular properties, dipole moment and verloop steric parameters), atomic indices (like shape, connectivity and topological indices) and Vamp

electrostatic properties (total energy, HOMO, LUMO, heat of formation, etc)[5]. To reduce data redundancy, pair wise correlation analysis was carried out [6]. Among the highly intercorrelated descriptors the one that had high correlation with biological activity was kept and other was discarded. This process was repeated number of times and finally three descriptors were retrieved that were highly correlated with biological activity and were not having inter correlation among each other.

Linear Regression Analysis (MLR & PLS): To develop QSAR models, stepwise MLR analysis with leave-one-out (LOO)[7] cross-validation was applied to the training set. The molecules of the series were divided randomly into training set (19 molecules) and test set (10 molecules) with descriptors retrieved by data reduction. Training set was used to build linear models so that an accurate relationship could be found between structure and biological activity[8]. The test set of twelve molecules was not used to develop the regression model but served to check the predictive power of the developed model. In addition to MLR, partial least squares (PLS)[9] analysis was also performed to check the predictive ability and robustness of the developed model.

Nonlinear Regression Analysis (FFNN): 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 4-2-1[10]. The results were visualized on a 2D plot of output node against input (dependency plots).

RESULT AND DISCUSSION

Linear Regression Analysis: Multiple linear regression (MLR) and partial least squares (PLS) were used to derive the QSAR equations. The statistically significant model (final 19 molecules in training set) was constructed from the training set by using 4 parameters. The final regression equation obtained from MLR analysis represented as (Equation 1)

$$\mathbf{Y = 0.066 \times X1 - 0.039 \times X2 + 0.155 \times X3 - 0.543 \times X4 - 8.360} \qquad \textbf{\textit{-Equation 1}}$$

This best model was selected on the basis of various statistical parameters such as coefficient of determination (r^2), predictive power of model (r^2_{cv})standard deviation (SD), sequential Fisher test (F) and test for statistical significance (t). The value of r^2 should always be greater than 0.6 (a

good model should have an $r^2 > 0.9$) and the value of r^2_{cv} could fall into three categories: $r^2_{cv} > 0.6$: The model is fairly good, $0.4 < r^2_{cv} < 0.6$: The model is questionable, $r^2_{cv} < 0.4$: The model is poor.

$r = 0.94$, $r^2 = 0.89$, $r^2_{cv} = 0.86$, $F = 36.04$, $S = 0.41$, predictive r^2 for test set = 0.84

PLS analysis was also performed on the same data set to check the soundness of the MLR model. The resulted r^2_{cv} value of 0.86 clearly demonstrates the high predictive ability of the developed PLS model (Equation 2)

$$\mathbf{Y = 0.066 \times X1 - 0.039 \times X2 + 0.155 \times X3 - 0.543 \times X4 - 8.360} \qquad \textbf{\textit{-Equation 2}}$$

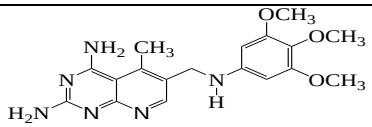
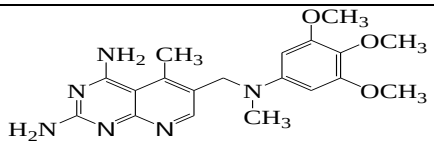
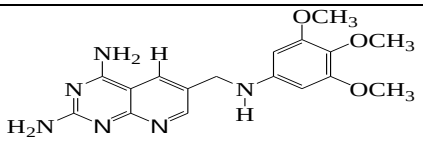
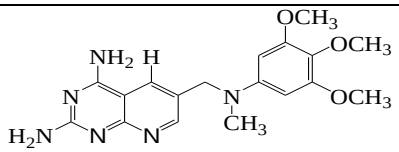
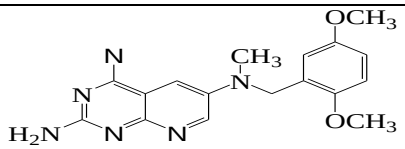
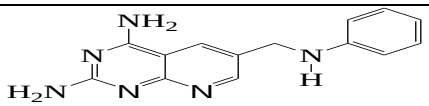
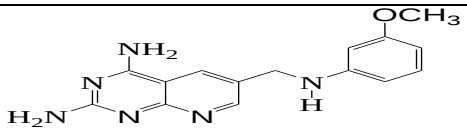
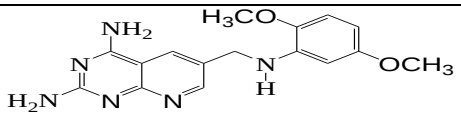
Where X1 is Molecular volume (Whole Molecule), X2 is Lipole Y component (Whole Molecule), X3 is Lipole Z component (Whole Molecule) and X4 is Kappa 1 index Lipole X component (Whole Molecule) in both the equations. Statistical significance = 0.99, $r^2_{cv} = 0.86$, Fraction of variance explained = 0.89, r^2 for test set = 0.83.

Since for a well defined problem, both MLR and PLS should generate comparable result, the r^2_{cv} values of MLR and the PLS models were evaluated and it was found that both models have comparable r^2_{cv} value of 0.86 and 0.86 for MLR and PLS respectively. The predictive ability of the model was also validated using the external test set of 13 compounds in context of minimum difference between the actual and predicted biological activity values of MLR and PLS analysis for training and test which is shown in table 1 and 2 and their respective plots are depicted in figure 1 and 2.

Nonlinear Regression Analysis: The neural network models were used to study the type of relationships between the molecular descriptors and biological data. The results of present study reveal that both the techniques can be used with greater efficiency to develop predictive models, though the data under consideration can vary the statistics of the developed model. Best RMS fit obtained for the model is 0.0588 at 988 cycles. The predictive power was judged from the plot of predicted versus experimental affinities of training and test set of compounds for model illustrated in figure 3 and values were depicted in table 1 and 2. Results were visualized on a 2D plot of output node against input (figure: 5a-d). The input descriptors were the same as used for multivariate regression. The close analysis of all the plots reveals that the relationship between biological activity and four descriptors is linear and analogous to MLR and PLS analysis. $r^2_{training}$=0.87 and r^2_{test}=0.73. Compound no. 2a, 11 (table 1) & 3, 5, 11, 13 (table 2)

showed very close actual and predicted biological activities corresponding graph to them is depicted in figure 4.

Table 1. 2,4-Diamino-6-(arylaminomethyl) pyrido [2,3-d] pyrimidine analogues with their actual & predicted values

Compound's name	Compound structure	Actual activity (Log 1/c)	Predicted activity		
			MLR	PLS	FFNN
1a*		-1.934	-2.358	-2.331	-2.187
1b		-1.113	-1.019	-0.956	-1.185
2a#		-3.176	-3.062	-3.049	-2.960
2b*		-2.380	-2.669	-2.644	-2.503
3*		-1.924	-1.787	-1.768	-1.707
4		-3.346	-3.809	-3.841	-3.834
5		-2.716	-3.347	-3.369	-3.329
6		-3.785	-3.546	-3.559	-3.535

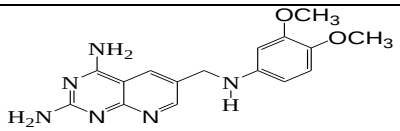
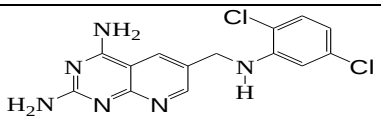
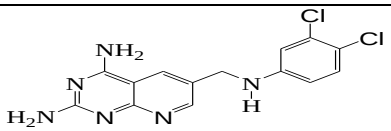
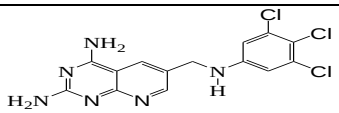
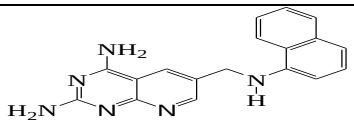
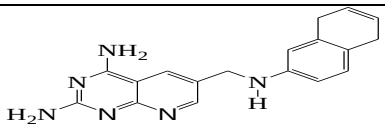
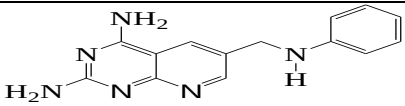
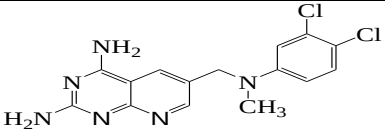
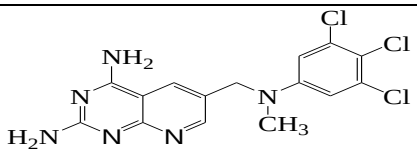
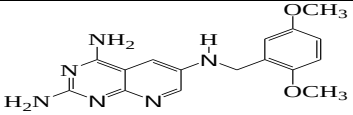
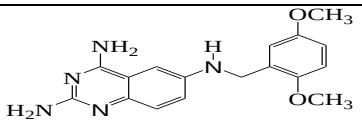
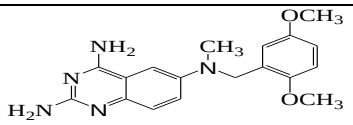
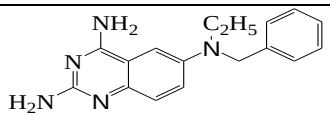
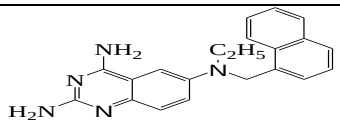
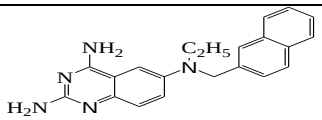
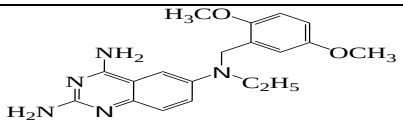
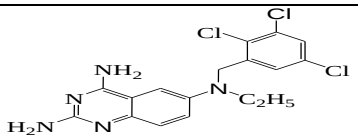
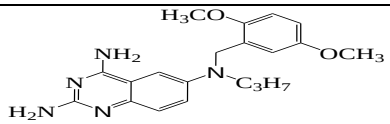
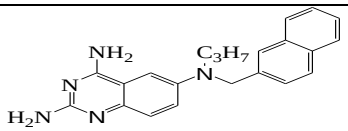
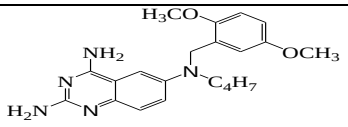
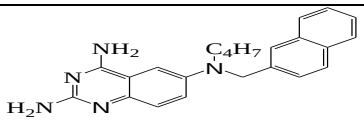
7		-2.633	-3.220	-3.222	-3.160
8		-2.643	-2.933	-2.984	-2.880
9		-2.397	-2.221	-2.273	-2.056
10		-2.544	-1.723	-1.787	-1.621
11#		-1.845	-1.895	-1.820	-1.899
12		-3.301	-3.361	-3.284	-3.491
13		-5	-4.319	-4.277	-4.479
`14*		-2.117	-2.799	-2.806	-2.868
15*		-2.292	-2.936	-2.836	-3.042

Table 2. Series of N9-Substituted 2,4-Diaminoquinazoline analogues with their actual & predicted biological values

Compound's name	Compound structure	Actual activity (Log 1/c)	Predicted activity		
			MLR	PLS	FFNN
1		-3.579	-3.190	-3.190	-3.103

3#		-3.662	-3.472	-3.458	-3.447
4*		-1.939	-1.263	-1.231	-1.325
5#		-1.322	-1.352	-1.332	-1.406
6		-0.397	-0.823	-0.831	-1.101
7		-0.857	-0.697	-0.657	-0.975
8		-0.995	-1.299	-1.285	-1.348
9*		-0.838	-0.029	0.0205	-0.756
10*		-1.579	-0.914	-0.899	-1.114
11*#		-1.431	-1.329	-1.404	-1.404
12*		-1.397	-0.659	-0.640	-0.990
13#		-1.623	-1.594	-1.675	-1.648

*Compounds included in Test set. #Compounds showing very close actual and predicted values.

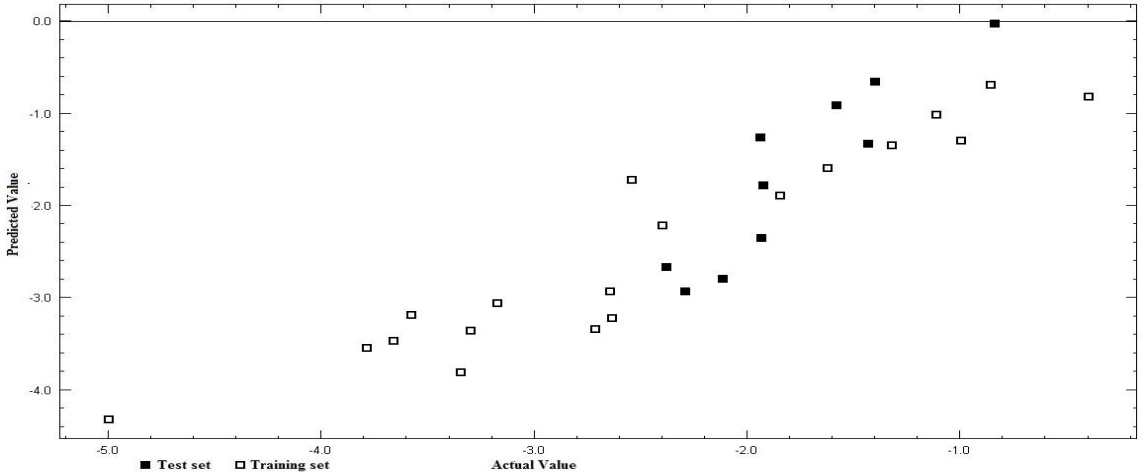


Fig. 1

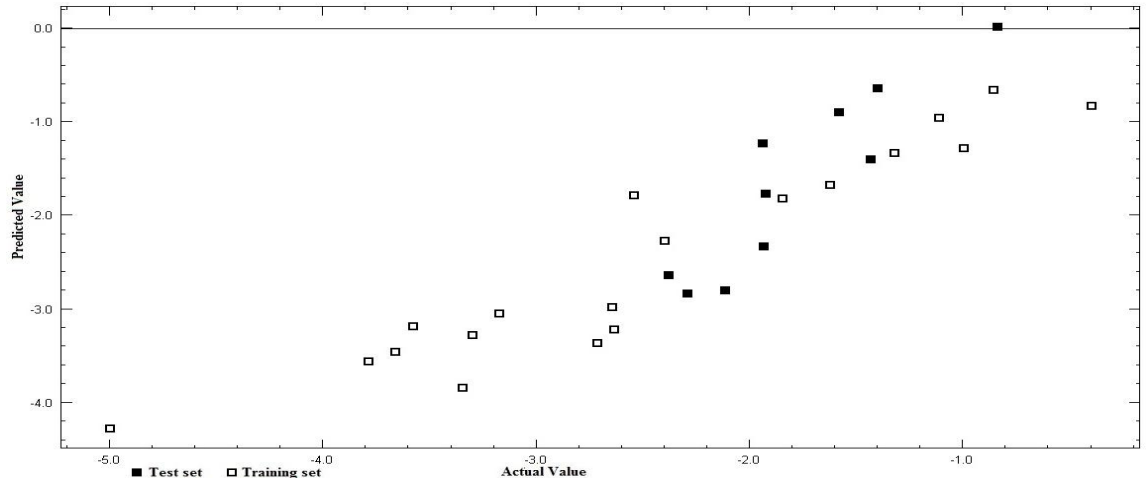


Fig. 2

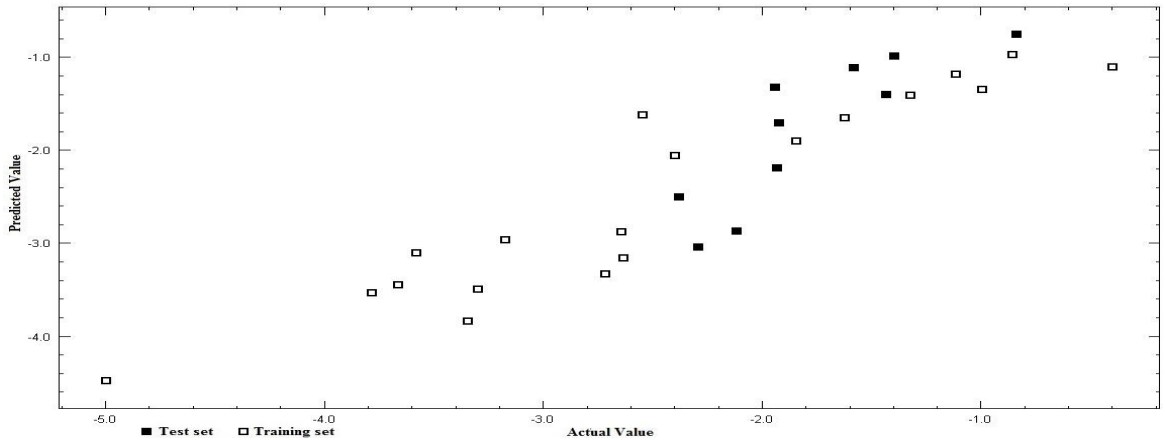


Fig. 3

Actual vs. predicted activity for the training and test set of compounds derived from MLR **(Fig. 1)**, PLS **(Fig. 2)** and FFNN **(Fig. 3)** analysis.

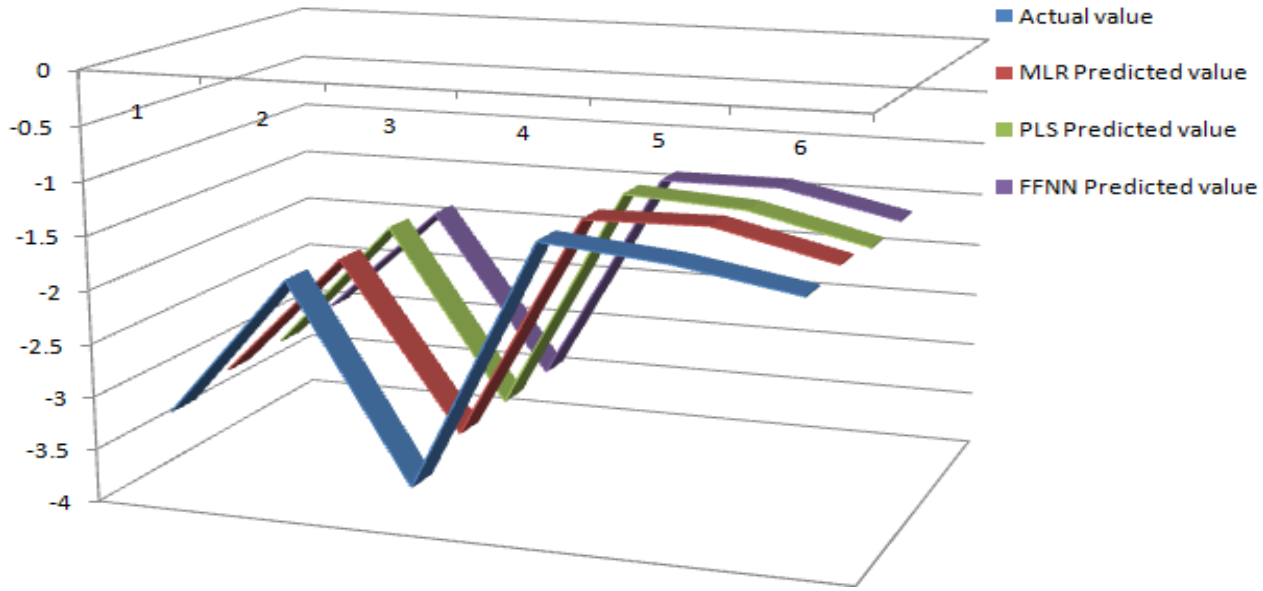
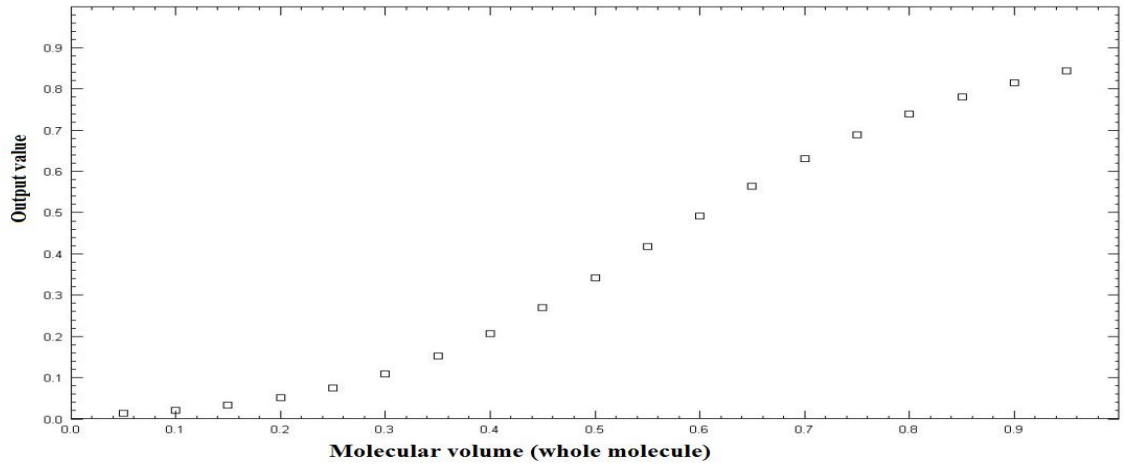
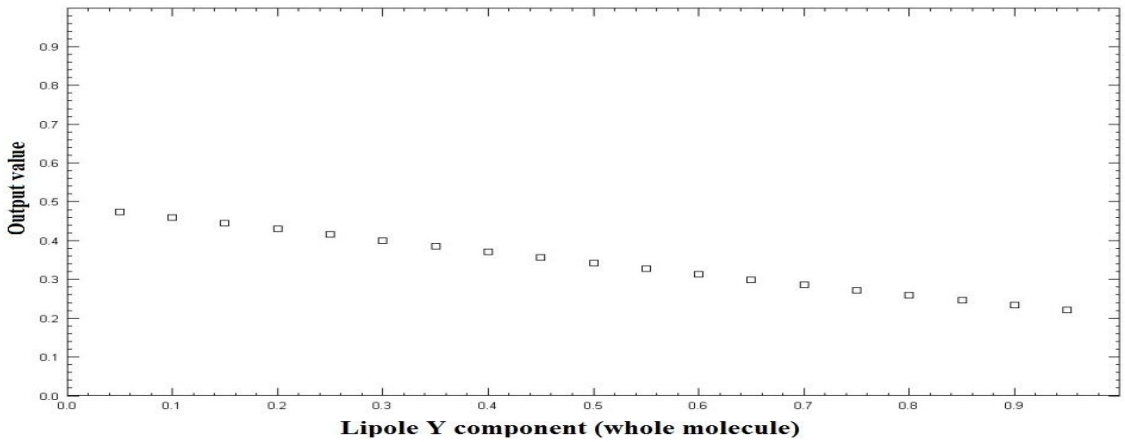


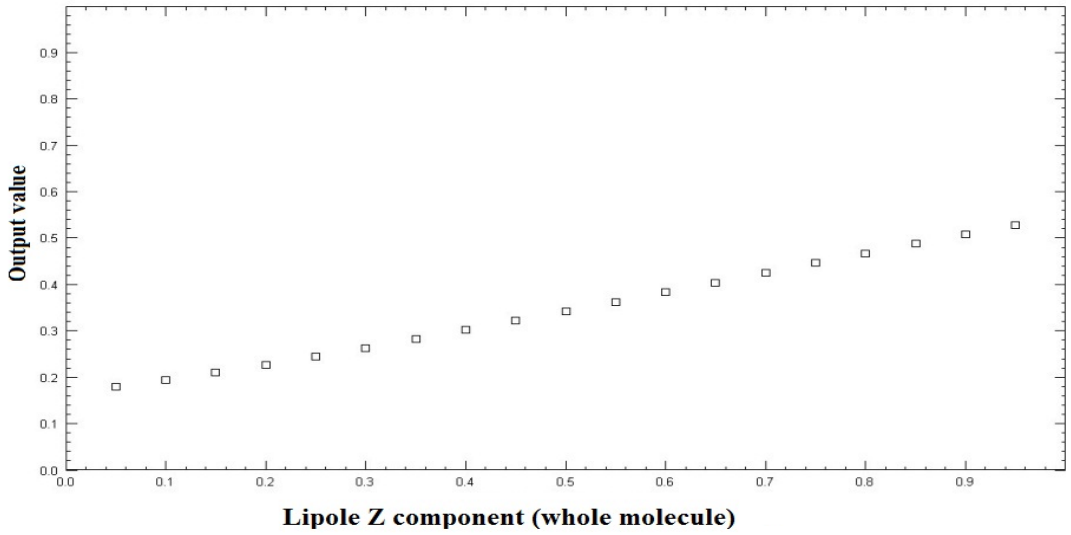
Fig. 4 Graph showing comparison between actual and predicted activity for compounds (2a, 11from **table 1** & 3, 5, 11, 13 from **table 2**) derived from MLR, PLS and FFNN analysis.



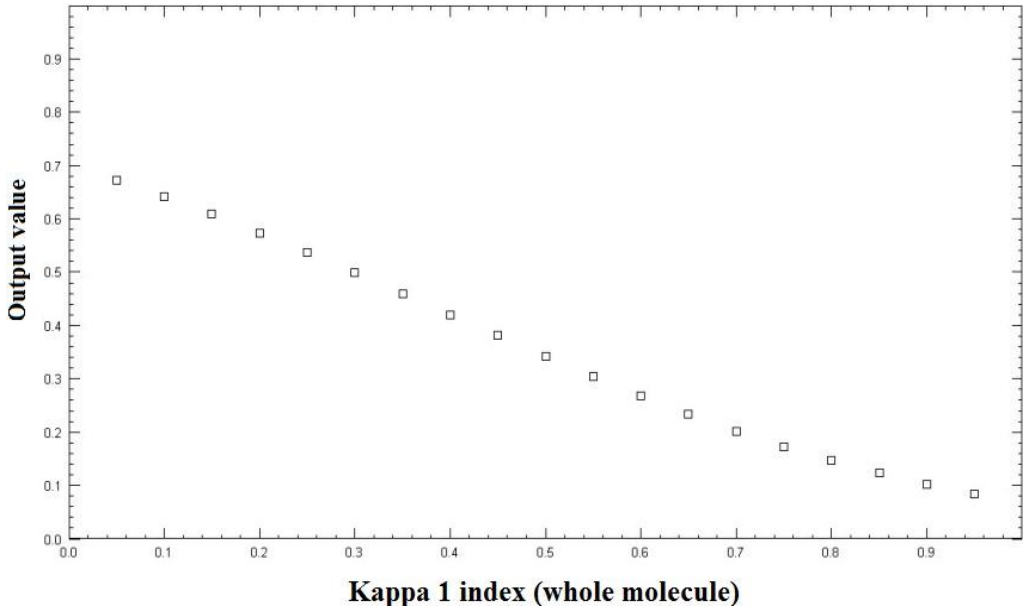
5a



5b



5c



5d

Figure 5a-5d. Dependency plots between biological activity and physicochemical parameters

CONCLUSION: The MLR, PLS and FFNN were employed to study *P. crainii* DHFR inhibitory activity of pyrimidine and qunazoline derivatives. Highly predictive QSAR models were obtained using MLR, PLS and FFNN. All models were validated using external test set of 10 compounds. All three different statistical approaches (MLR, PLS, FFNN) generated nearly same results for each QSAR model and allow us to estimate additionally the quality of prediction. Findings of present study will certainly aid in the design of more potent *P. crainii* DHFR inhibitors with improved activity and reduced mechanism based side effects of traditional *P. crainii* DHFR inhibitors.

ACKNOWLEDGEMENT: We gratefully acknowledge support for this research from Indian Council of Medical Research (ICMR), New Delhi.

REFERENCES

1. Behbahani, R. & Moshfeghi M. (1995). Therapeutic approaches for AIDS-related toxoplasmosis. *Annals of Pharmacotherapy, 29*, 760-768.
2. Kovacs, J.A. & Allegra, C.J. (1989). Characterization of de novo folate synthesis in *Pneumocystis carinii* and *Toxoplasma gondii*: potential for screening therapeutic agents. *Journal of Infectious Diseases, 160*, 312-320.
3. Gangjee, A. & Adair, O.O. (2003). Synthesis and Biological Evaluation of 2,4-Diamino-6-(arylaminomethyl) pyrido [2,3-d] pyrimidines as inhibitors of *Pneumocystis crainii* and *Toxoplasma gondii* Dihydrofolate Reductase and as Antiopportunistic infection and Antitumor Agents. *Journal of Medicinal Chemistry, 46*, 5074–5082.
4. Gangjee, A. & Adair, O.O. (2008). N9-Substituted 2,4-Diamino quinazolines: Synthesis and Biological Evaluation of lipophilic inhibitors of *Pneumocystis crainii* and *Toxoplasma gondii* Dihydrofolate Reductase. *Journal of Medicinal Chemistry, 51*, 6195–6200.
5. Mishra, R. & Agarwal, A. (2012). QSAR analysis of antitoxoplasma agents. In *Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives* pp. 131-135. Springer-Verlag, Berlin.
6. Agarwal, A. & Mishra, R. (2012). A QSAR study investigating the potential anti-leishmanial activity of cationic 2-Phenylbenzofurans. In *Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives* pp. 137-141. Springer-Verlag, Berlin.
7. JinCan, C. & Li, Q. (2008). A QSAR study and molecular design of benzothiazole derivatives as potent anticancer agents. *Science in China - Series B***:** *Chemistry, 51*, 111-119.
8. Golbraikh, A. & Shen, M. (2003**).** Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design, 17***,** 241–253.
9. Yu, Y. & Su, R. (2009). Comparative QSAR modeling of antitumor activity of ARC-111 analogues using stepwise MLR, PLS, and ANN techniques. *Medicinal Chemistry Research, 19*, 1233-1244.
10. Cramer, R.D. (1993). Partial Least Squares (PLS): its strengths and limitations. *Perspectives in Drug Discovery and Design, 1*, 269–278.