

## QSAR analysis of some phthalimide analogues based inhibitors of HIV-1 integrase

Ruchi Bansal,<sup>a,†</sup> C. Karthikeyan,<sup>b,†</sup> N S Hari Narayana Moorthy,<sup>b</sup> and Piyush Trivedi<sup>b,\*</sup>

<sup>a</sup> Department of Pharmacy, Shri G.S. Institute of Technology and Science, 23, Park Road, Indore-452003, Madhya Pradesh, India

<sup>b</sup> School of Pharmaceutical Sciences, Rajiv Gandhi Technological University, Airport Bypass Road, Gandhi Nagar, Bhopal-462036, Madhya Pradesh, India

E-mail: [drtrivedislab@yahoo.co.in](mailto:drtrivedislab@yahoo.co.in), [nshnm@yahoo.co.in](mailto:nshnm@yahoo.co.in)

---

### Abstract

A quantitative structure activity relationship (QSAR) study has been performed on phthalimide analogues based inhibitors of HIV-1 integrase to understand the structural features influencing the affinity of these inhibitors towards the enzyme. The compounds in the selected series were characterized by molecular descriptors calculated using the QSAR software Dragon and molecular modeling software ChemOffice 2001. QSAR models were derived by stepwise multiple regression analysis employing the method of least squares. The best QSAR model describing the HIV-1 integrase inhibitory activity of phthalimide analogues was selected on the basis of statistical significance and predictive ability as gauged by cross-validation procedure and external test-set method. The generated QSAR models revealed that increased HIV-1 inhibitory potency of tricyclic phthalimide derivatives could be achieved by increasing the overall lipophilicity of the molecules and by incorporating halogen substituents in the benzylic aromatic ring attached to the phthalimido nitrogen atom. Additionally, the model also suggests that the increase in the molecular flexibility by incorporation of rotatable bonds is conducive for HIV-1 inhibitory activity of phthalimide derivatives whereas increase in molecular branching appears to be detrimental to the activity.

**Keywords:** QSAR, AIDS, HIV-1 integrase, phthalimide, Dragon

---

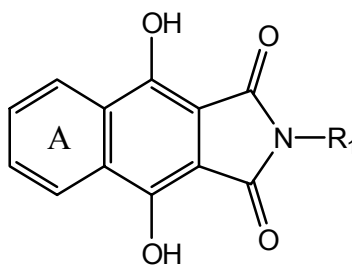
<sup>†</sup> These two authors contributed equally to this work.

## Introduction


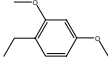

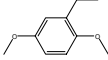

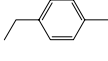

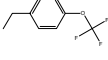

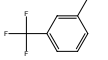

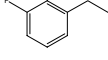

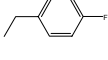

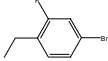

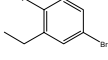

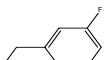

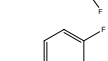

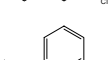

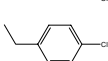

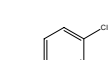



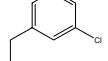

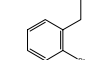

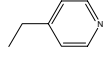

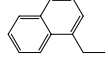

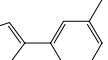

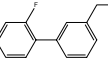

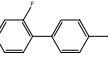
The alarming spread of the acquired immuno-deficiency syndrome (AIDS) epidemic has stimulated the discovery of therapeutic agents to inhibit the replication of the causative virus, human immunodeficiency virus (HIV).<sup>1</sup> The advanced understanding of the viral cell cycle has made it possible to define the targets to interrupt the life cycle of the virus. Among them, one such target is the viral integrase (IN), which is responsible for the integration of proviral DNA into the host cell DNA. It catalyzes two distinct reactions: terminal cleavage at each 3' end of the proviral DNA removing a pair of bases and strand transfer which results in the joining of each 3' end to 5' phosphates in the target DNA.<sup>2</sup> As these biochemical reactions are essential for the life cycle of the virus, integrase represents an attractive target for treatment of HIV infections. What makes IN an especially attractive target for drug design is that there are no known mammalian counterparts to this enzyme, so toxicity is expected to be rather low. Furthermore, integration is an essential part of the viral replication cycle and the IN region of the pol gene is more conserved than either the RT or PR coding regions.<sup>3</sup> Therefore, the inhibition of IN catalytic activities offers a promising anti-retroviral drug target. While a large number of compounds that inhibit integrase have already been identified,<sup>4</sup> only a handful displayed antiviral activity in cell culture. Currently there are no approved drugs that target IN. However, to date, six integrase inhibitors are under preclinical or clinical studies in AIDS patients namely AR-188, S-1360, L-880,812, JTK-303, FZ-41, MK-0518.<sup>5,6</sup> As more IN inhibitors enter human drug trials, there is a growing need for the design of novel lead compounds with diverse structural scaffolds and promising pharmacokinetic properties to counteract the difficulties observed with first-generation IN inhibitors.<sup>7</sup> Considering the recent interest in HIV-1 integrase inhibitors and for progression of design and development of such inhibitors, a Quantitative Structure-Activity Relationship (QSAR) investigation of novel series of HIV-1 inhibitory phthalimide analogues reported by Verschuere *et al*<sup>8</sup> is carried out. A QSAR study is performed on these series in order to analyze the physicochemical and structural requirements of these inhibitors to exhibit optimal inhibitory potency of HIV-1 integrase enzyme which will in turn help in rationalizing the design of these molecules as integrase inhibitors.






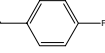

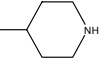
## Experimental Section

**Dataset:** The dataset consist of structurally diverse compounds reported for HIV-1 integrase inhibitory activity. The selected series comprises of forty-two tricyclic phthalimide analogues reported by Verschuere *et al*.<sup>8</sup> (Table 1). The biological activity values of two compounds (41 and 42) in the series are not well defined; hence they were not used in QSAR modeling. Further, Compound 9 and 40 were stereoisomers, since the applied descriptors are unable discriminating stereoisomers, only one of the two compounds preferentially most potent (Comp. No. 9) was considered for QSAR study.

**Table 1.** Structural modification and HIV-1 integrase inhibition data of tricyclic phthalimide analogues

Compound number	Ring A	R <sub>1</sub>	Inhibition of HIV-1 integrase enzyme (pIC <sub>50</sub> )
1			6.42
2			6.59
3			5.44
4			6.68
5		-CH <sub>3</sub>	4.31
6			4.98
7			5.62
8			5.66
9			5.98
10			5.00
11			6.25
12			5.85
13			6.09
14			6.25
15			5.70
16			6.37

17			6.37
18			6.11
19			5.80
20			5.69
21			6.37
22			6.25
23			6.389
24			6.36
25			6.45
26			6.43
27			6.75
28			6.66
29			6.38
30			6.95
31			6.73
32			6.20
33			5.17
34			5.85
35			6.17
36			6.30
37			6.00
38			5.84

39			5.69
40			5.69
41			< 4.00
42			< 4.00

The HIV-1 integrase inhibitory activity of compounds in the series is reported as  $\text{pIC}_{50}$  values where  $\text{IC}_{50}$  refers to experimentally determined concentration required to inhibit 50% of integrase strand transfer activity. The compounds in the selected series were randomly divided into two sets with 31 compounds used as a training set in developing regression models and the remaining 8 as validation set in the prediction of biological activity.

The molecular structures of the compounds in selected series were sketched using Chem Draw ultra module of CS ChemOffice 2001<sup>9</sup> molecular modeling software. The sketched structures were then transferred to Chem3D module for generation of three dimensional structure (3D). The geometries of generated 3D structures were pre-optimized using MM2 force field as implemented in the Chem3D module of CS ChemOffice 2001. These molecular geometries were refined using the quantum chemical program package MOPAC 6.0 applying the AM1 parameterization together with eigenvector following geometry optimization procedure. The gradient norm 0.001 kcal/Å was used to calculate electronic, geometric and energetic parameters for the isolated molecules. The optimized geometries of the molecules were used to compute the necessary quantum chemical descriptors available in the MOPAC server of Chem3D module. Further, the molecular output was also used for the calculation of some selected descriptors available in the software DRAGON.<sup>10</sup> The molecular descriptors employed in the present study are summarized in Table 2.

Variable-selection for the QSAR modeling was carried out by stepwise linear regression method using statistical program SYSTAT (version 10.2).<sup>11</sup> The program employs a stepwise technique, *i.e.*, only one parameter at a time was added to a model and always in the order of most significant to least significant in terms of F-test values. Statistical parameters were calculated subsequently for each step in the process, so the significance of the added parameter could be verified. The goodness of the correlation is tested by the regression coefficient ( $R^2$ ), the F-test and the standard error of estimate (SEE). The t-test and the level of significance of each coefficient, as well as the confidence limits of the regression coefficient, are also reported. The squared correlation coefficient (or coefficient of multiple determination),  $R^2$ , is a measure of the fit of the regression model. Correspondingly, it represents the part of the variation in the observed (experimental) data that is explained by the model. The correlation coefficient values closer to 1.0 represent the better fit of the model. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the model (*i.e.*, the variance not explained by the model). High values of the F-test indicate that the model is statistically

significant. The standard error is measured by the error mean square,  $s^2$ , which expresses the variation of the residuals or the variation about the regression line. Thus, the standard error measures the model error. If the model is correct, it is an estimate of the error of the data variance. The t-test measures the statistical significance of the regression coefficients. The higher t-test values correspond to the relatively more significant regression coefficients.

To further validate the model, other tests were performed for the descriptors, the pairwise correlations and the variance inflation factors (VIF). The VIF values, defined as  $(1-R^2)^{-1}$ , were calculated to identify whether excessively high multicollinear coefficients existed among the descriptors; a VIF greater than 10 is indicative of multicollinearity. The Z-score method was adopted for the detection of outliers. Z-score can be defined as absolute difference between the value of the model and the activity field, divided by the square root of the mean square error of the data set. Any compound which shows a value of Z-score higher than 2.5, during generation of a particular QSAR model is considered as outlier

The model, which passed the statistical diagnosis with as few descriptors as possible was chosen. When adding of another descriptor in stepwise addition did not improve significantly the statistics of a model, it was determined that the optimum subset of descriptor for QSARs had been achieved.

Besides deriving quantitative models of statistical significance, an important aspect of QSAR modeling is validating the model since a good statistical fit does not guarantee the predictive ability of the model. In view of above, the internal consistency of the selected models was assessed by cross-validation method following a leave-one-out scheme using the in-house program VALSTAT.<sup>12</sup> In this method, one data point is systematically deleted from the dataset and a QSAR model is constructed on the basis of reduced dataset and the model is subsequently used to predict the removed data point. The procedure was repeated until a completed set of predicted values is generated. The validation parameters calculated are squared cross-correlation coefficient ( $Q^2$ ), standard deviation of sum of square of difference between predicted and observed values ( $S_{PRESS}$ ) and standard deviation of error of prediction ( $S_{DEP}$ ).  $Q^2$  values greater than five and low  $S_{PRESS}$  and  $S_{DEP}$  values ( $<0.5$ ) can be considered as a proof of the high predictive ability of the QSAR models.

Finally, the derived QSAR models were used for the prediction of the activity values of the compounds in the test set and the external validation parameter, predictive  $r^2$  ( $r^2_{pred}$ ) was calculated for evaluating the predictive capacity of the model. A value of  $r^2_{pred}$  greater than 3 indicates the good predictive capacity of the QSAR model.

**Table 2.** Classification and description of the calculated molecular descriptors

Functional families of the descriptors	Descriptor definition
Constitutional descriptors	Number of rotatable bonds (RBN), Rotatable bond fraction (RBF), Number of double bonds (nDB), Number of triple bonds (nTB) and Number of aromatic bonds (nAB).
Topological descriptors	Number of Nitrogen atoms (nN), Number of Oxygen atoms (nO), Number of Sulfur atoms (nS), Number of Fluorine atoms (nF), Number of Chlorine atoms (nCl), Number of Bromine atoms (nBr), Molecular weight (MW), Connectivity index chi-0 ( ${}^0\chi$ ), Connectivity index chi-1 ( ${}^1\chi$ ), Connectivity index chi-2 ( ${}^2\chi$ ), Connectivity index chi-3 ( ${}^3\chi$ ), Connectivity index chi-4 ( ${}^4\chi$ ), Connectivity index chi-5 ( ${}^5\chi$ ), Valence connectivity index chi-0 ( ${}^0\chi_v$ ), Valence connectivity index chi-1 ( ${}^1\chi_v$ ), Valence connectivity index chi-2 ( ${}^2\chi_v$ ), Valence connectivity index chi-3 ( ${}^3\chi_v$ ), Valence connectivity index chi-4 ( ${}^4\chi_v$ ), Valence connectivity index chi-5 ( ${}^5\chi_v$ ), First Zagreb index M1 (ZM1), Second Zagreb index M2 (ZM2), Wiener W index (W), Balaban distance connectivity index (J), Kier symmetry index (S0K), 1-path Kier alpha-modified shape index (S1K), 2-path Kier alpha-modified shape index (S2K), 3-path Kier alpha-modified shape index (S3K), Kier flexibility index (PHI), Balaban-type index from Z weighted distance matrix (Barysz matrix) (JhetZ), Balaban-type index from mass weighted distance matrix (JhetM), Balaban-type index from van der Waal's weighted distance matrix (JhetV), Balaban-type index from electronegativity (JhetE), Eccentricity (ECC) and Distance/Detour index (D/D).
Functional group counts	Number of Primary carbon atoms ( $sp^3$ ) (nCp), Number of secondary carbon atoms ( $sp^2$ ) (nCs), Number of tertiary carbon atoms ( $sp^3$ ) (nCt), Number of aromatic ethers (nRORPh), Number of sulfonamide (nSO <sub>2</sub> N), Number of unsubstituted aromatic C( $sp^2$ ) (nCaH), Number of substituted aromatic C( $sp^2$ ) (nCaR), Number of halogen atoms on aromatic ring (nPhX), Number of hydrogen bond acceptor atoms (nHacc), Number of hydrogen bond donor atoms (nHdon), Number of ring secondary C( $sp^3$ ) (nCrH2) and Number of ring tertiary C( $sp^3$ ) (nCrHR).
Molecular properties	Molecular refractivity (MR), Topological Polar surface area (PSA), log of the octanol/water partition coefficient (MlogP), Unsaturation index (Ui) and Hydrophilic factor (Hy).
Quantum chemical descriptors	Dipole moment (AM1_Dipole), Energy of highest occupied molecular orbital (AM1_HOMO), Energy of lowest unoccupied molecular orbital (AM1_LUMO), Repulsion Energy (AM1_RE), Total energy (AM1_E), Electronic energy (AM1_Eele) and Heat of formation (AM1_HF).

## Results and Discussion

Among all possible multiple combinations of describing parameters obtained for modeling HIV-1 integrase inhibitory activity of tricyclic phthalimide analogues, the best statistical accuracy was reached with four descriptors. The best tetra- parametric model generated for the series is summarized below, along with the statistical parameters.

$$pIC_{50} = [6.31 (\pm 1.15)] + RBF [11.6 (\pm 4.34)] + MlogP [0.23 (\pm 0.19)] + nPhX [0.43 (\pm 0.18)] + Jhete [-0.96 (\pm 0.56)]. \quad (1)$$

$N=31$ ,  $R=0.87$ ,  $R^2 = 0.76$ ,  $SEE=0.29$ ,  $F=20.15$ ,  $P < 0.001$ ,  $Q^2 = 0.63$ ,  $S_{PRESS} = 0.36$ ,  $S_{DEP} = 0.33$ ,  $r^2_{pred} = 0.61$ .

Compound **3** behaved as an outlier ( $Z$  score =  $-2.76$ ), and hence was removed, and the resultant model was:

$$pIC_{50} = [6.24 (\pm 1.01)] + RBF [11.69 (\pm 3.8)] + MlogP [0.24 (\pm 0.16)] + nPhX [0.44 (\pm 0.16)] + Jhete [-0.92 (\pm 0.49)] \quad (2)$$

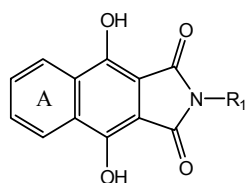
$N = 30$ ,  $R = 0.90$ ,  $R^2 = 0.81$ ,  $SEE = 0.25$ ,  $F = 27.19$ ,  $P < 0.001$ ,  $Q^2 = 0.66$ ,  $S_{PRESS} = 0.34$ ,  $S_{DEP} = 0.31$ ,  $r^2_{pred} = 0.62$ .

In the QSAR models given above,  $N$  is the number of data points,  $R$  is correlation coefficient,  $R^2$  is squared correlation coefficient,  $SEE$  is standard error of estimate,  $F$  represents Fischer ratio between the variances of calculated and observed activities,  $P$  is the probability value. The figures given in the parentheses with  $\pm$  sign in the model are 95% confidence limits. The statistical quantities  $Q^2$ ,  $S_{PRESS}$ ,  $S_{DEP}$  are based on the leave-one-out method and correspond to cross-validated squared correlation coefficient, standard deviation based on the predicted residual sum of squares and standard deviation of error of prediction respectively. The molecular descriptors used in the selected QSAR models are defined in Table 2. The values of descriptors in QSAR models derived for the series are tabulated in Table 3. Table 4 records the  $t$ -values and VIF values of the descriptors in the selected models derived. Pair wise correlations between the descriptors used in the QSAR models obtained for series is summarized in Table 5. The Tables 6 and 7 represent the  $pIC_{50}$  estimates of compounds in the training and test set of series obtained using model 1 and 2 along with the experimental data.


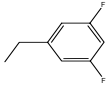

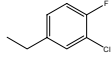

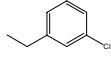

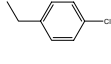

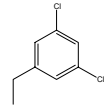

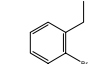

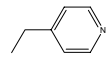

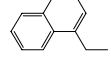

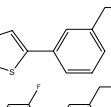

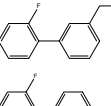

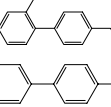

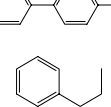

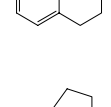

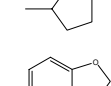

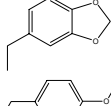

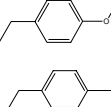

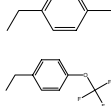

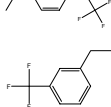

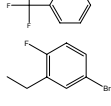

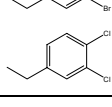

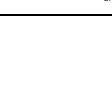


**Table 3.** Descriptors values used for the formulation of model 1

## Training Set



Compound Number	Ring A	R <sub>1</sub>	RBF	Jhete	nPhX	MlogP
1			0.1	2.087	1	1.72
2			0.12	2.096	1	1.945
3			0.11	2.094	1	0.731
4			0.11	2.097	1	0.731
5		-CH <sub>3</sub>	0.11	2.919	0	-0.969
7			0.06	1.605	0	0.329
8			0.09	2.012	0	0.597
9			0.12	2.171	0	0.614
10			0.05	1.986	0	-0.705
12			0.11	1.65	0	-0.235
13			0.14	2.102	0	-0.134
15			0.17	2.114	0	-0.641
16			0.17	2.081	0	-0.641
17			0.17	2.047	0	-0.641
18			0.17	2.077	0	-0.641
22			0.11	2.103	1	0.496
23			0.11	2.081	1	0.496
24			0.11	2.086	2	0.847

26			0.11	2.09	2	0.614
27			0.11	2.067	2	0.731
28			0.11	2.099	1	0.614
29			0.11	2.077	1	0.614
31			0.11	2.083	2	0.847
32			0.11	2.12	1	0.731
33			0.11	2.134	0	-0.615
34			0.09	1.847	0	1.515
35			0.09	1.747	0	1.157
36			0.08	1.75	1	1.595
37			0.08	1.677	1	1.595
38			0.08	1.679	0	1.489
39			0.120	1.963	0	0.164
Test Set						
6			0.05	2.257	0	0.104
11			0.10	1.799	0	-0.467
14			0.14	2.023	0	-0.134
19			0.12	2.071	0	0.614
20			0.14	1.907	0	0.213
21			0.12	2.005	0	0.961
25			0.11	2.106	2	0.847
30			0.11	2.064	2	0.847

**Table 4.** Descriptor *t*-values and Variance inflation factor (VIF) values for QSAR models

	Descriptors	t test	VIF
Model 1	Intercept	11.27316	-
	RBF	5.54529	1.3580
	Jhete	-3.55795	1.4416
	nPhX	4.83994	1.4342
	MlogP	2.52952	1.9420
Model 2	Intercept	12.73961	-
	RBF	6.34664	1.3580
	Jhete	-3.88932	1.4421
	nPhX	5.70848	1.4248
	MlogP	3.00267	1.9390

The statistical details of the QSAR model given above speak for its good statistical quality. The  $R^2$  value is above 0.8, which suggest that a good percentage of the total variance in biological activity is accounted by the model. Low value of standard error of estimate ( $< 0.3$ ) indicates the accuracy of the statistical fit. All the values of the *t*-statistic are significant which confirms the significance of each descriptor. The *F*-statistic (on 4 and 25 degrees of freedom) for this model is 26.7 (compared to the critical value of 4.62 at the 0.05 level of significance). The calculated *F* value for the generated for the QSAR model exceed the tabulated *F* value by large margin as desired for a meaningful regression. Furthermore, the calculated *F* value also determines a confidence limit superior to 99% for this model. The correlation matrix given by Table 5 and the variance inflation factors values less than 5 indicates the absence of multicollinearities in the model. The stability of model 2 as judged by leave-one-out procedure is fairly good ( $Q^2 > 0.6$ ) suggesting that the models will be useful for meaningful predictions. Further support in this regard is obtained from the low values of the cross-validation parameters  $S_{PRESS}$  and  $S_{DEP}$ . Very good agreement between experimental data and model computation is achieved using Model 1 as expressed in Table 3 and Figure 1. Furthermore, the predictive potential of model is good as judged by the  $r^2_{pred}$  value = 0.62. The predicted activity values for the compounds in the test set, along with their corresponding experimental activity values, are recorded in Tables 6 and 7. The predicted  $pIC_{50}$  values of the compounds in the test set are in agreement with the corresponding experimental values (Figure 2).

**Table 5.** Correlation matrix for the descriptors used in Model 1

	pIC <sub>50</sub>	RBF	Jhete	nPhX	MlogP
pIC <sub>50</sub>	1.00	0.24	-0.27	0.60	0.47
RBF	0.24	1.00	0.38	-0.12	-0.47
Jhete	-0.27	0.38	1.00	0.09	-0.42
nPhX	0.60	-0.12	0.09	1.00	0.45
MlogP	0.47	-0.47	-0.42	0.45	1.00

The best tetra- parametric equation obtained for modeling HIV-1 integrase inhibitory activity of tricyclic phthalimide analogues comprises of the following descriptor terms MlogP,<sup>13</sup> RBF,<sup>14</sup> Jhete,<sup>15</sup> and nPhX.<sup>14</sup> The molecular descriptor MlogP refers to the Moriguchi log of the octanol/water partition coefficient of the molecule and is considered as a measure of lipophilicity of a molecule. The positive coefficient of the descriptor in model 2 suggests that increase in the overall lipophilicity of the molecule will in turn increase the HIV-1 integrase inhibitory activity of phthalimide derivatives. The constitutional descriptor RBF in the QSAR corresponds to rotatable bond fraction in the molecule. The positive term associated with the descriptor in model 2 indicates that fractional increase in the rotatable bonds in the molecule is conducive for the HIV-1 integrase inhibitory activity exhibited by phthalimide derivatives. The functional group count descriptor nPhX in the model represents the number of halogen atoms bonded to carbon atoms in the aromatic ring. The descriptor bears a positive coefficient in model 2, which suggest that increase in the number of halogen substituents in the aromatic ring will lead to corresponding increase in the HIV-1 integrase inhibitory potency of phthalimide derivatives. The last descriptor included into the model is topological descriptor Balaban-type index derived from electronegativity weighted distance matrix (Jhete). The descriptor Jhete carries a negative weight in the model 2, which suggests that molecular branching and presence of highly electronegative atoms in the molecule are disfavored for the HIV- 1 integrase inhibition phthalimide analogues.

Interestingly, compound number 3 behaves as a statistical outlier in case of the selected QSAR model. The outlying behavior of the compound may be owing to the fact that it incorporates pyridazinyl moiety in ring A (Table 1) and exhibits low HIV-1 integrase inhibitory potency despite the presence of benzylic ring with a bromine substitution.

**Table 6.** Leave- one- out predicted values for HIV-1 integrase inhibition (training set)

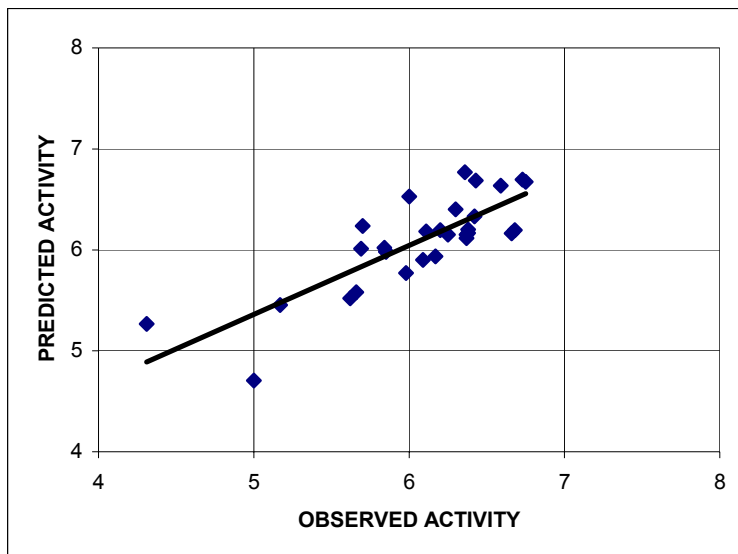
Compound Number	pIC <sub>50</sub>	Model 1	Model 2
1	6.42	6.27904	6.33018
2	6.59	6.57255	6.63344
3	5.44	6.2198	Outlier
4	6.68	6.1575	6.1937
5	4.31	5.12996	5.26611
7	5.62	5.52365	5.52058
8	5.66	5.5612	5.57983
9	5.98	5.74362	5.77037
10	5	4.69746	4.70561
12	5.85	5.99275	5.98607
13	6.09	5.88372	5.89939
15	5.7	6.22356	6.23672
16	6.37	6.10365	6.11503
17	6.37	6.14192	6.15167
18	6.11	6.17191	6.18316
22	6.25	6.11577	6.15101
23	6.38	6.13191	6.16595
24	6.36	6.7049	6.76713
26	6.43	6.62651	6.68695
27	6.75	6.61495	6.67325
28	6.66	6.12908	6.16429
29	6.38	6.16463	6.19954
31	6.73	6.63395	6.69375
32	6.2	6.15724	6.19567
33	5.17	5.44088	5.45403
34	5.85	5.95693	5.97888
35	6.17	5.92187	5.93421
36	6.3	6.3696	6.40205
37	6	6.4974	6.5276
38	5.84	6.00653	6.0198
39	5.69	5.99574	6.0123

**Table 7.** Predicted values for HIV-1 integrase inhibition (Test Set)

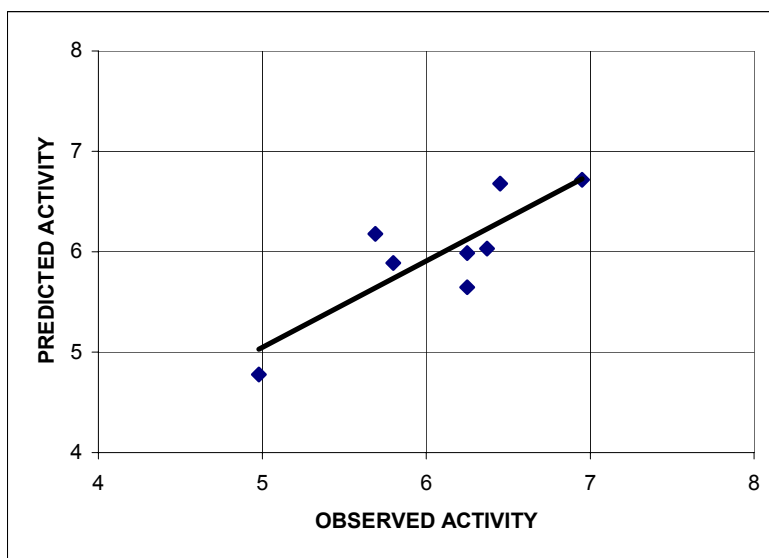
Compound Number	pIC <sub>50</sub>	Model 1	Model 2
<b>6</b>	4.98	4.75443	4.77729
<b>11</b>	6.25	5.64853	5.64747
<b>14</b>	6.25	5.97686	5.98801
<b>19</b>	5.8	5.86793	5.88754
<b>20</b>	5.69	6.16766	6.17701
<b>21</b>	6.37	6.01067	6.03055
<b>25</b>	6.45	6.62795	6.67883
<b>30</b>	6.95	6.66831	6.71745

From the molecular descriptors incorporated in the QSAR model, one may conjecture that molecular flexibility and hydrophobicity predominantly govern the integrase inhibitory activity of phthalimides under study. Molecular flexibility increases with the number of flexible bonds in the molecule and the importance associated with flexible bond might be owing to the fact that they play an important role in the orientation of pharmacophoric groups in the active site of the enzyme. Hydrophobic substituents in the molecule might influence enzyme-drug affinity through non-specific interactions with hydrophobic region in the active site of the enzyme. Furthermore, it appears that the halogen substitution in the phenyl ring plays a significant role molecule-enzyme affinity, a fact reflected in increased integrase inhibitory potency exhibited by molecules with halogen substituents in Table 1 (e.g., Compounds number 22 to 32). Additionally, the negative weight associated with the topological descriptor *j*hete emphasizes that presence of bulky groups and electronegative atoms in the molecule disfavors the HIV-1 integrase inhibitory affinity of the title compounds.

Summarizing the above discussion, the present study gives rise to QSARs with good statistical significance and predictive capacity for HIV-1 integrase inhibitory activity of phthalimide derivatives. For the dataset of 39 phthalimide analogues with well-defined HIV-1 integrase inhibitory activity, the HIV-1 integrase inhibitory potency appears to be influenced by structural components and overall lipophilicity of the molecule. The interpretation of the generated QSAR revealed that increased HIV-1 inhibitory potency of tricyclic phthalimide derivatives could be achieved by increasing the overall lipophilicity of the molecules and by incorporating halogen substituents in the benzylic aromatic ring attached to the phthalimido nitrogen atom. Additionally, the model also suggests that the increase in the molecular flexibility by incorporation of rotatable bonds is conducive for HIV-1 inhibitory activity of phthalimide derivatives whereas increase in molecular branching appears to be detrimental to the activity.



**Figure 1.** Scatter Plot between observed activity and predicted activity of Model 2 (training set)



**Figure 2.** Scatter Plot between observed activity and predicted activity of Model 2 (test set)

## Acknowledgements

One of the authors, Ms. Ruchi Bansal, likes to thank the All India Council for Technical Education, New Delhi for providing a fellowship. The authors wish to thank the Director, Shri G.S. Institute of Technology and Science, Indore for providing the necessary facilities for undertaking this research work.

## References and Notes

1. Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chemaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Byun-Vezinet, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. *Science* **1983**, *220*, 868.
2. Brown, P. O. *Curr. Top. Microbiol. Immunol.* **1990**, *157*, 19.
3. Meadows, D. C.; Gervay-Hague, J. *Chem. Med. Chem.* **2006**, *1*, 16.
4. Pommier, Y.; Neamati, N. *Adv. Virus Res.* **1999**, *52*, 427.
5. Pommier, Y.; Allison, A. J.; Christophe, M. *Nat. Rev. Drug Discov.* **2005**, *4*, 236.
6. Philippe, C. *Recent Patents on Anti-Infective Drug Discov.* **2006**, *1*, 1.
7. Dayam, R.; Sanchez, T.; Neamati, N. *Chem. Med. Chem.* **2006**, *1*, 238.
8. Verschuere, W. G.; Dierynck, I.; Amssoms, K. I. E.; Hu, L.; Boonants, P. M. J. G.; Pille, G. M. E.; Daeyaert, F. F. D.; Hertogs, K.; Surleraux, D. L. N. G.; Wigerinck, P. B. T. P. *J. Med. Chem.* **2005**, *48*, 1930.
9. *CS Chem Office* version 7.0, Cambridge Soft corporation, software publishers Association, 1730 M street, NW, Suite 700, Washington D.C. 20036, **2002**, 452.
10. DRAGON web version 3.0 is a QSAR software developed by Milano Chemometrics and QSAR Research Group, Dipartimento di Scienze dell'Ambiente e del Territorio Università degli Studi di Milano - Bicocca.
11. SYSTAT 10.2 version, Supplied by SYSTAT SOFTWARE INC (SSI).
12. Gupta, A. K.; Babu, M. A.; Kaskhedikar, S. G. *Indian J. Pharm. Sci.* **2004**, *66*, 396.
13. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, *Chem. Pharm. Bull.* **1992**, *40(1)*, 127.
14. Todeschini, R.; Consonni, V. Wiley-VCH: Weinheim, 2000.
15. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.