INTERNATIONAL JOURNAL OF RESEARCH IN PHARMACY AND CHEMISTRY

Available online at www.ijrpc.com

Research Article

IN-SILICO MOLECULAR MODELING OF NITRIC OXIDE SYNTHASES INHIBITORS

Ruchi Mishra*, Sarvesh Paliwal, Ankita Agarwal and Ruchi Choudhary

Department of Pharmacy, Banasthali University, Banasthali -304022, Rajasthan, India.

ABSTRACT

Global physicochemical descriptor based QSAR models were developed using multiple linear regression (MLR), partial least squares (PLS) and feed forward neural network (FFNN) techniques for a set of 71 molecules as NOS Inhibitor belonging to 2-Imidazol-1-ylpyridine derivatives synthesized against acute and chronic inflammatory diseases. Leave out one row method is used to validate the developed model. The MLR and PLS generated excellent models with good predictive ability and all the statistical values, such as r, r^2, r^2_{cv}, r^2 (test set), F and S values were 0.88, 0.78, 0.76, 0.73, 40.42 and 0.26 for MLR and r^2_{cv}, r^2 (test set) and statistical significance value were 0.75, 0.70 and 0.99 for PLS respectively, were satisfactory. The model developed from feed forward neural network (FFNN) technique also showed good correlation value of r^2= 0.87. The results obtained from this study provide insights regarding role of Dipole moment Z component (whole molecule), Total Lipole (whole molecule), Number of H-bond acceptors (whole molecule) and VAMP polarization XY (whole molecule) in determining the NOS inhibitory activity.

Keywords: QSAR, TSAR, MLR, PLS, 2-Imidazol-1-ylpyridine derivatives.

INTRODUCTION

Nitric oxide synthases (NOSs) are a family of enzymes that catalyze the production of nitric oxide (NO) from L-arginine. NOS is one of the most regulated enzymes in biology. There are three known isoforms, two are constitutive (cNOS) and the third is inducible (iNOS). Cloning of NOS enzymes indicates that, cNOS include both brain constitutive (NOS1) and endothelial constitutive (NOS3), the third is the inducible (NOS2) gene [1]. Studies have shown that production of NO by iNOS is implicated in a variety of acute and chronic inflammatory diseases (e.g., sepsis, septic shock, vascular dysfunction in diabetes, asthma, arthritis, multiple sclerosis, and inflammatory diseases of the gut) [2]. NOS inhibition may be promising for the treatment of anaphylactic shock, however, the failure of a phase III trial of some inhibitors such as L-NMMA in septic shock, indicates that other approaches need to be pursued for the successful treatment of septic shock. High hopes have been placed on selective iNOS inhibitors as one of the possible remedies for this condition [3]. A combination of selective iNOS inhibition and β adrenergic agonists, such as isoprenaline, in treatment of endotoxic shock has been reported to have beneficial effects [4].

Development of new iNOS inhibitors as novel chemotherapeutics directed against acute and chronic inflammatory diseases. Here, we report the QSAR studies of 2-Imidazol-1-ylpyridine derivatives as inhibitors of iNOS, and we include some QSAR parameters of the structure-activity relationship. In view of iNOS inhibitors and their toxicities we decided to perform optimization of existing leads using sophisticated computer aided drug design technique like QSAR. Quantitative structure activity relationship provides the guidelines for making structural

changes in the compound so that drugs of higher potency can be obtained. A successful QSAR model generates statistically significant relationships between chemical structure and biological activity [1].

In the present paper, we describe our attempt to investigate the relationship between the various physiochemical parameters and anti-inflammatory activity of 2-Imidazol-1-ylpyridine derivatives that may be helpful in development of potent anti-inflammatory agents. The present study has been performed employing equation based techniques and Comparison of the performance of developed multivariate regression models with FFNN in order to obtain robust model with least possibilities of over fitting and chance correlation.

The focus of the present work is the analysis of important physiochemical descriptors and structural contribution at level of atoms and groups that will further help in the design of more potent iNOS inhibitors.

MATERIAL AND METHODS

The structures and anti-cancer activities of 71 compounds of 2-Imidazol-1-ylpyridine derivatives[5] (Table 1) 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) [6]. To reduce data redundancy, pairwise correlation analysis was carried out [7]. 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 five descriptors were retrieved that were highly correlated with biological activity and were not having intercorrelation among each other.

Model Development

Linear Regression Analysis

To develop QSAR models, stepwise MLR analysis with leave-one-out (LOO) [8] cross-validation was applied to the training set. The molecules of the series were divided randomly into training set (55 molecules) and test set (16 molecules) with descriptors retrived by data reduction. Training set was used to build linear models so that an accurate relationship could be found between structure and biological activity [9]. The test set of six 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) [10] analyses were also performed to check the predictive ability and robustness of the developed model.

Neural Network Analysis

FFNN is a layered system of processing units that are interconnected to facilitate the ordered transfer and data processing. It is reported that sometimes FFNN is superior to MLR in providing accurate predictions. By definition, MLR assumes a linear relationship between binding affinities and molecular descriptors, or incorporated explicitly [11]. In contrast, FFNN makes no assumption about the linearity of a problem. The major advantage of FFNN lies in the fact that QSAR can be developed without having to specify the analytical form of a particular correlation model.

In this work, the neurons were arranged in a three-layered forward feed neural network model: an input layer (molecular descriptor values used in the final MLR and PLS models), a hidden layer, and an output layer (antagonistic activity). The Monte Carlo algorithm was used to select a better set of starting weights within the default constrained limits. The ratio (ρ) between the number of input variables and the number of hidden neurons, which is critical to the predictive power of the FFNN, was set to close to 2 to prevent the problems of overfitting or memorizing data. 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. The results were visualized on a 2D plot of output node against input (dependency graph).

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 was constructed from the training set by using 4 parameters.

In order to improve the predictivity of the model, 6 potential outliers namely 11d (r), 11c, 11v, 12q, 19f (s) and 12y (which exhibited high residual value and were two far away from regression line) were identified and deleted.

The final regression equation obtained from MLR analysis (final 49 molecules in training set) after deleting the outliers is represented as equation 1

$$Y = -0.169 \times X1 - 0.200 \times X2 + 0.496 \times X3 + 0.077 \times X4 - 2.66 \quad \textit{–Equation 1}$$

Where X1= Dipole moment Z component (whole molecule), X2= Total Lipole (whole molecule), X3= Number of H-bond acceptors (whole molecule) and X4= VAMP polarization XY (whole molecule).

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 [12]:

- $r^2_{cv} > 0.6$: 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 = 0.88, r^2 = 0.78, r^2_{cv} = 0.76, F = 40.42, S = 0.26 and predictive r^2 for test set = 0.73

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 0.75 clearly demonstrates the high predictive ability of the developed PLS model (equation 2).

$$Y = -0.154 \times X1 - 0.226 \times X2 + 0.444 \times X3 + 0.076 \times X4 - 2.18 \quad \textit{-Equation 2}$$

Where X1= Dipole moment Z component (whole molecule), X2= Total Lipole (whole molecule), X3= Number of H-bond acceptors (whole molecule) and X4= VAMP polarization XY (whole molecule)

Statistical significance = 0.99, r^2_{cv} = 0.75, Fraction of variance explained = 0.77, Predictive r^2 for test set = 0.70

Since for a well defined problem, both MLR and PLS should generate comparable results[13], the r^2_{cv} values of MLR and the PLS models were evaluated and it was found that both the models have comparable r^2_{cv} value of 0.75 and 0.75 for MLR and PLS respectively. The predictive ability of the model was also validated using the external test set of 16 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 2, 3 and their respective plots are depicted in figure 1, 2, 3 and 4.

Feed forward neural network 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. The 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 5-6. 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 & test set were given by the trained neural network architecture ($r^2_{training}$=0.87 and r^2_{test}=0.64). The dependency plots obtained in FFNN through TSAR software are given in figure: 7-10. The actual and predicted biological activity values of FFNN analysis for training and test which is shown in table 2, 3. 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.

Analysis of Descriptors

The Dipole descriptor is a 3D electronic descriptor that indicates the strength and orientation behavior of a molecule in an electrostatic field. Both the magnitude and the components (X, Y and Z) of the dipole moment

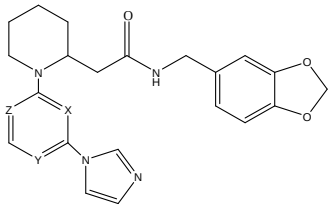
are calculated. It is estimated by utilizing partial atomic charges and atomic coordinates. Dipole properties have been correlated to long range ligand-receptor recognition and subsequent binding. Dipole moment Z component (Whole molecule) is negatively correlated with the biological activity in linear analysis. Dependency plot of neural network analysis also shows negative correlation (figure 7). So with decrease in dipole moment of molecule there will be increase in biological activity.

Lipophilicity is defined by the partitioning of a compound between an aqueous and a nonaqueous phase. P is defined as the ratio of substance concentrations in the organic and aqueous phases of a two-compartment system under equilibrium conditions. Total lipole (Whole molecule) is negatively correlated with the biological activity. Dependency plot of neural network analysis also shows negative correlation (figure 8). So with decrease in total lipole of molecule there will be increase in biological activity.

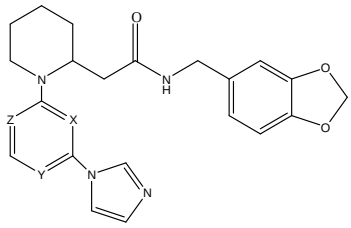
The number of Hydrogen bond acceptors is structural descriptors. This descriptor calculates the number of hydrogen bond acceptors. The number of Hydrogen bond acceptors is positively correlated biological activity, so with increase in number of Hydrogen bond acceptors there will be increase in biological activity. The dependency plot of neural network of Hydrogen bond acceptors also shows positive correlation with biological activity (figure 9).

VAMP polarization Descriptors, the semi-emperical quantum mechanics engine, provides energy, orbitals (HOMO/LUMO), multipoles etc. VAMP polarization XY (Whole molecule) is positively correlated biological activity, so with increase in VAMP polarization there will be increase in biological activity. The dependency plot of neural network of VAMP polarization also shows positive correlation with biological activity (figure 10).

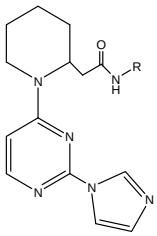
Table 1: Structure and biological activity data of 2-Imidazol-1-ylpyrimidine analogues



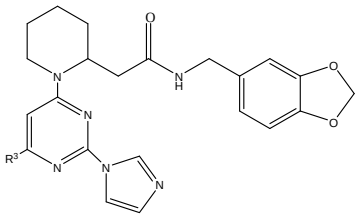
Comp. Name	R^2	IC_{50} Values	Comp. Name	R^2	IC_{50} Values
11a	$COCH_3$	1	11n	$COCH_2OBn$	0.50
11b	CH_2Ph	1	11o	CO-2-furan	0.67
11c	COPh	0.5	11p	CO_2Ph	1.4
11d	$COOCH_3$	0.38	11q	COOBn	0.8
11e	$CONHCH_3$	1	11r	$(CH_2)_2CHMe_2$	1.0
11f	SO_2CH_3	0.7	11s	COO^iPr	1.0
11g	H	0.55	11t	6-F-2-pyridine	0.28
11h	CH_3	0.48	11u	CH_2COOEt	0.93
11i	CH_2-1-naphthalene	69	11v	C(=NH)Me	33
11j	CONHPh	0.87	11w	dansyl	1.1



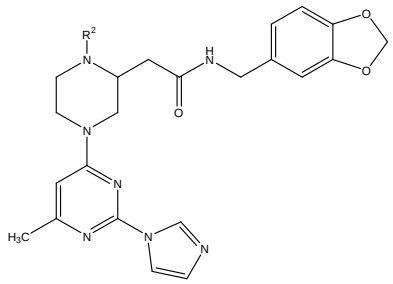
Comp. Name	X	Y	Z	IC_{50} Values
15	N	CH	N	52
16	CH	N	N	10
17	CH	CH	CH	77



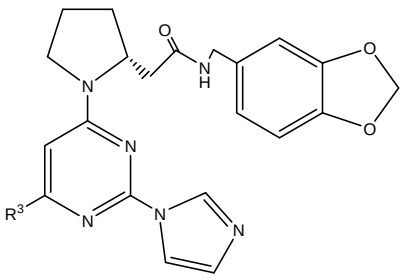
Comp. Name	R	IC_{50} Values	Comp. Name	R	IC_{50} Values
12a	CH_2-benzodioxolane	0.5	12k	CH_2Ph-4-CH_3	23
12b	CH_2Ph-4-OCH_3	0.8	12l	CH_2Ph-4-NH_2	270
12c	CH_2Ph-3-OCH_3	3.7	12n	CH_2Ph-4-$N(CH_3)_2$	66
12d	CH_2Ph-2-OCH_3	283	12o	CH_2Ph-4-CF_3	150
12e	CH_2Ph-3,4-OCH_3	1.5	12q	CH_2Ph-4-NO_2	168
12g	CH_2Ph	25	12s	CH_2Ph-4-SO_2CH_3	17
12h	CH_2Ph-4-Cl	5.7	12t	CH_2-2-furan	63
12i	CH_2Ph-4-F	3.0	12u	CH_2-2-pyridine	539
12j	CH_2Ph-4-OCF_3	19	12v	$(NCH_3)CH_2Ph$	885
12w	$CHPh_2$	1805	12bb	$(CH_2)_2Ph$-3,4-OCH_3	12
12y	2-benzimidazole	11	12cc	$(CH_2)_2Ph$-4-OCH_3	3.5
12aa	Ph-3,4-OCH_3	66			



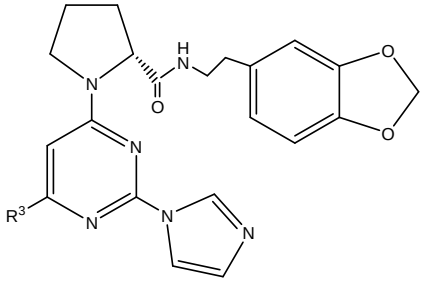
Comp. Name	R^3	IC_{50} Values	Comp. Name	R^3	IC_{50} Values
18a	Cl	0.49	18d	CF_3	180
18b	Ph	67	18e	imidazole	35
18c	Me	4.1			



Comp. Name	R^2	IC_{50} Values	Comp. Name	R^2	IC_{50} Values
19a	H	100	19g	CO_2CH_3	5.4
19b	CH_3	12	(S)-11d	-	0.24
19c	Bn	25	(R)-11d	-	34
19d	SO_2CH_3	12	(R)-19f	-	0.75
19e	Ac	8.5	(S)-19F	-	300
19f	$COCH(CH_3)_2$	1.9			



20



21

Comp. Name	R^3	IC_{50} Values
20a	H	0.28
20b	Me	0.12
20c	Et	0.13
21a	H	0.48
21b	Me	0.29
21c	Et	0.52

Table 2: Actual and predicted activity of training set obtained by MLR, PLS and FFNN

Comp. Name	Actual activity	Predicted activity		
		MLR	PLS	FFNN
11f	0.154	-0.133	-0.388	-0.279
11h	0.318	0.363	0.477	0.376
11i	-1.83	-1.831	-2.028	-2.088
11j	0.060	0.008	0.089	0.004
11k	0.017	-0.919	-0.946	-0.744
11l	-0.431	0.174	0.086	-0.033
11m	0.236	0.647	0.659	0.451
11n	0.301	0.403	0.262	0.133
11o	0.173	0.659	0.542	0.398
11p	-0.146	-0.286	-0.433	-0.232
11q	0.096	-0.646	-0.755	-0.300
11r	0	0.170	0.242	0.119
11s	0	-0.354	-0.312	-0.076
11t	0.552	0.083	0.080	0.595
11u	0.031	0.698	0.707	0.097
11w	-0.041	0.512	0.554	-0.062
12a	0.301	-0.272	-0.241	-0.184
12aa	-1.819	-1.07	-1.085	-1.458
12b	0.096	-0.731	-0.626	-1.000
12bb	-1.079	-0.549	-0.598	-0.788
12c	-0.568	-0.447	-0.318	-0.542
12cc	-0.544	-0.824	-0.758	-0.386
12d	-2.451	-1.523	-1.441	-2.065
12k	-1.361	-1.36	-1.236	-1.597
12l	-2.43	-2.93	-2.683	-2.412
12n	-1.819	-1.465	-1.367	-1.658
12o	-2.176	-2.615	-2.576	-2.794
12s	-1.230	-1.842	-1.852	-2.391
12t	-1.799	-0.931	-0.844	-1.234
12u	-2.731	-2.308	-2.25	-2.575
12v	-2.946	-2.679	-2.61	-2.709
12w	-3.256	-2.399	-2.509	-3.084
15	-1.716	-1.245	-1.118	-1.475
16	-1	-0.555	-0.504	-0.726
17	-1.886	-2.062	-1.946	-2.202
18c	-0.612	-0.926	-0.906	-1.276
18d	-2.25	-2.083	-2.218	-2.885
18e	-1.54	-1.335	-1.405	-1.469
19a	-2	-1.34	-1.393	-1.415
19b	-1.07	-1.370	-1.389	-1.305
19c	-1.39	-1.347	-1.220	-1.249
19d	-1.07	-1.54	-1.762	-0.988
19e	-0.929	-1.267	-1.387	-0.859
19f	-0.278	-0.631	-0.675	-0.493
19g	-0.732	-0.572	-0.791	-0.479
20a	0.552	-0.127	-0.096	0.457
20b	0.920	0.042	0.078	0.574
21a	0.318	0.094	0.156	0.384
21b	0.537	0.181	0.241	0.423

Table 3: Actual and predicted activity of test set obtained by MLR, PLS and FFNN

Comp. Name	Actual activity	Predicted activity		
		MLR	PLS	FFNN
11a	0	-0.459	-0.380	-0.144
11b	0	-0.904	-0.980	-0.794
11d	0.420	0.222	0.295	0.104
11e	0	-0.821	-1.010	-0.805
11g	0.259	-0.666	-0.556	-0.108
12e	-0.176	-1.032	-1.008	-1.243
12h	-0.755	-2.228	-2.201	-2.528

12i	-0.477	-1.527	-1.43	-1.769
12j	-1.278	-2.766	-2.94	-3.365
18a	0.309	-1.183	-1.295	-1.894
18b	-1.82	-3.530	-3.917	-3.604
20c	0.886	-0.559	-0.570	-0.776
21c	0.283	0.0328	0.0414	0.335
19f (r)	0.124	-0.562	-0.602	-0.418
12g	-1.397	-1.221	-1.12	-1.382
11d (s)	0.619	0.222	0.295	0.104

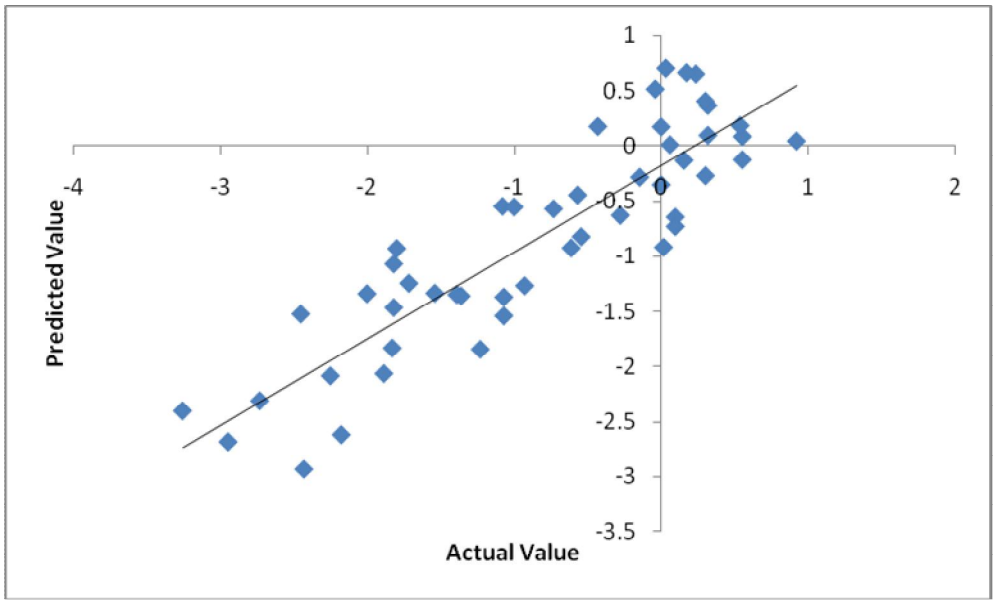


Fig. 1: Actual vs. predicted activity for the training set of compounds derived from MLR analysis

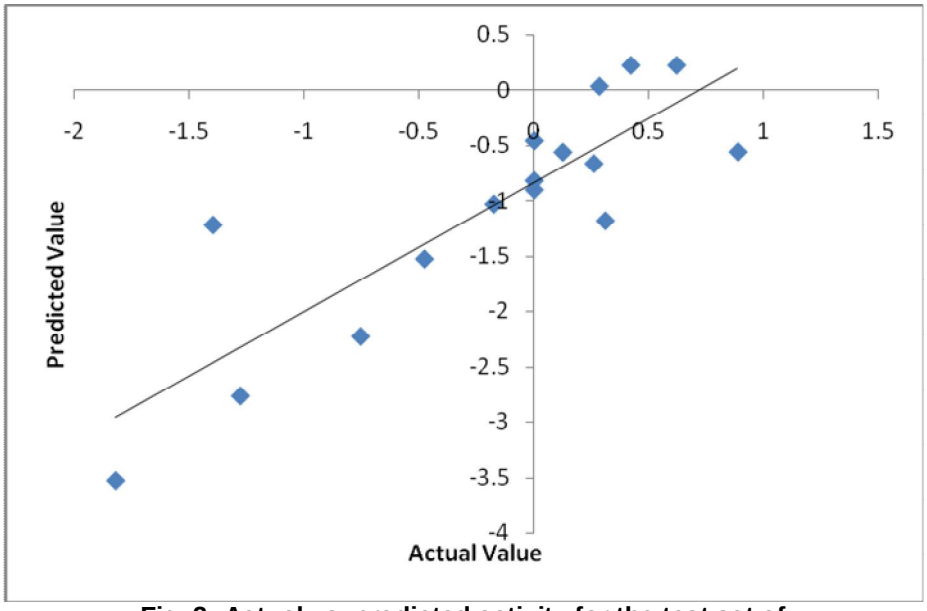


Fig. 2: Actual vs. predicted activity for the test set of compounds derived from MLR analysis

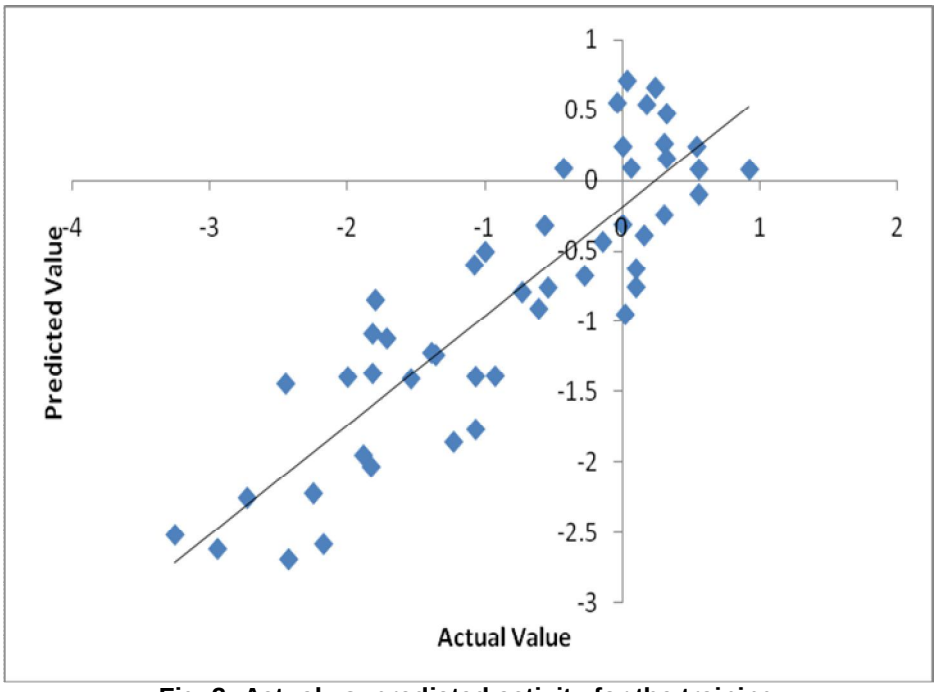


Fig. 3: Actual vs. predicted activity for the training set of compounds derived from PLS analysis

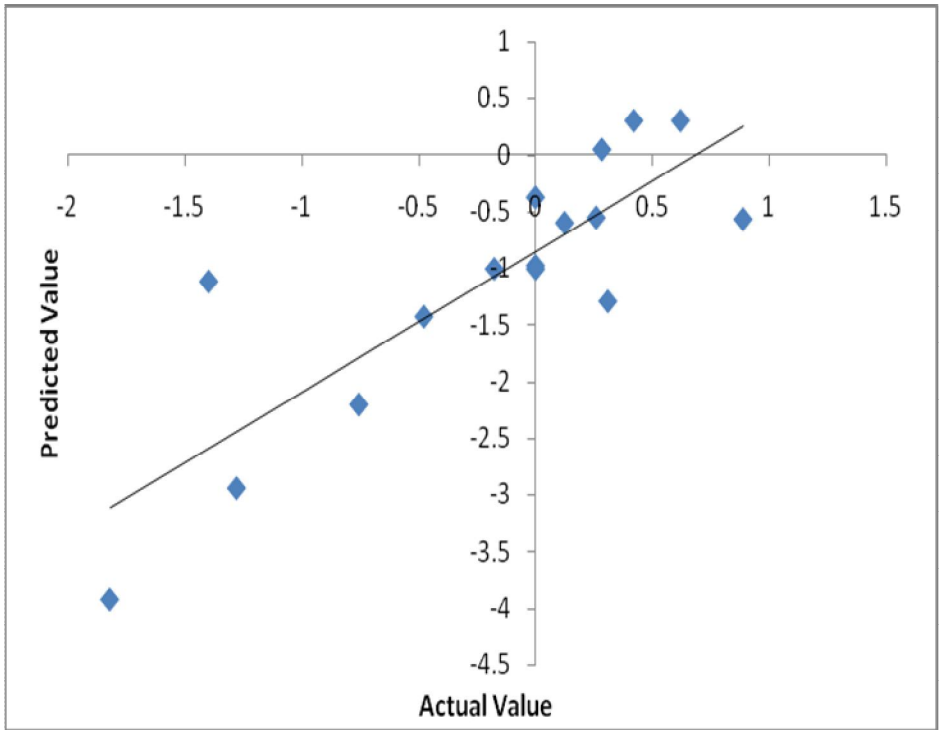


Fig. 4: Actual vs. predicted activity for the test set of compounds derived from PLS analysis

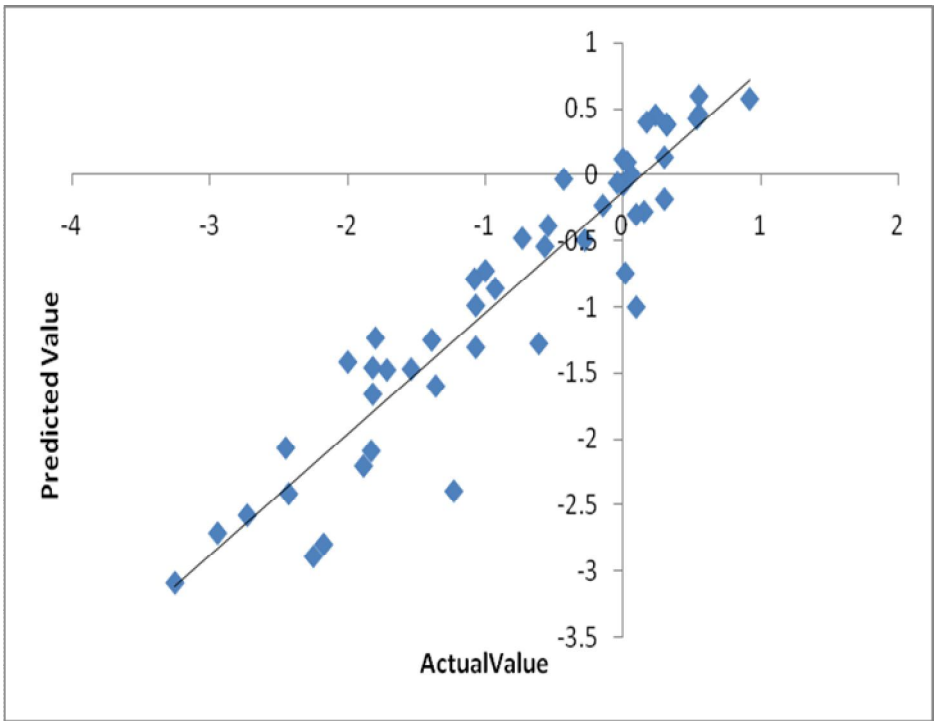


Fig. 5: Actual vs. predicted activity for the training set of compounds derived from FFNN analysis

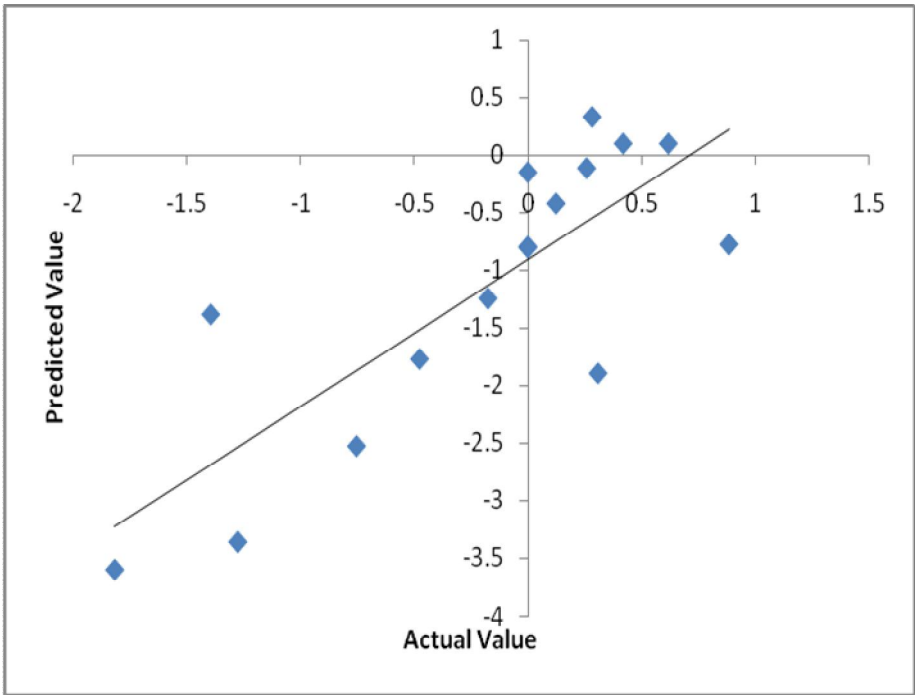


Fig. 6: Actual vs. predicted activity for the test set of compounds derived from FFNN analysis

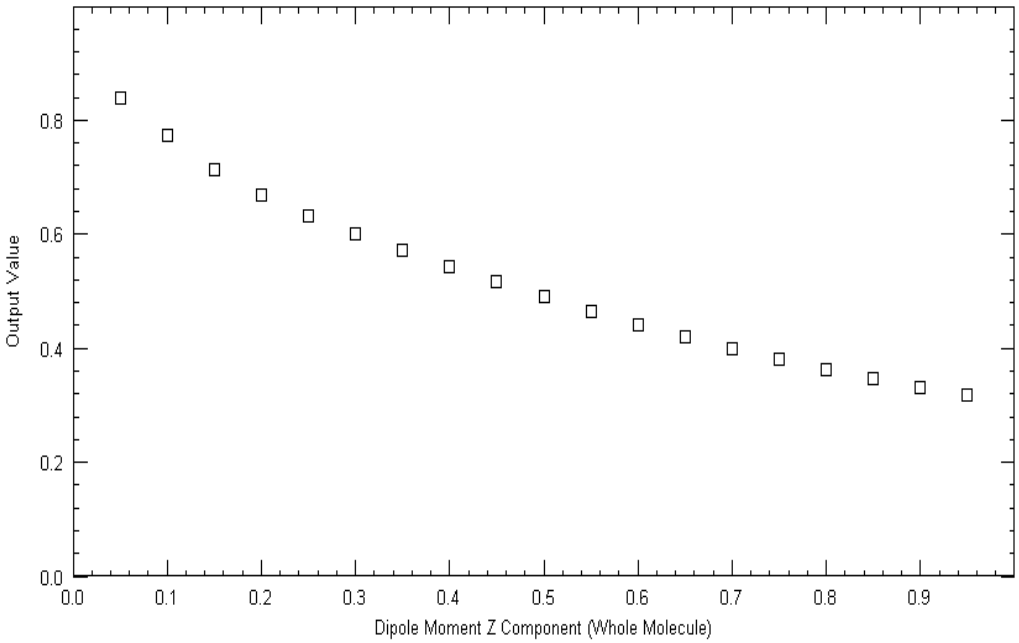


Fig. 7: Dependency plot between biological activity and Dipole Moment Z Component (whole molecule) used to built the final model

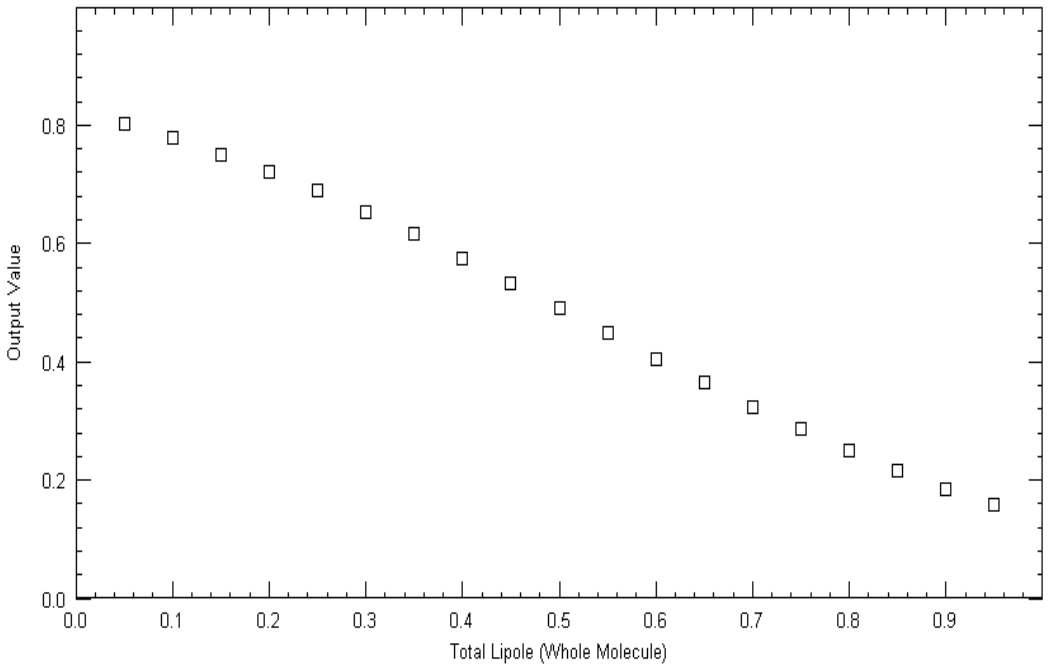


Fig. 8: Dependency plot between biological activity and Total Lipole (whole molecule) used to built the final model

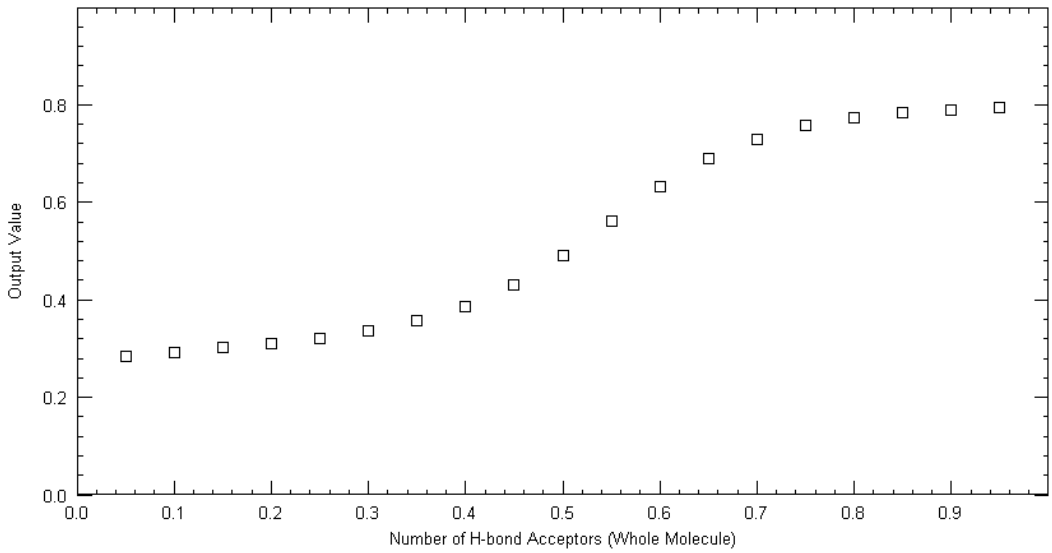


Fig. 9: Dependency plot between biological activity and Number of H-bond Acceptors (whole molecule) used to built the final model

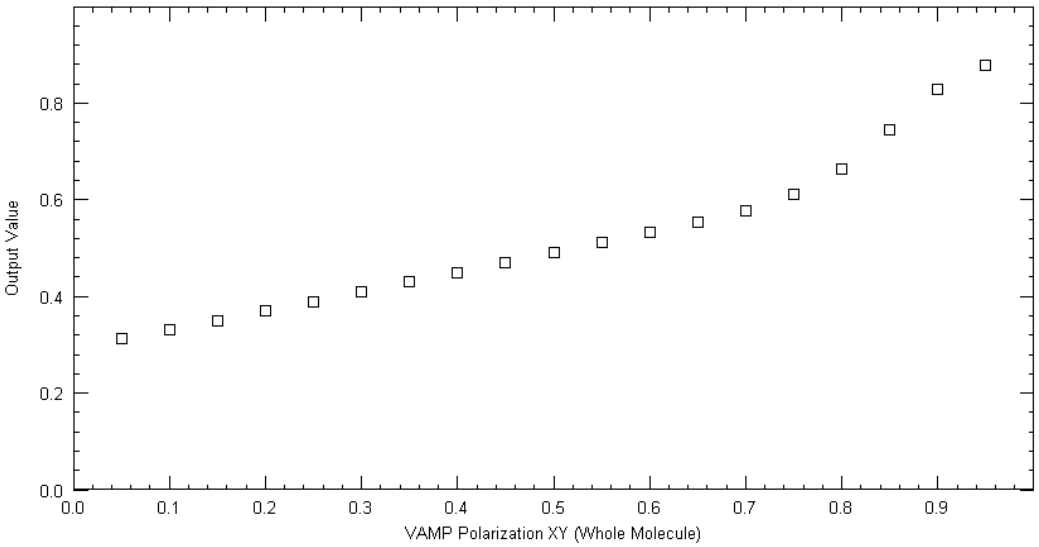


Fig. 10: Dependency plot between biological activity and VAMP Polarization XY (whole molecule) used to built the final model

CONCLUSION

The MLR, PLS and FFNN were employed to study the iNOS inhibitory activity of 2-Imidazol-1-ylpyridine derivatives. Highly predictive QSAR models were obtained using the MLR, PLS and FFNN. All the models were validated using external test set of 16 compounds. All three different statistical approaches (MLR, PLS and FFNN) generated nearly the same results for each QSAR model and allow us to estimate additionally the quality of prediction. The findings of present study will certainly aid in the design of more potent iNOS inhibitors with improved

activity and reduced mechanism based side effects of traditional iNOS inhibitors.

ACKNOWLEDGEMENT
Computational resources were provided by Banasthali University, and the authors thank the Vice Chancellor, for extending all the necessary facilities.

REFERENCES

1. Stuehr DJ. Mammalian nitric oxide synthases. Biochim. Biophys. Acta. 1999;1411 (2-3): 217–30.
2. Kroncke KD and Fehsel K. Inducible nitric oxide synthase in human diseases. Clin. Exp. Immunol. 1998; 113(2): 147–156.
3. Suliburk JW, Helmer KS and Kennison SD. Time dependent aggravation or attenuation of LPS induced gastric injury by nitric oxide synthase inhibition. J Surg Res. 2005; 129(2): 265-271.
4. Bryk R and Wolff DJ. Mechanism of inducible nitric oxide synthase inactivation by aminoguanidine and L-N6-(1-iminoethyl)lysine. Biochemistry. 1998; 37: 4844-4852.
5. Davey DD and Adler M. Design, synthesis, and activity of 2-imidazol-1-ylpyrimidine derived inducible nitric oxide synthase dimerization inhibitors. J. Med. Chem. 2007; 50: 1146-1157.
6. Mishra R, Agarwal A and S Paliwal. QSAR Analysis of Anti-Toxoplasma Agents. In Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Edited by Khemani LD, Srivastava MM, Srivastava S. Springer-Verlag Berlin Heidelberg, New York. 2012; 131-135.
7. Agarwal A, Mishra R and Paliwal S. A QSAR Study Investigating the Potential Anti-Leishmanial Activity of Cationic 2-Phenylbenzofurans. In Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Edited by Khemani LD, Srivastava MM, Srivastava S. Springer-Verlag Berlin Heidelberg, New York. 2012; 137-141.
8. JinCan C, Li Q and Yong S. A QSAR study and molecular design of benzothiazole derivatives as potent anticancer agents. Sci China Ser B-Chem. 2008; 51: 111 - 119.
9. Golbraikh A, Shen M, Xiao Z and Xiao YD. Rational Selection of Training and Test Sets for the Development of Validated QSAR Models. Journal of Computer-Aided Molecular Design. 2003; 17: 241–253.
10. Yu Y, Su R and Wang L. Comparative QSARmodeling of antitumor activity of ARC-111 analogues using stepwise MLR, PLS, and ANN techniques. Med. Chem. Res. 2009; DOI 10.1007/s00044-009-9266-9.
11. Tetko IV, Livingstone DJ and Luik AI. Neural network studies. 1. Comparison of overfitting and overtraining. J Chem Inf Comput Sci. 1995; 35: 826-833.
12. Prasad RY, Kumar RP, Smiles JD and Babu AP. QSAR studies on chalcone derivatives as antibacterial agents against Bacillus pumilis. Arkivoc. 2008; 11: 266-276.
13. Cramer RD. Partial Least Squares (PLS): its strengths and limitations. Perspect Drug Discov Des. 1993; 1: 269–278.