



## Study of 2D-QSAR model for the development of Anti-HIV agents in reference of the integrase inhibitor

Priyanka Jain\*, Neha Garg, Sudesh Neyol, D.Kishore  
Dept. of Chemistry, Banasthali University, Rajasthan-304022, (INDIA)

### ABSTRACT

A qualitative structure activity relationship (QSAR) study has been performed on the series of 42 compounds derived from 3-aryl-1, 1-dioxo-, 1, 4, 2 and 4-chloro n-(4-oxopyrimidine-2-yl)-2-mercapto benzene sulfonamides. This series of integrase inhibitors were synthesized to assess their potential against HIV. The multiple linear regression analysis (MLR) and partial least squares (PLS) feed forward neural networking (FFNN) generated excellent models with good predictive ability. In reference to this the statistical value multiple linear regression (MLR)  $r$  0.89,  $r^2$  0.80 and  $r^2$  cross validation 0.71 and a comparable partial least squares (PLS) model  $r^2$  0.77  $r^2$ cv 0.65 and feed forward neural networking  $r^2$  0.95 was obtained. The developed model was further validated by leave-one-out method of cross-validation and prediction of test set. The study indicated that the Anti-HIV activity could largely be explained by kier chiv4 (path/cluster) index (whole molecule), inertia moment 1 length (whole molecule), H-bond donors (whole molecule) and VAMP heat of formation (whole molecule).

© 2014 Trade Science Inc. - INDIA

### KEYWORDS

HIV;  
Quantitative structure  
activity relationship;  
Tools for structure activity  
relationship;  
Diabetes.

### INTRODUCTION

The successful use of highly active antiretroviral therapy (HAART) can dramatically suppress human immunodeficiency virus (HIV)- 1 viral replication and effect significant immune reconstitution<sup>[1-3]</sup>. This structural motif contained in the molecule possessed metal-chelating functions, and it is postulated that compounds bearing these functional groups interact with divalent metals within the active site of HIV-1 integrase. Human immunodeficiency virus type 1 (HIV-1) encodes three enzymes which are required for viral replication such as the reverse transcriptase, protease, and integrase (IN). Although drugs targeting reverse transcriptase and protease are widely used and have shown effectiveness particu-

larly when employed in combination, toxicity and development of resistant strains have limited their usefulness<sup>[1]</sup>.

The virally encoded integrase protein is an essential enzyme in the life cycle of the HIV-1 virus and represents an attractive and validated target in the development of therapeutics against HIV infection. Drugs that selectively inhibit this enzyme, when used in combination with inhibitors of reverse transcriptase and protease, are believed to be highly effective in suppressing the viral replication. Among the HIV-1 integrase inhibitors, the  $\alpha$ -diketo acids (DKAs) represent a major lead for anti-HIV-1 drug development. In this study, novel bifunctional quinolonyl diketo acid derivatives were designed, synthesized, and tested for their inhibitory ability against HIV-1 integrase. The compounds are

## Full Paper

potent inhibitors of integrase activity.

Integrase catalyzes two reactions that are required for the insertion of the reverse-transcribed viral genome into the host DNA<sup>[3-5]</sup>. In the first reaction, endonucleolytic cleavage, the terminal two 3' nucleotides are removed from the U3 and U5 regions at each end of the HIV-1 DNA. After 3' end processing, integrase catalyzes strand transfer between the recessed viral DNA ends and the cellular DNA<sup>[6]</sup>. The sequence specific oligonucleotide representing the U3 or U5 end of HIV-1 is referred to as the viral end or donor substrate, and the nonspecific oligonucleotide that mimics the cellular DNA is termed the target substrate. In vitro, integrase also catalyzes a disintegration reaction. In this reaction integrase excises the viral DNA end and joins adjacent target sequences from a branched oligonucleotide. Because this substrate mimics an integration intermediate, dis-integration is referred to as the reverse of integration<sup>[7]</sup>.

## MATERIAL AND METHODS

A data set of 42 compounds has been taken as a series from published article<sup>[8]</sup>. All structure of these 3-aryloxy-1,1-dioxo-,1,4,2 and 4-chloro n-(4-oxopyrimidine-2-yl)-2-mercapto benzene sulfonamide derivatives were constructed using Chem Draw and transferred to Chem 3D to convert them into 3D structures and then transferred to standalone module of Discovery Studio (version 2.0) and were loaded via .mol files into the work sheet of TSAR (version 3.3; Oxford Molecular, Oxford, UK). Most stable structure for all the compounds were generated and used for calculating various topological, steric, and electronic descriptors. All the calculated descriptor values were considered as independent variable and biological activity as dependent variable. The various descriptors studied are shown in TABLE 1.

TABLE 1 : Various descriptors

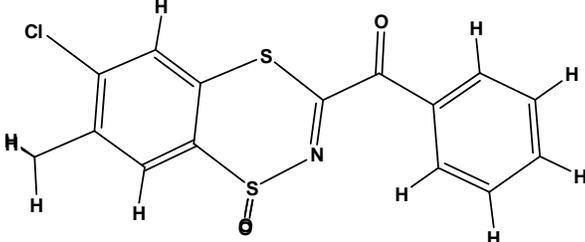
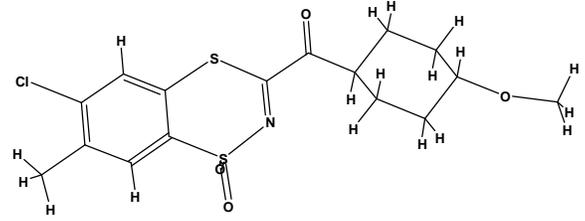
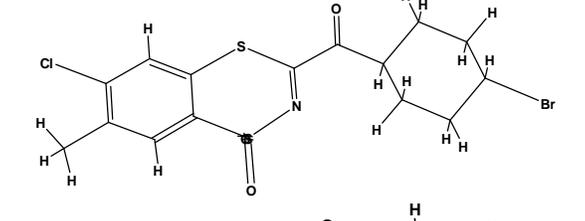
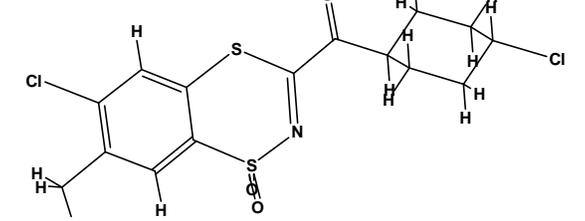
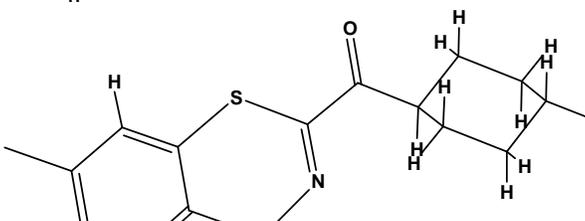
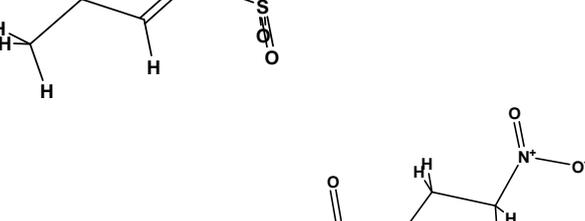
Descriptor Class	Descriptors name
Bulk	Structural mass, Structural surface area, Structural volume.
Verloop	Verloop L, Verloop B1, Verloop B2, Verloop B3, Verloop B4, Verloop B5.
Inertia	Inertia moment 1 size, inertia moment 1 length, inertia moment 2 size, inertia moment 2 length, inertia moment 3 size, Inertia moment 3 length, Ellipsoidal volume.
Log P	Total dipole, log P, lipole component, molar refractivity, substituents bond lipole.
Shape	Kappa 1, Kappa 2, Kappa 3, K-alpha1, K-alpha 2, K-alpha 3, molecular flexibility, rotatable bond count.
Topological indices	Radices, balaban, wiener.
Cosmic parameter	Optimize structures, includes bond potential, bond angle potential, torsional potential, electrostatic interaction, Vander waal interaction.
Vamp	Bond length, bond angle, torsions, total energy, electronic energy, LUMO eigen value, nuclear repulsion, ionization potential, overall atomic Charge.

## DATA SET PREPARATION AND DATA REDUCTION

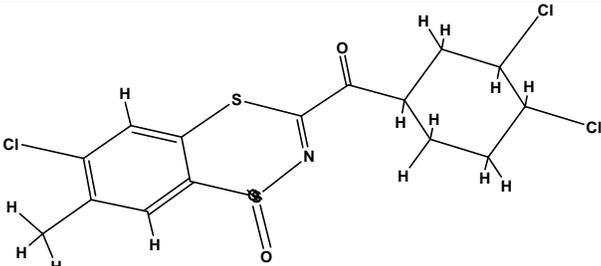
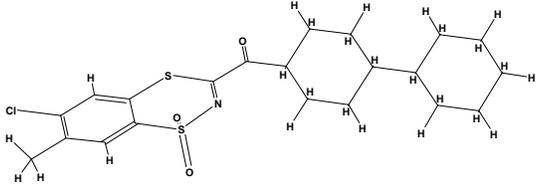
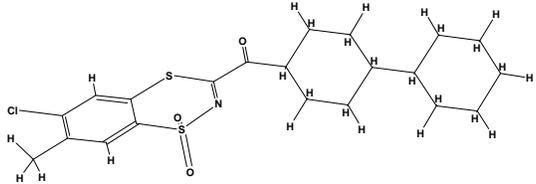
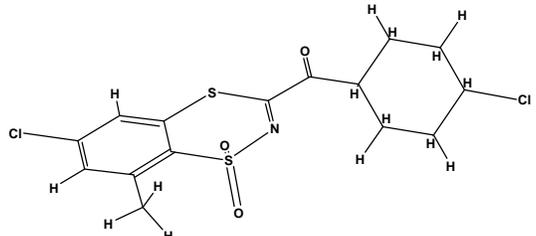
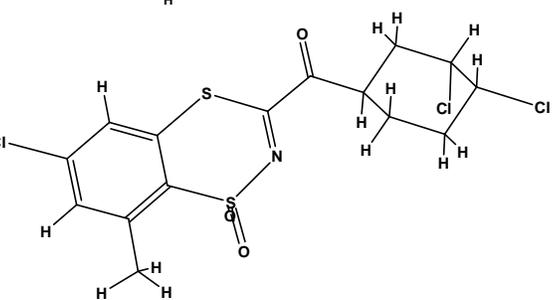
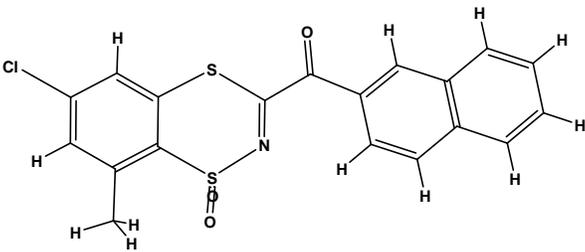
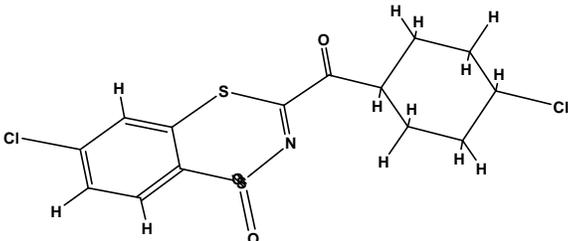
All the 42 compounds were randomly divided into training set and test set. 36 compounds were included in training set and were used to develop regression models, while the remaining 6 compounds were used in test set for the prediction of biological activity. Tools Structure Activity Relationship software was used to generate Quantitative Structure Activity relationship models by multiple linear regression analysis, partial least square and neural networking. Cross validation was performed

using leave-one out (LOO) method. Statistical measures were used, number of moles in regression, r<sup>2</sup> correlation-coefficient, F-test (Fischer's value) for statistical significance, S-standard deviation. The values of descriptors for every compound were checked to ensure that value of each descriptor was calculated for each structure and that there are sufficient variations in these values. The descriptors for which values were not calculated and/or had constant value for every structure in the data were discarded. The correlation matrix was generated to study the data patterns and to reduce the data redundancy<sup>[11]</sup>. The correlation terms involved in correlation matrix indicates extent of colinearity. The

TABLE 2 : Structure and IC50 values of Integrase derivatives

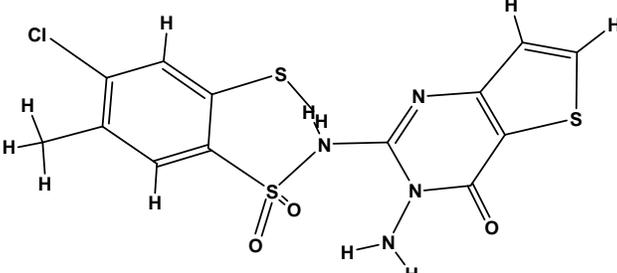
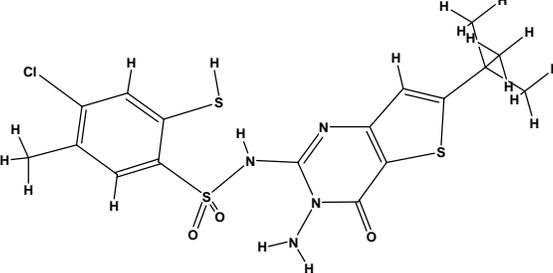
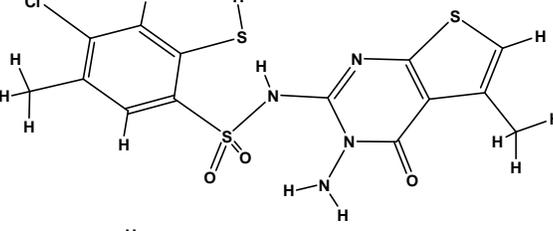
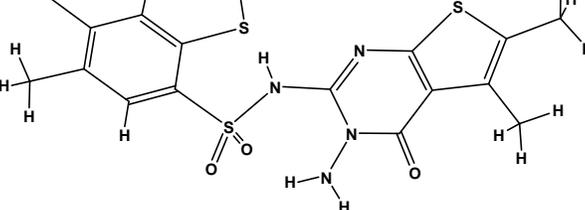
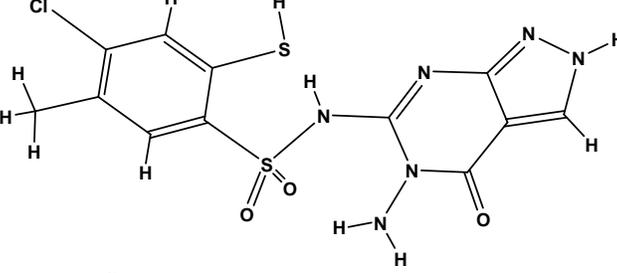
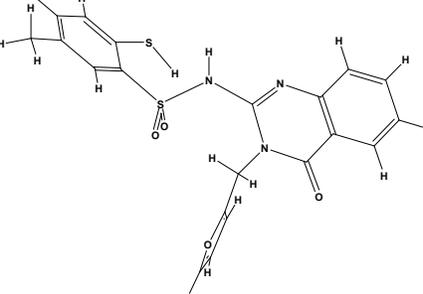
Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
1		4.92	4.618	4.6724	5.0816
2		4.52	4.3171	4.3753	4.5428
3		4.52	4.7535	4.8322	4.53
4		4.52	4.5389	4.6056	4.5793
5		4.24	4.2213	4.2744	4.2091
7*		4.24	4.5739	4.6042	4.3958

## Full Paper

Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
8		4.82	4.7055	4.7882	4.763
9		4.96	4.7055	4.7882	4.763
10*		4.55	4.8626	4.859	4.4462
11		4.57	4.5203	4.5848	4.5328
12		4.77	4.6915	4.7738	4.7729
13		4.66	4.8491	4.9148	4.6583
14*		4	4.4834	4.4914	4.3597

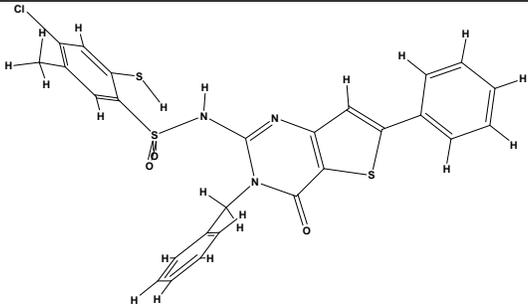
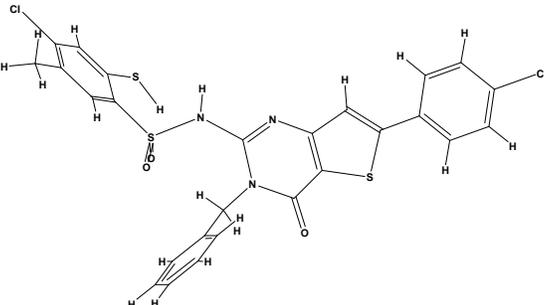
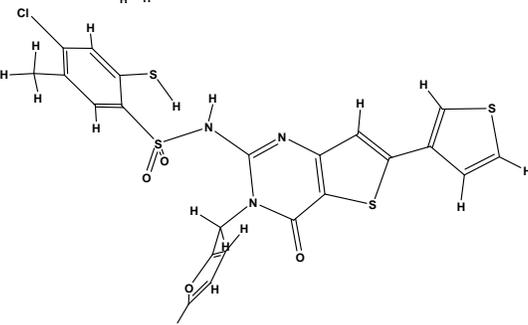
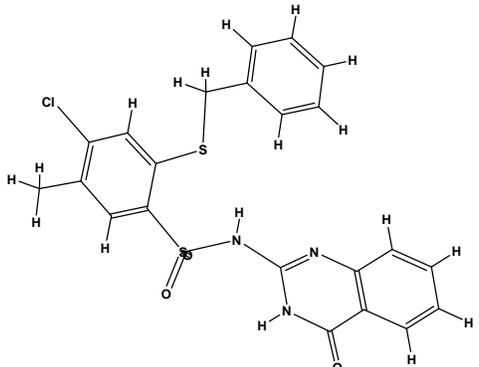
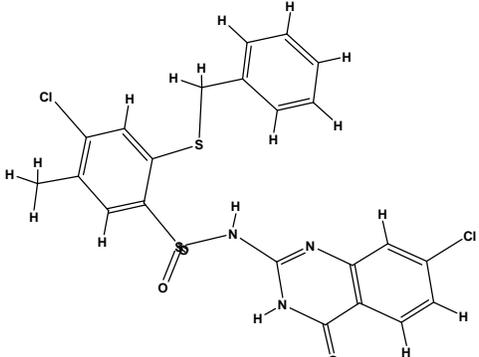
Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
15		4.6	4.4511	4.5046	4.6125
16		4.72	4.595	4.6682	4.7294
17		5.1	4.7958	4.8548	5.0786
18		4	4.1234	4.1411	4.0416
19		4	3.9364	3.9581	4
20		4.16	3.956	3.9646	4.0472

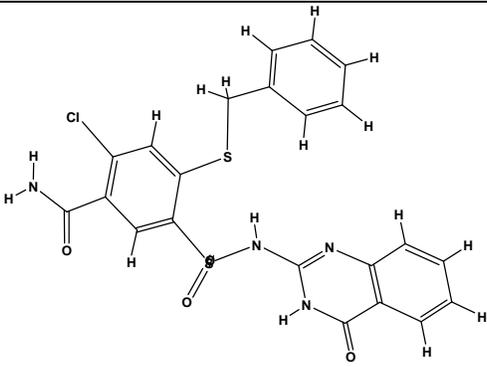
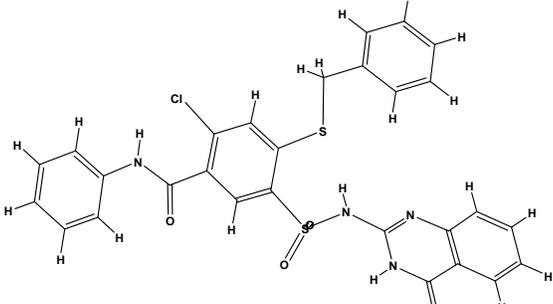
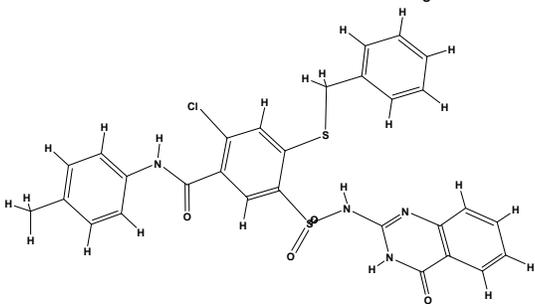
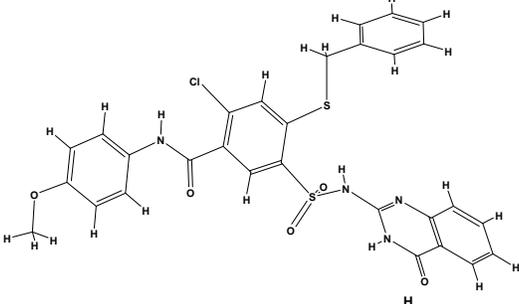
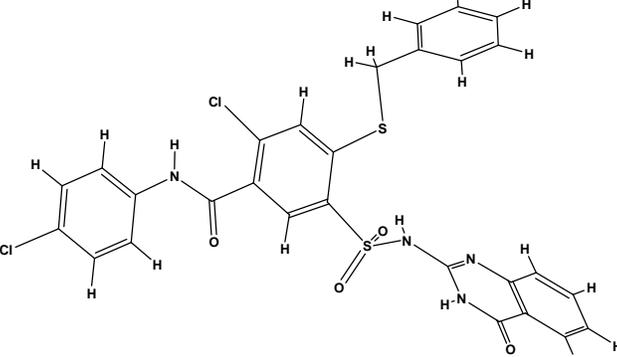
## Full Paper

Name of compound	Structures	Actual activity pIC <sub>50</sub>	Predicted MLR	Predicted PLS	Predicted NN
21		4.18	4.0661	4.0825	4.0524
22*		4.22	4.5828	4.7915	4.7939
23		4	3.9985	4.0157	4.0766
24		4.25	4.0754	4.1041	4.216
25		4	3.9787	3.9781	4.0351
26		4.11	4.2562	4.2873	4.0962

Name of compound	Structures	Actual activity pIC <sub>50</sub>	Predicted MLR	Predicted PLS	Predicted NN
27*	<p>Chemical structure of compound 27*: A benzothiazole ring system substituted with a chlorine atom and a methyl group, connected via a sulfonamide linkage to a benzothiazole ring system.</p>	4.10	4.4341	4.4502	4.321
28	<p>Chemical structure of compound 28: A benzothiazole ring system substituted with a chlorine atom and a methyl group, connected via a sulfonamide linkage to a benzothiazole ring system, which is further substituted with a phenyl ring.</p>	4.28	4.3852	4.4229	4.3054
29	<p>Chemical structure of compound 29: A benzothiazole ring system substituted with a chlorine atom and a methyl group, connected via a sulfonamide linkage to a benzothiazole ring system, which is further substituted with a phenyl ring and a methyl group.</p>	4.21	4.3036	4.3391	4.1894
30	<p>Chemical structure of compound 30: A benzothiazole ring system substituted with a chlorine atom and a methyl group, connected via a sulfonamide linkage to a benzothiazole ring system, which is further substituted with a phenyl ring and a methyl group.</p>	4.66	4.7784	4.8582	4.6732

## Full Paper

Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
31		5.05	4.7655	4.8248	5.0972
32*		5.22	4.8762	4.9762	4.9837
33		5.10	4.755	4.8154	5.09
34		4	4.1553	4.18	4.0964
35		4.33	4.1057	4.1307	4.143

Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
36		4	3.7657	3.771	4.0808
37		4.14	4.0977	4.116	4.0505
38		4	4.1255	4.1482	4.0573
39		4	4.01	4.0308	4.0734
40		4	4.2	4.2268	4.0465

## Full Paper

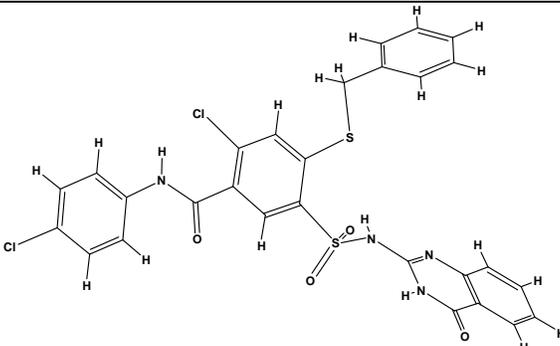
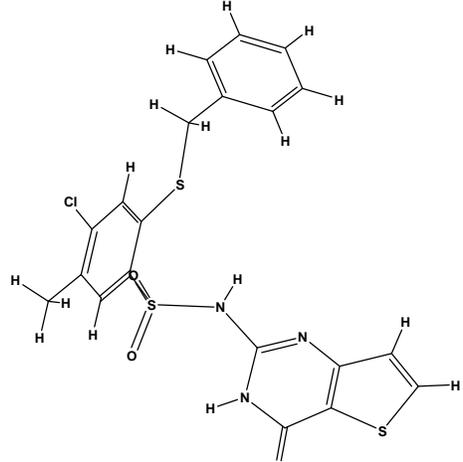
Name of compound	Structures	Actual activity pIC50	Predicted MLR	Predicted PLS	Predicted NN
41		4	4.145	4.1691	4.0988
42		4	4.1611	4.1882	4.1361

TABLE 3 : Statistical test and their values obtained after performing multiple linear regression analysis of Integrase inhibitor

Statistical Tests	S value	f value	F probability	r value	(r) <sup>2</sup> value	r <sup>2</sup> (CV)
Values	0.182877	17.38	1.46613e-008	0.901615	0.812909	0.66

S value – standard error; F value – Fischer test; r value – regression coefficient; r<sup>2</sup>value – square of regression coefficient; r<sup>2</sup> (CV) value – cross validation

term close to 1 indicated high colinearity, while the value below 0.5 indicated that no colinearity exist between the two parameters. As there were large numbers of descriptors, which may lead to over fitting of data and low predictivity of the model, data reduction was done. Among the highly inter-correlated parameters the one that showed low correlation with the biological activity (IC50 value) was discarded while the other was kept. This process was repeated for each and every set of two consecutive parameters and finally descriptors that were highly correlated to the biological activity but did not have any correlation among each other were retained. These were kier chiv4 (path/cluster) index (whole molecule), inertia moment 1 length(whole molecule), H-bond donors (whole molecule) and VAMP heat of formation(whole molecule) which were used to per-

form the Multiple Linear Regression analysis. An Artificial Neural Network (ANN) consisted of “neurons” or “hidden units” that received data from the outside process the data and output of a signal. It was reported that Artificial Neural Network is superior to Multiple Linear Regression in providing accurate prediction. The major advantage of Artificial Neural Network lies in the analytical from a particular correlation model<sup>[12]</sup>. In this work the neurons were arranged in three layered Feed Forward Neural Networking (FFNN) model : an input layer (molecular descriptor values used in the final Multiple Linear Regression and Partial Least Square models) an hidden layer and an output layer. The ratio between the number of input variables and the number of hidden neurons, which is critical to the predictive power of the Feed Forward Neural Net-

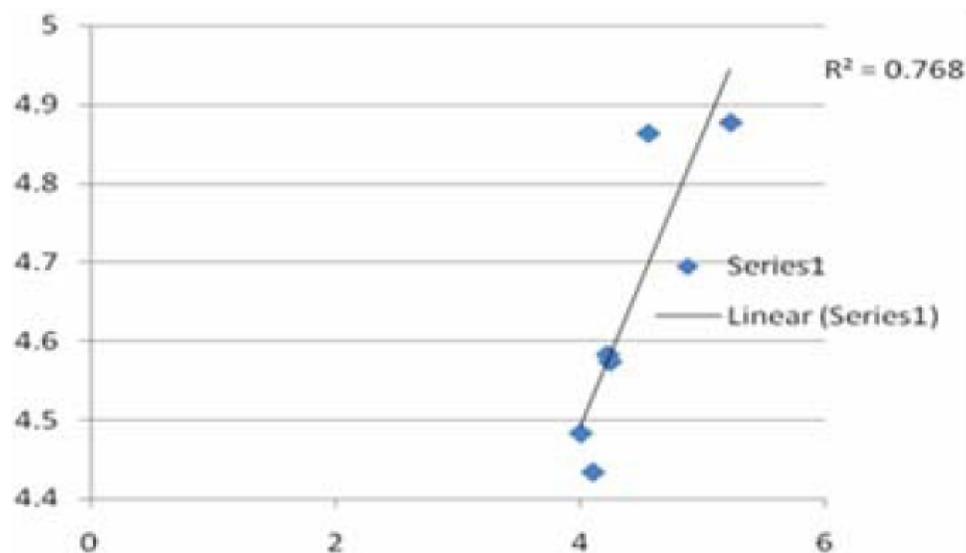


Figure 1 : Graph of actual vs. predictive activity for the training set of compounds derived from MLR analysis

working (FFNN) was set to close to prevent the problems of over fitting or memorizing data. The results were visualized on a 2D plot of output node against input (dependency graph) The Feed Forward Neural Networking (FFNN) architecture was set to no of cycles 1633. The output values for these data were predicted simultaneously as the training progressed and the error reported. The best RMS was 0.053 and test RMS was 0.088. The (FFNN) R2 of training set 0.954 R2 of test set 0.583.

## RESULT AND DISCUSSION

After performing the Multiple Linear Regression and Partial Least Square analysis on training set, regression equation relating biological activity with possible combination of the descriptors were obtained.

### Multiple linear regression analysis of Intigrase in-

### hibitor

For the data set of Intigrase inhibitor derivatives, the final model generated using Multiple Linear Regression analysis included 2 independent variables. The best model generated using Multiple Linear Regression analysis, after deleting 6 potential outliers had  $r^2$  value of 0.81 and  $r^2$  (CV) of 0.66  $r^2$  is a measure of goodness of prediction of the model. Value of the  $r^2$  closer to 1.0 indicated good correlation while the value of  $r^2$  (CV) above 0.6 indicated the good internal predictive capability of the developed model. Many other statistical tests were also performed on the training set to assure that the model formed is sound. For example, F test indicates the degree of statistical confidence so its higher value 17.38 depicts that the model formed is statistically significant. The standard deviation of the data  $s$ , show how far the activity value are spread about their average. Its lower value (0.1828) indicates that the model is sound. The

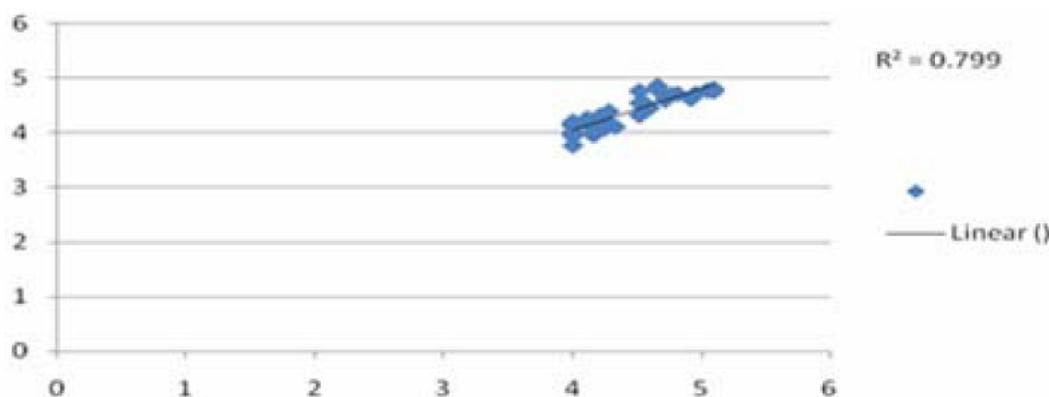


Figure 2 : Graph of actual vs. predictive activity for the test set of compounds derived from MLR analysis

## Full Paper

usefulness of the model was also evaluated by checking its statistical stability using predictive and residual sum of square. The values of these tests were 0.802652 and 1.43075 respectively. The QSAR equation obtained after MLR analysis is a linear model which related variations in biological activity to variations in the value of computed properties for a series of compounds.

### Partial least square

PLS analysis was also performed on the same data

set to check the soundness of model. The result obtained after PLS were comparable to that the MLR analysis as depicted in the graphs of actual activity vs predicted activity for the training set of compound (Figure 3) and test set (Figure 4).

### Final equation of MLR and PLS analysis:

$$Y = 0.98832613 \cdot X_1 + 10.413659 \cdot X_2 + 0.63229358 \cdot X_3 - 25.6665 \cdot X_4 - 0.54472023 \cdot X_5 - 0.3817353 \cdot X_6 + 19.325712$$

Kierchiv4 (path/cluster) index (whole molecule)

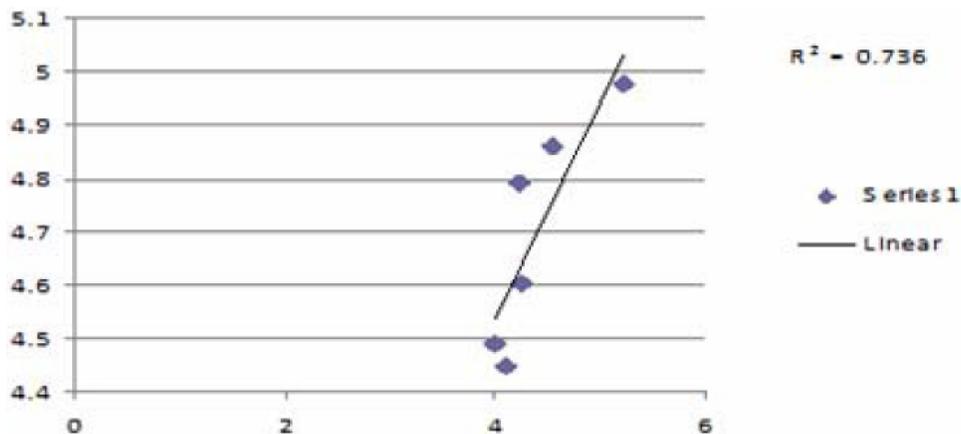


Figure 3 : Graph of actual vs. predictive activity for the training set of compounds derived from PLS analysis

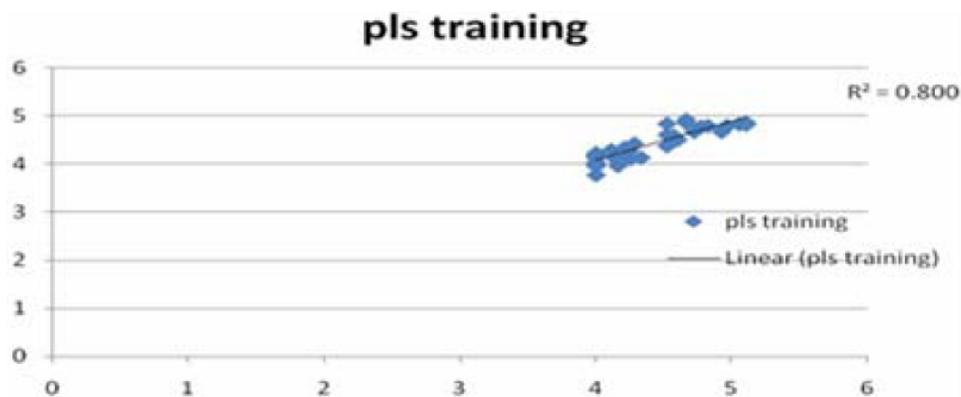


Figure 4 : Graph of actual vs. predictive activity for the test set of compounds derived from PLS analysis

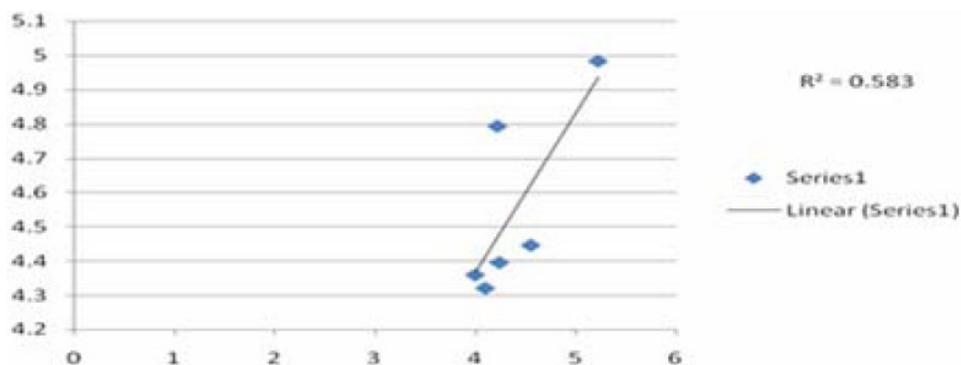


Figure 5 : Graph of actual vs. predictive activity for the test set of compounds derived from feed forward neural networking (FFNN) analysis

**TABLE 4 : Statistical test and their values obtained after performing partial linear regression analysis of nicotinic acid derivative**

Statistical Tests	Statistical significance	E statistic	(r) <sup>2</sup> value	r <sup>2</sup> (CV)
Values	0.8401	0.64463	0.77758	0.65799

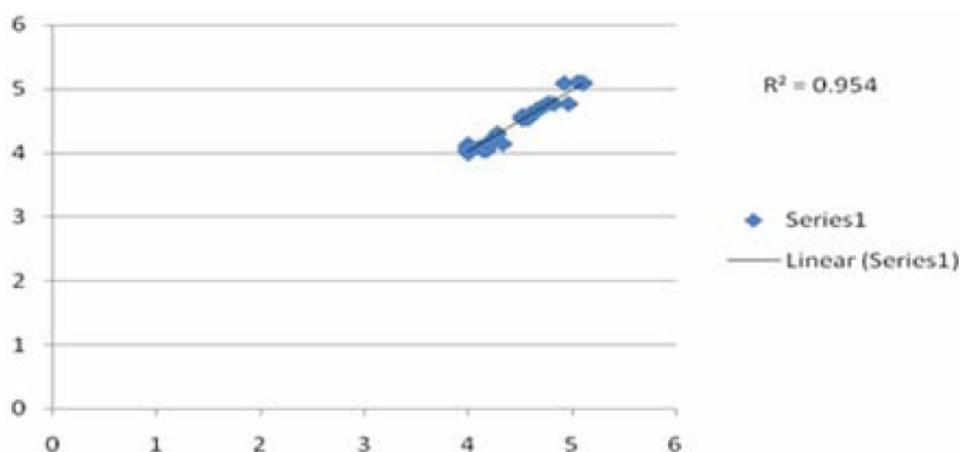
**TABLE 5 : Statistical test and their values obtained after performing neural network analysis of nicotinic acid derivative**

Cycles	Best RMS fit	Test RMS fit	R <sup>2</sup> of training set	R <sup>2</sup> if test set
1633	0.053132	0.088	0.954	0.583

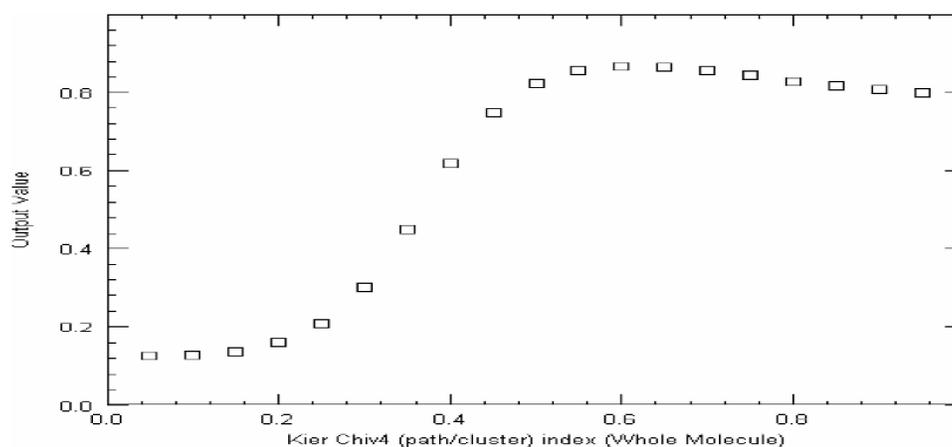
Inertia moment 1 length (whole molecule)  
 Number of H-bond donors (whole molecule)  
 VAMP heat of formation (whole molecule)

### Neural network

Feed-forward neural network: Neural networking was performed on the same data set. The inputs for the neural network were the descriptors obtained above, while the outputs were the log 1/EC50 values. The statistic obtained for the FFNN treatment of the integrase inhibitor derivatives data were N = 33, input columns



**Figure 6 : Graph of actual vs. predictive activity for the training set of compounds derived from feed forward neural network (FFNN) analysis**



**Figure 7 : Dependency plots between biological activity and kierchiv4 (path/cluster) index (whole molecule) entered in the final model**

(descriptors) = 3, net configuration = 3-3-1 (3 input nodes, 3 processing nodes, 1 output node), with the training root mean square 0.053132 and test root mean square 0.088.

The values of r<sup>2</sup> obtained for training and test were 0.954 and 0.583 respectively. This indicated good predictive ability of the model obtained Figure 5 illus-

trates the actual log 1/EC50 (x axes) vs predicted log 1/EC50 (y axes) for FFNN model, for the training set of compounds. Whereas, Figure 6 shows the actual log 1/EC50 (x axes) vs predicted log 1/EC50 (y axes) for FFNN model, for the test set of compounds. The actual and predicted activities of training and test set of compounds obtained by FFNN analysis are given

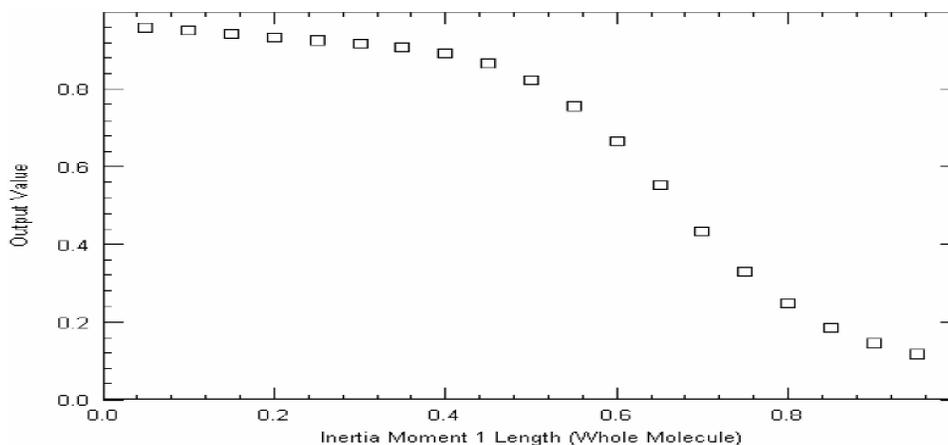


Figure 8 : Dependency plots between biological activity and inertia moment 1 length (whole molecule) entered in the final model

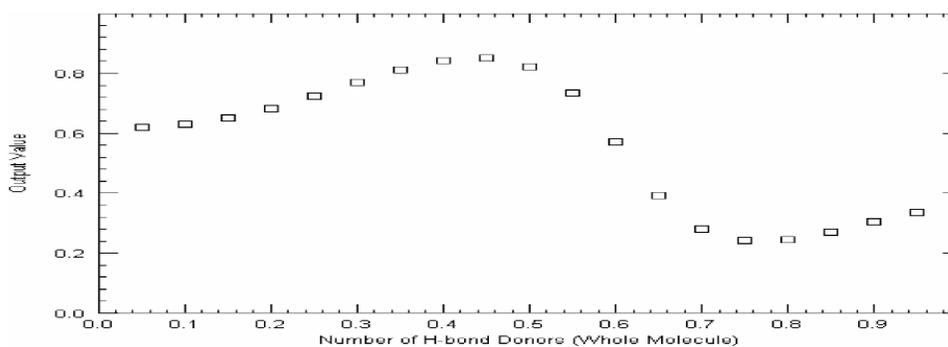


Figure 9 : Dependency plots between biological activity and number of H-bond donors (whole molecule) entered in the final model

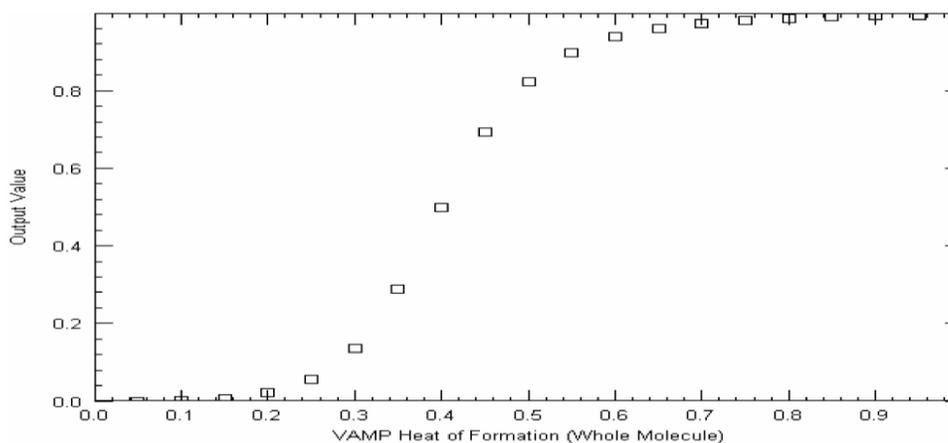


Figure 10 : Dependency plots between biological activity and VAMP heat of formation (whole molecule) entered in the final model

in TABLE 7.

The four descriptors entered in the final model were Kier chiv4 (path/cluster) index (whole molecule), Inertia moment 1 length (whole molecule), Number of H-bond donors (whole molecule) and VAMP heat of formation (whole molecule). These descriptors define to characterize the steric information and structural com-

plexity (cluster index), length, hydrogen character and chemical properties show the sub graph of the molecules.

Kier chi v4 (cluster index) (whole molecule) is a topological descriptor. It is the fourth order cluster valence connectivity index, which encodes steric information and structural complexity. It describes mainly

TABLE 6 : Actual and predicted activity of training set of compounds obtained by MLR

Compound	Actual activity pIC50	Predicted activity
1	4.92	4.618
2	4.52	4.3171
3	4.52	4.7535
4	4.52	4.5389
5	4.24	4.2213
8	4.82	4.7055
9	4.96	4.7055
11	4.57	4.5203
12	4.77	4.6915
13	4.66	4.8491
15	4.6	4.4511
16	4.72	4.595
17	5.1	4.7958
18	4	4.1234
19	4	3.9364
20	4.16	3.956
21	4.18	4.0661
23	4	3.9985
24	4.25	4.0754
25	4	3.9787
26	4.11	4.2562
28	4.28	4.3852
29	4.21	4.3036
30	4.66	4.7784
31	5.05	4.7655
33	5.10	4.755
34	4	4.1553
35	4.33	4.1057
36	4	3.7657
37	4.14	4.0977
38	4	4.1255
39	4	4.01
40	4	4.2
41	4	4.145
42	4	4.1611

TABLE 7 : Actual and predicted activity of test set of compounds obtained by MLR

Compound	Actual activity pIC50	Predicted activity
7*	4.24	4.5739
10*	4.55	4.8626
14*	4	4.4834
22*	4.22	4.5828
27*	4.10	4.4341
32*	5.22	4.8762

structural properties, such as the extent or degree of branching in a molecule. Kier chi v4 cluster index is highly sensitive to change in branching and their value rapidly increases the degree of branching. Kier chi v4

TABLE 8 : Actual and predicted activity of training set of compounds obtained by PLS

Compound	Actual activity pIC50	Predicted activity
1	4.92	4.6724
2	4.52	4.3753
3	4.52	4.8322
4	4.52	4.6056
5	4.24	4.2744
8	4.82	4.7882
9	4.96	4.7882
11	4.57	4.5848
12	4.77	4.7738
13	4.66	4.9148
15	4.6	4.5046
16	4.72	4.6682
17	5.1	4.8548
18	4	4.1411
19	4	3.9581
20	4.16	3.9646
21	4.18	4.0825
23	4	4.0157
24	4.25	4.1041
25	4	3.9781
26	4.11	4.2873
28	4.28	4.4229
29	4.21	4.3391
30	4.66	4.8582
31	5.05	4.8248
33	5.10	4.8154
34	4	4.18
35	4.33	4.1307
36	4	3.771
37	4.14	4.116
38	4	4.1482
39	4	4.0308
40	4	4.2268
41	4	4.1691
42	4	4.1882

TABLE 9 : Actual and predicted activity of test set of compounds obtained by PLS

Compound	Actual activity pIC50	Predicted activity
7*	4.24	4.6042
10*	4.55	4.859
14*	4	4.4914
22*	4.22	4.7915
27*	4.10	4.4502
32*	5.22	4.9762

## Full Paper

**TABLE 10 : Actual and predicted activity of training set of compounds obtained by feed forward neural networking (FFNN)**

Compound	Actual activity pIC50	Predicted activity
1	4.92	5.0816
2	4.52	4.5428
3	4.52	4.53
4	4.52	4.5793
5	4.24	4.2091
8	4.82	4.763
9	4.96	4.763
11	4.57	4.5328
12	4.77	4.7729
13	4.66	4.6583
15	4.6	4.6125
16	4.72	4.7294
17	5.1	5.0786
18	4	4.0416
19	4	4
20	4.16	4.0472
21	4.18	4.0524
23	4	4.0766
24	4.25	4.216
25	4	4.0351
26	4.11	4.0962
28	4.28	4.3054
29	4.21	4.1894
30	4.66	4.6732
31	5.05	5.0972
33	5.10	5.09
34	4	4.0964
35	4.33	4.143
36	4	4.0808
37	4.14	4.0505
38	4	4.0573
39	4	4.0734
40	4	4.0465
41	4	4.0988
42	4	4.1361

(cluster index) (whole molecule) is positively correlated with the biological activity in the regression equation, this shows that substitution in lead compound with substituent which have higher branching and atom connectivity will lead to increase in biological activity.

Inertia moment 1 length subst. 1 is the principal mo-

**TABLE 11 : Actual and predicted activity of test set of compounds obtained by feed forward neural networking (FFNN)**

Compound	Actual activity pIC50	Predicted activity
7*	4.24	4.3958
10*	4.55	4.4462
14*	4	4.3597
22*	4.22	4.7939
27*	4.10	4.321
32*	5.22	4.9837

ment of inertia of a molecule. This descriptor depends only on the principal moments of inertia of a molecule. The moment of inertia is a geometrical descriptor which characterizes the mass distribution in a molecule and the susceptibility of the molecule to different rotational transitions. As inertia moment length subst. 1 is positively correlated with the biological activity in the regression equation, this shows that substitution in lead compound with substituents which have higher mass of distribution or moment of inertia will lead to increase in biological activity.

Number of H-bond donors (whole molecule) is the electromagnetic attractive interaction of a polar hydrogen atom in a molecule or chemical group and an electronegative atom, such as nitrogen, oxygen or fluorine, from another molecule or chemical group. It is not a covalent chemical bond. Number of H-bond donors (whole molecule) is negatively correlated with the biological activity in the regression equation; this shows that hydrogen bonding in lead compound with substituents which have low number of H bond donors will lead to increase in biological activity.

VAMP heat of formation (whole molecule) shows the chemical properties. VAMP heat of formation (whole molecule) is negatively correlated with the biological activity in the regression equation; this shows that chemical properties in lead compound with substituents which have low chemical properties will lead to increase in biological activity.

## CONCLUSION

The niacin receptor agonist is an important target for anti-integrase therapy with huge therapeutic potential for their agonist. In the present QSAR study, multiple linear regression (MLR), partial least square

(PLS) and feed forward neural network (FFNN) analysis were used to develop the QSAR models and to predict the biological activity from computationally derived molecular descriptors for a set of compound of integrase inhibitor derivatives. The results obtained from the MLR and PLS and FFNN analysis was comparable and indicated good predictive ability of the model obtained. Also, considering the facts that the QSAR model was able to reproduce the experimental facts, it was validated by the appropriate statistical procedures and it generated important information about the geometrical and structural requirements for the desired biological activity, it can be used for the effective design of new, more potent integrase inhibitor. The QSAR approach to design has much to offer to the medicinal chemist. They are increasingly used to predict a wide range of activities and toxicities of drugs and environmental pollutants. However, it requires, for successful application, reliable bioactivity data for a carefully selected training set of compounds from which the QSAR model is developed and proper validation to the reliability of the prediction. Validation through external test set was also performed to check the predictive ability of the developed model. A good R<sup>2</sup> value of 0.81 was obtained for the test set, by MLR analysis.

### ACKNOWLEDGEMENT

Authors are thankful to Banasthali University, Banasthali India for providing funds for software required in this study.

### REFERENCES

- [1] B. Autran, G. Carcelain, T.S. Li, C. Blanc, D. Mathez, R. Tubiana et al.; Positive effects of combined antiretroviral therapy on CD4<sup>+</sup> T cell homeostasis and function in advanced HIV disease. *Science*, **277**(5322), 112-6 (1997).
- [2] N.G. Pakker, D.W. Notermans, R.J. de Boer, M.T. Roos, F. de Wolf, A. Hill et al.; Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nat Med.*, **4**(2), 208-14 (1998).
- [3] J.K. Wong, H.F. Gunthard, D.V. Havlir, Z.Q. Zhang, A.T. Haase, C.C. Ignacio et al.; Reduction of HIV-1 in blood and lymph nodes following potent antiretroviral therapy and the virologic correlates of treatment failure. *Proc. Natl. Acad. Sci.*, **94**(23), 12574-9 (1997).
- [4] Z.Q. Zhang, D.W. Notermans, G. Sedgewick, W. Cavert, S. Wietgreffe, M. Zupancic et al.; Kinetics of CD4<sup>+</sup> T cell repopulation of lymphoid tissues after treatment of HIV-1 infection. *Proc. Natl. Acad. Sci.*, **95**(3), 1154-9 (1998).
- [5] J.A. Grobler, K. Stillmock, B. Hu, M. Witmer, P. Felock, A.S. Espeseth et al.; Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci.*, [www.pnas.org/cgi/doi/10.1073/pnas.092056199](http://www.pnas.org/cgi/doi/10.1073/pnas.092056199), **99**(10), 6661-6 (2002).
- [6] C. Marchand, A.A. Johnson, R.G. Karki, G.C. Pais, X. Zhang, K. Cowansage et al.; Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S). *Mol. Pharmacol.*, **64**(3), 600-9 (2003).
- [7] D.J. Hazuda, S.D. Young, J.P. Guare, N.J. Anthony, R.P. Gomez, J.S. Wai et al.; Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science*, **305**(5683), 528-32 (2004).
- [8] J.A.O. Meara, C. Yoakim, P.R. Bonneau, B.S. Michael, M.G. Cordingley, R. Deziel, L. Doyon, J. Duan, M. Garneau, I. Guse, S. Landry, E. Malenfant, J. Naud, W.W. Ogilvie, B. Thavonekham, B. Simoneau; Novel 8-Substituted Dipyridodiazepinone Inhibitors with a Broad-Spectrum of Activity against HIV-1 Strains Resistant to Non-nucleoside Reverse Transcriptase Inhibitors. *J. Med. Chem.*, **48**, 5580-5588 (2005).
- [9] T. Imamichi; Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Curr. Pharm. Des.*, **10**, 4039-4053 (2004).
- [10] S.A. Chow, K.A. Vincent, V. Ellison, P.O. Brown; *Science*, **255**, 723-726 (1992).
- [11] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**, 1373-1396 (2003).
- [12] H.T. Siegelmann, E.D. Sontag; Turing computability with neural nets. *Appl. Math. Lett.*, **4**, 77-80 (1991).