Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Archives of Applied Science Research, 2010, 2 (1) 134-142
(http://scholarsresearchlibrary.com/archive.html)

ISSN 0975-508X
CODEN (USA) AASRC9

Prediction of Anti-HIV Activity of Non-Nucleoside Inhibitors of Human Immuno Deficiency Virus -I Derivatives: Molecular Modelling Approach

Mukesh Chandra Sharma[*], Smita Sharma[a], D. V. Kohli[b], S. C. Chaturvedi[c]

**School of Pharmacy, Devi Ahilya Vishwavidyalaya, Indore (M.P), India*
[a] Department of Chemistry, Yadhunath Mahavidyalya, Bhind (M.P), India
[b] Department of Pharmaceutical Sciences, Dr. Hari Singh Gour University, Sagar (M.P), India
[c] Shri Arvindo Institute of Pharmacy, Ujjain Road, Indore (M.P), India

---

Abstract

In the present work, quantitative structure activity relationship studies were performed on a series of Thiotetrazole alkynyl acetanilide derivatives as anti-HIV agents. 2D QSAR studies through Multiple Linear Regression (MLR) and Partial Least Statistical (PLS) led to three statistically significant models for Non-Nucleoside Inhibitors of human immuno deficiency virus -I derivatives (all with r^2>0.75, F>> tabulated value, and chance correlation <0.001) having acceptable statistical quality and predictive potential as indicated by the value of cross validated squared coefficient (q^2>0.80).Stepwise multiple linear regressions analysis was applied to identify the structural and physicochemical requirements for anti-HIV activity, which was further evaluated for statistical significance and predictive power by internal and external validation.

Key words: 2D- QSAR; anti-HIV, Acetanilide derivatives.

Introduction

There has been no complete and successful chemotherapy [1,2] developed so far for the treatment of the acquired immunodeficiency syndrome (AIDS), which has been the most challenging worldwide medical problem[3] and also has become a major worldwide pandemic [4]. The human immunodeficiency virus type 1 (HIV-1), a retrovirus of the lentivirus family, has been recognized as the attractive causal agent of the AIDS [5]. Many studies have been devoted to anti-HIV drugs since the separation of HIV-I in 1983[6]. Acquired immune deficiency syndrome (AIDS) is caused by human immunodeficiency virus (HIV). Three essential enzymes, reverse transcriptase (RT), protease (PR) and integrase (IN), are required in the HIV life cycle. Although commercial drugs inhibiting RT and PR have been developed, their efficiency is limited due to side effects and drug resistance. Therefore, it is necessary to search for other

antiviral targets. This makes the HIV-1 IN to be an attractive target for new developments in anti-AIDS therapy. Drugs targeted to IN would be a valuable complement to RT and PR inhibitors. Combination therapy of RT and PR drugs has been proved to be effective in reducing viral load and HIV-1 mortality and morbidity .The ability of the DNA of HIV to become incorporated into the genome of the host is a hallmark of retroviral infection. This event is necessary and it ensures the indefinite survival of the virus in the host as long as the infected pool of cells remains alive. Integrase, which is coded for by the virus, is responsible for this event in the viral lifecycle [7, 8]. Two main categories of HIV RT inhibitors have been discovered to date. The first category of inhibitors is nucleoside analogues (e.g., AZT, 3TC, DDI, DDC) and the second category of inhibitors is nonnucleoside analogues. Nevirapine, delaviridine and efavirenz are the only non nucleoside reverse transcriptase inhibitors (NNRTI) that have received regulatory approval with several NNRTI (MKC442, Troviridine, S-1153, AG1549, PNU142721, ACT and HBY1293/GW420867X) are currently undergoing clinical trials. Efavirenz was the first potent anti-HIV drug to be approved by FDA and studies have shown that efavirenz penetrates into the cerebrospinal fluid, a common viral sanctuary. The therapeutic efficacy of the drug is mainly restricted due to the development of viral resistance associated with mutations that includes K103N, L1001 and Y188L [9].

Materials and Methods

The present study aimed at elucidating the structural features of benzyl amide-ketoacid derivatives required for HIV-integrase inhibition and to obtain predictive two and three-dimensional quantitative structure activity relationship (2D-QSAR, 3D-QSAR) models, which may guide the rational synthesis of novel inhibitors. This is accomplished by combining one of the stochastic search methods such as multiple regressions (MLR), partial least square regression (PLSR). The generated models would give insight to the influence of various interactive fields on the activity and thus, can help in designing and forecasting the inhibition activity of novel molecules.

Experimental Work

All the molecular modelling studies (2D QSAR) were performed using Molecular Design Suite supplied by the VLife Sciences, Pune on Compaq PC having Pentium IV processor and windows XP operating system. The structures were sketched using the 2D draw application and converted to 3D structures. Energy minimization and geometry optimization was conducted using Merck Molecular Force Field (MMFF) method with Root Mean Square (RMS) gradient set to 0.01 Kcal/mol Å and iteration limit to 10000.

1.1 Data set and molecular modelling (For 2D QSAR)

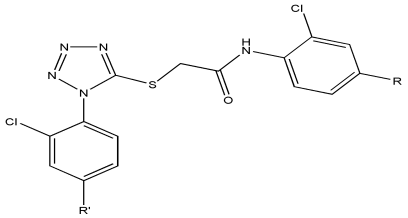
Computational work was performed by using the software Molecular Design Suite which was supplied by the V-Life Sciences, Pune (V-Life MDS 3.5) [10].The all twenty six compounds were manually divided into training set (seventeen compounds) and test set (nine compounds) according to Alexander Golbraikh et.al [11] who recommend that training and test sets must satisfy the following criteria (i) representative points of the test set must be close of those to the training set, (ii) representative points of the training set must be close to the representative points of the test set, and (iii) training set must be diverse. This approach resulted in selection of compound and as test set, and remaining twenty three compounds as training set. The unicolumn statistics of test and training sets, further reflected the right selection of test and training sets as maximum of training set was more than maximum of test set; and minimum of training set was less than the minimum of test set. This showed that the test set was interpolative i.e. derived within the minimum – maximum range of the training set. The average and standard deviation of

the training and test set provided insight to the relative difference of mean and point density distribution (along mean) of the two sets. The structures of thiotetrazole alkynyl acetanilide derivatives were constructed using the 2D draw application and converted to 3D structures by sending them to MDS. Energy minimization and geometry optimization was conducted using Merck Molecular Force Field (MMFF) method with Root Mean Square (RMS) gradient set to 0.01 Kcal/mol A^0 and iteration limit to 10000. The basis of energy minimization is that the drug binds to effectors/receptors in the most stable form i.e. minimum energy form. All the 2D descriptor like heat of formation, dipole moment, local charges, different alignment independent topological descriptors, elemental count including bromine count, fluorine count, path count, constitutional descriptors, chemical descriptors like molar volume, molecular surface area, hydrophobicity, hydrogen acceptor count, hydration energy and molecular polarizability were calculated for these geometrically optimized structures. The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute to QSAR. The whole twenty six compounds were subjected to regression analysis using MLR as model building method coupled with stepwise variable selection. QSAR equations were generated by using – 1og IC_{50} values as dependent variable and various parameter values as independent variables.

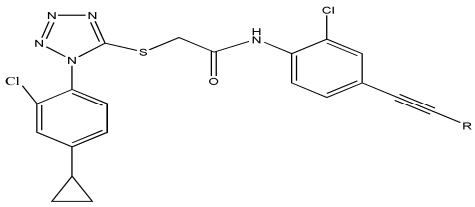
Regression analysis was carried out for cytotoxicity activity and the best model cross-validated. Cross correlation limit was set as 0.5, number of variable in final equation as 5, and term selection criteria as r^2, F-test 'in' as 4 and F-test 'out' as 3.99. Variance cut off was set to 0 and scaling as auto scaling, number of random iteration was set to 10. Following statistical parameters were considered to compare the generated QSAR models: correlation coefficient (r), squared correlation coefficient (r^2), predicted r^2 (pred_ r^2), and Fischer's value (F). In order to validate the generated QSAR models Leave One out (LOO) method was used indicated as value of q^2 (cross- validated explained variance) which is a measure of internal predictive ability of the model. The method resulted in three statistically significant models (Model 1, 2, and 3) considering the term selection criteria as r^2. The statistical significance of these models was further supported by the 'fitness plots' obtained for each model which is a plot of observed Vs predicted activity of training and test set compounds and provides an idea about how well the model was trained and how well it predicts the activity of external test set The nearness of observed to predicted activity presented will also aids to this fact.

The contribution charts for all the significant models are presented which gives the percentage contribution of the descriptors used in deriving the model. The descriptors used in deriving the QSAR models mentioned herein included both, physiological and alignment independent topological descriptors[12], Alignment independent descriptors are the molecular descriptors based upon a count statistic of the topological distance matrix where encoding of molecule is done by computing many selective count statistics (histograms) reflecting the distribution of different atom types and bond types in the molecule. The descriptors also incorporate geometric features of molecules by weighting the topological distance counts with the geometric distance. It is invariant to both translation and rotation. As a result, it does not require the alignment of the structures under study.

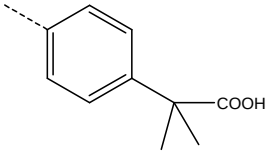
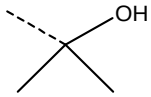
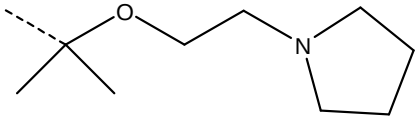
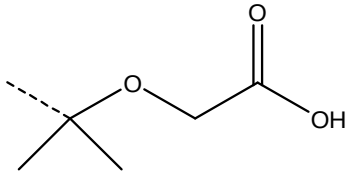
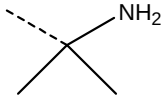
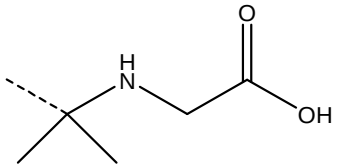
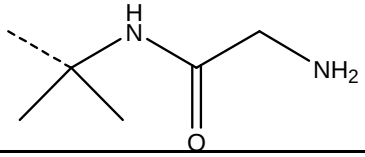
Table-1

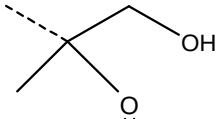
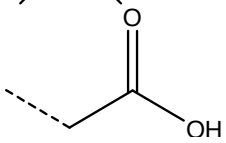
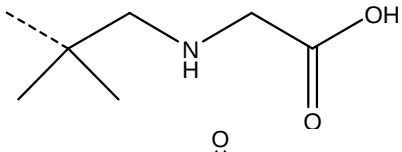
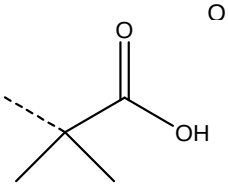
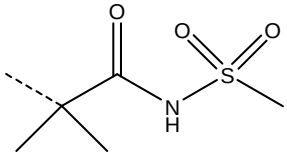
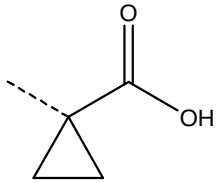
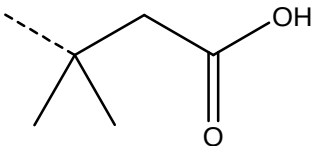


Com-1-10



Com-11-26

Compound	R'	R	IC_{50}(nM)	Log IC_{50}
1	t-Bu	COOH	5	0.69897
2	c-Pr	COOH	16	1.20412
3	t-Bu	SO_2NH_2	5.8	0.763428
4	c-Pr	SO_2NH_2	5	0.69897
5	t-Bu	SO_2Me	3.1	0.491362
6	t-Bu		10	1
7	t-Bu	CH_2CH_2COOH	10	1
8	t-Bu	$C\equiv CCO_2H$	2.9	0.462398
9	t-Bu	$C\equiv CCH_2CH_2OH$	5.5	0.740363
10	c-Pr	$C\equiv CCH_2CH_2OH$	2	0.30103
11		H	44	1.643453
12		CH_2OH	20	1.30103
13		COOH	47	1.672098
14			17	1.230449
15			80	1.90309
16			23	1.361728
17			38	1.579784
18			30	1.477121
19			7.9	0.897627

20		17	1.230449
21		32	1.50515
22		23	1.361728
23		28	1.447158
24		19	1.278754
25		43	1.633468
26		26	1.414973

Results and Discussion

The 2D QSAR study of twenty four compounds (divided into seven test and seventeen training) for HIV integrase inhibitory activity through MLR, PLR and PCR analysis using V-Life MDS resulted in following statistically significant model, considering the term selection criteria as r^2.The statistically significant model (Model 1) with coefficient of determination (r^2) = 0.8231 (which corresponds to value of r =0.7532) was considered as The model showed an internal predictive power (q^2=0.8471) of 84% and predictivity for external test set (pred_r^2 =0.8129) about 81%. This model indicates the positive contribution of hydrogen count. External coordinates of the dipole moment contributes positively thus more electronegative groups should be included such as -I,-F,-Cl,-OH etc.

Model 1

Log p (log_{IC50}) =+0.5432 RotatableBondCount+ 0.6574 H-AcceptorCount-0.3218 chi3 - 5.7584 +0.6391 Quadrupole

n = 17; Degree of freedom = 11; r2 = 0.8231; q2 = 0.8471 F test = 31.1121; r2 se = 0.3363; q2 se = 0.6525; pred_r2 = 0.7532; pred_r2se = 0.8129

The same data set were subjected to Partial Least Square Regression method resulted in coefficient of correlation as 0.8231, internal predictive power as 82 % with poor external predictivity. Hydrogen count and external coordinates of the dipole contributes positively (same as MLR method), Quadrupole3 contributes positively towards biological activities.

Model 2

Log p (log_{IC50}) = + 0.7492 Polar Surface Area Including Pand + 2.3021 chi3Cluster+0.3821 T_O_O_6 --0.2341chiV3Cluster+2.321 Rotatable BondCount

n = 17; Degree of freedom = 13; r2 = 0.9291; q2 = 0.6895; F test = 76.8003; r2 se = 0.5321; q2 se = 0.4880; pred_r2 = 0.6476; pred_r2se = 0.6749

Principle Component Regression analysis with the same data set resulted in coefficient of correlation as 0.9291, internal predictive power as 48.80% with good external predictivity of 59.9%. Hydrogens count contributes in the same manner as above two. Rotatable Bond Count defines the total no. of carbon connected with four single bonds and gives negative contribution towards activity.

Model 3

Log p (log_{IC50}) = +3.36534 Hydrogens Count + 0.2651 Dipole Moment+ 2.3021 chi5Cluster

n = 17; Degree of freedom = 15; r2 = 0.7954; q2 = 0.3893; F test = 27.8056 r2 se = 0.4661; q2 se = 0.6153; pred_r2 = 0.7564; pred_r2se = 0.3940

Principle Component Regression analysis with the same data set resulted in coefficient of correlation as 0.7954, internal predictive power as 79.54 % with good external predictivity of 75.64 %. Hydrogens count contributes in the same manner as above two. Chi5Cluster defines the total no. of carbon connected with four single bonds and gives negative contribution towards activity.

Table 3: Inter correlation matrix of the independent descriptor

	Rotatable Bond count	H-Acc Coun	chi3	Quadrupole3
Rot Bond count	1.000			
H-Acce.count	0.4352	1.000		
chi3	0.2163	0.7521	1.000	
Quadrupole3	0.3031	0.6712	0.4218	1.000

Table-2 Predicted biological activity and LOO predicted activity with their variance in comparison to the observed biological activity

Compound	Observed pIC_{50}		Calculated pIC_{50}	
		Model 1	Model 2	Model 3
1	0.69897	0.684741	0.735643	0.655333
2	1.20412	0.621657	0.735643	0.547936
3	0.763428	0.584381	0.735643	0.633775
4	0.69897	0.521297	0.735643	0.526828
5	0.491362	0.706653	0.735643	0.682349
6	1	1.061696	0.735643	1.00112
7	1	0.783422	0.735643	0.87587
8	0.462398	0.642356	0.735643	0.459672
9	0.740363	0.690514	0.735643	0.841289
10	0.30103	0.62743	0.735643	0.732814
11	1.643453	1.530838	1.441886	1.515274
12	1.30103	1.352369	1.431201	1.319338
13	1.672098	1.349789	1.415983	1.121938
14	1.230449	1.475124	1.428458	1.325058
15	1.90309	1.660087	1.431785	1.897504
16	1.361728	1.569972	1.412182	1.267585
17	1.579784	1.475673	1.433659	1.483413
18	1.477121	1.409832	1.416418	1.414897
19	0.897627	0.897626	1.419189	0.892526
20	1.230449	1.516027	1.429207	1.470558
21	1.50515	1.415446	1.418219	1.397136
22	1.361728	1.409832	1.416418	1.414897
23	1.447158	1.515157	1.415476	1.40511
24	1.278754	1.438326	1.413325	1.409065
25	1.633468	1.478366	1.418077	1.518788
26	1.414973	1.580433	1.417382	1.425319

Figure 1: Plot between predicted pIC_{50} and observed pIC_{50} values of compounds of training set

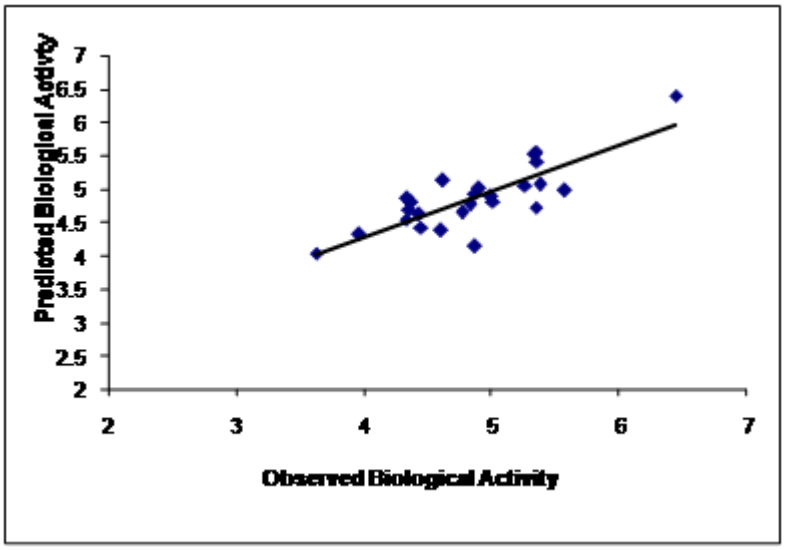
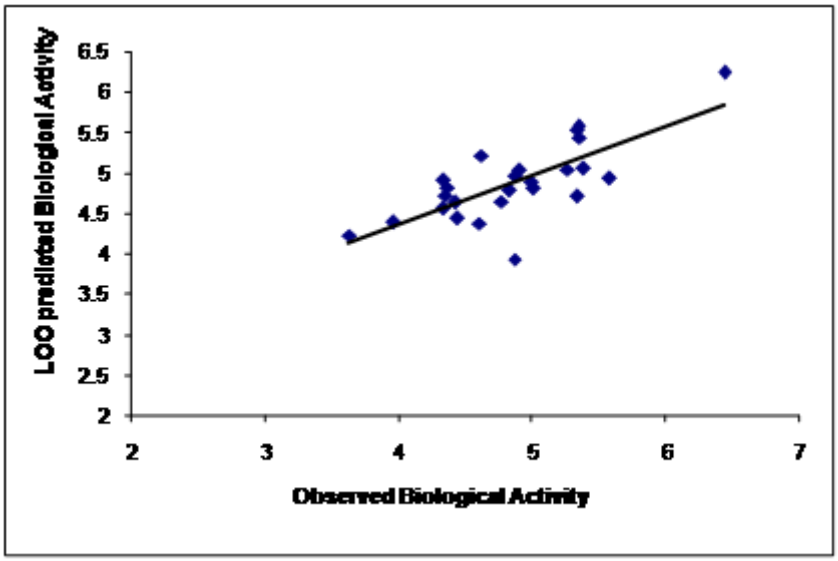


Table 4 Unicolumn statistics of Training and Test sets

Activity (pic50)	Average	Maximum	Minimum	Std. Dev	Sum
Training set	1.3728	3.2011	0.05463	0.2164	15.4326
Test set	0.2179	0.65834	0.03841	0.3318	8.4680

Figure 2: Plot between LOO predicted pIC_{50} and observed pIC_{50} values of compounds of training set



Conclusions

The present studies were aimed at deriving predictive 2D QSAR models capable of elucidating the structural requirements for antibacterial translation inhibitors. The 2D QSAR model resulted in the r2 = 0.8231and q2 = 0.6525 through multiple regression (MLR), r2=0.9291 and q2= 0.7896 through Partial least square method (PLSR) and r2 = 0.9291and q2 = 0.6895 through partial component analysis (PCR). All 2D models indicated the positive contribution of dipole, polarizability and hydrogen counts towards the activity. 3D results revealed that less bulky substituent group was preferred in that region. Freedom of an amide C-N bond compared to an acyclic C-C bond is required for the activity; presence of less bulky groups on benzene ring would improve the activity and selectivity of benzyl amide-ketoacid derivatives as HIV integrase inhibitors. The developed models also possess promising predictive ability as discerned by testing on the external test set and should be useful to elucidate the relationship between compound structure and biological activities and to facilitate design of more potent and selective agents.

Acknowledgement

The authors are thankful to Mr.Amit Badi for given valuable suggestions and Vlife Science Technologies Pvt. Ltd, 1, Akshay 50, Anand Park, Aundh, Pune, India to provide the software.

References

[1] J.T. Leonard; K. Roy *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 4467-74.
[2] J.T. Leonard; K. Roy *Bioorg. Med. Chem.*, **2006**, *14*, 1039-46.
[3] B. Hemmateenejad; S.M.H.Tabaei; F.Namvaran *J. Mol. Struct.*, **2005**, *732*, 39-45.
[4] M.A.Chamjangali;M. Beglari; G.Bagherian *J.Mol.Graph.Model.*,**2007**,*26*, 360-67.
[5] J.T. Leonard; K. Roy *Eur. J. Med. Chem.*, **2008**, *43*, 81-92.
[6] J.L.Andersen; E. Le. Rouzic; V. Planelles *Exp. Mol. Path.*, **2008**, *85*, 2-10.
[7] G.A.Barbaro; A.Scozzafava; C.T.Mastrolorenzo *Supuran.Curr. Pharm. Des.*, **2005,** 11, 1805-1843.
[8] Y.Yazdanpanah; D.Sissoko; M.Egger; Y.Mouton; M.Zwahlen; G.Chene *B.M.J.*, **2004**,328, 249-253.
[9]S.D.Young;S.F.Britcher;L.O.Tran;L.S.Payne;W.C.Lumma;T.A.Lyle;J.R.Huff;P.S.Anderson; D.B.Olsen;S.S.Carroll *Antimicrob. Agents. Chemother.*,**1995,** 39, 2602.
[10] Vlife MDS software package, version 3.5, supplied by Vlife science technologies Pvt. Ltd,1, Akshay 50, Anand Park, Aundh, Pune, India 411007.
[11] A.Golbraikh; A.Tropsha *J. Comp. Aided Molecular Design.*, **2002**, 16,357-369.
[12] J. Balzarini *Curr. Top Med. Chem.*, ***2004,*** 4, 921.