

QSAR modeling of thymine based derivatives of HEPT series for anti-HIV compounds against HIV-1

Uttam K. Tripathi^{a*}, Indra P. Pandey^a, Laxmi Barelia^b, M. Sharma^b, S. P. Singh^a, Sangh Mitra^c and Jaya Dwivedi^c

^aDepartment of Chemistry, D. A. V. (P.G.) College, Dehradun-248 001, Uttarakhand, India

E-mail : uktripathi@hotmail.com

^bDepartment of Chemistry, S.M.S. Govt., Model Science College, Gwalior, Madhya Pradesh, India

^cDepartment of Chemistry, Bansthali Vidhyapeeth, Tank, Rajasthan, India

Manuscript received 17 January 2011, accepted 16 June 2011

Abstract : We performed 2D QSAR studies upon a series of 78 HEPT analogues, inhibitors of HIV reverse transcriptase; using the QSAR that imply analysis of correlation and multilinear regression; a significant collection of descriptors (lipophilicity, steric parameter, field effect, hydrogen acceptor and hydrogen donor) was used. The best QSAR model with good correlation coefficient ($r^2 = 0.792$) of high statistical significance ($>99.9\%$) well explained the variance in activity. The analysis reveal that R^1 should have much saturated carbon with branching, R^2 should be bulky with less number of H acceptors and R^3 should have less number of H_{acc} in support of biological activity. The presence of X = O or S is not contributing in biological activity to a significant level.

Keywords : 2D-QSAR, anti-HIV activity, HEPT analogous, substituent constants by Hansch.

Introduction

Most people are familiar with divesting effects of HIV, the human immunodeficiency virus. The virus, which is transmitted by blood to blood contact, may produce no symptoms for years but typically within 10 to 15 years destroy the T_4 -lymphocytes, the cell that play a key role on the immune system and causes a fatal disease i.e. AIDS (acquired immunodeficiency syndrome)¹. The resulting depletion in the level of essential immuno cells leave patients vulnerable to opportunistic infection that would not normally harm a healthy person².

The Human Immunodeficiency virus is a retrovirus because it is having an enzyme, able to transform RNA to DNA (reverse to the common process in cellular biology)³. The enzyme, reverse transcriptase, allows DNA to transcribe from RNA. Thus HIV can make copies of its own genome, as DNA, in human cells such as CD4 "helper" lymphocyte. The Viral DNA becomes integrated in the lymphocyte genome, and this is the basis for chronic HIV infection. Integration of the HIV genome into host cells is a formidable obstacle to any antiviral treatment that would not just suppress but also eradicate the infection. Nevertheless, modern treatment with combination

of nucleoside analogues and protease inhibitors has transformed the prognosis for carriers of HIV, usually achieving a sustained fall in virus concentration in blood and restoration of the main target cell (CD4 lymphocyte) to near normal levels.

HIV life cycle starts with a high affinity of its membrane via gp120 for the CD4 receptor (protein molecule) at the host cell's surface⁴. CD4 receptor is a protein molecule present on the surface of T-lymphocytes that are helper and inducer, respectively; of the immune response (T4-lymphocytes modulate the immune response). The next stage, the fusion between the virus and membrane of host cell, emerges through gp41 and the HIV genomic RNA. RT and genomic RNA enter the double DNA colaring. Thus, DNA migrates towards the nucleus in order to integrate in the host cell's chromosomes by the viral encoded integrase enzyme. The incorporation of the formed "provirus" into the host cell genome is permanent^{5,6}.

By contrast, the inherent variability of the HIV genome and the failure of human host to produce neutralizing antibodies to the virus, as well as technical difficulties and concern about safety, have continued to frustrate

attempts to make an effective vaccine too. This must not, however, allow efforts to develop and evaluate candidate vaccines to slacken.

A particular concern is that a useful candidate vaccine (probably a recombinant envelope vaccine developed in North America or Europe against the locally prevalent HIV-1 B subtype) would be ineffective in those parts of the world where other subtype predominate⁷.

Known human retroviruses are : HTLV-I, HTLV-II and HIV. Pathogenic human retrovirus (HTLV-I, HIV) have the same target T-lymphocytes. Thus Robert Gallo *et al.* named these agents human T-lymphotropic viruses⁸.

The derivatives of HEPT are used as non-nucleosidic reverse transcriptase inhibitors (NNRTI). These derivatives do not interact with the binding site of the DNA or RNA dependent DNA polymerase. This is why, it is expected not to determine the side effects. HEPT ligand interact uncompetitive with an allosteric site of the enzyme and don't affect in a direct way the substrate binding. Actually, NNRTI have a higher binding affinity to the ligand-enzyme complex than to a free enzyme. The HEPT ligand-enzyme interaction leads to enzymatic conformational variations meaning that the enzyme's active site has a decreased affinity to the natural substrate. This property is valid only regarding the HIV-1 RT; HEPT ligands are inactive against HIV-2 or other retroviruses. The NNRTI exclusively specify for the HIV-1 RT is due to the presence at the level of this enzyme (and not in the

case of RT or DNA-polymerases) of a flexible extreme hydrophobic pocket in which HEPT derivatives (different from natural substrate analogues) fit and can be bound^{9,10}.

Material and methods :

Literature data regarding quantitative studies upon chemical structure-biological activity relationship (QSAR)¹⁰⁻²⁰ in the class of HEPT derivative are strictly related to the presence or absence of atomic group in certain positions on the HEPT general structure (see Fig. 1). Our research is directed towards an extended HEPT series and is performed using the Hansch model with two dimensional structural descriptors adapted to multi-linear regression technique (MLR)²¹.

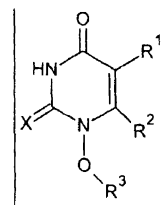


Fig. 1. General structure taken for QSAR study.

Results and discussion

Correlation of biological activity (dependent) with different combination of physicochemical parameter (independent) in the series of 78 studied ligand of Table 1. Several equations are obtained in order to correlate bio-

Table 1. The general structure to be analyzed, together with chemical constitution of R₁, R₂, R₃, X and experimental anti-HIV activity as BA_{obs} are presented in Table 1

Sl. no.	R ₁	R ₂	R ₃	X	BA _{obs}
1.	Methyl	2-Methylphenylthio	2-Hydroxyethyl	O	4.15
2.	Methyl	2-Nitrophenylthio	2-Hydroxyethyl	O	3.85
3.	Methyl	2-Methoxyphenylthio	2-Hydroxyethyl	O	4.72
4.	Methyl	3-Methylphenylthio	2-Hydroxyethyl	O	5.59
5.	Methyl	3-Ethylphenylthio	2-Hydroxyethyl	O	5.57
6.	Methyl	3- <i>tert</i> -Butylphenylthio	2-Hydroxyethyl	O	4.92
7.	Methyl	3-Trifluoromethylphenylthio	2-Hydroxyethyl	O	4.35
8.	Methyl	3-Fluorophenylthio	2-Hydroxyethyl	O	5.48
9.	Methyl	3-Chlorophenylthio	2-Hydroxyethyl	O	4.89
10.	Methyl	3-Bromophenylthio	2-Hydroxyethyl	O	5.24
11.	Methyl	3-Iodophenylthio	2-Hydroxyethyl	O	5.00
12.	Methyl	3-Nitrophenylthio	2-Hydroxyethyl	O	4.47
13.	Methyl	3-Hydroxyphenylthio	2-Hydroxyethyl	O	4.09
14.	Methyl	3-Methoxyphenylthio	2-Hydroxyethyl	O	4.66

Table-1 (contd.)

15.	Methyl	5-Dimethylphenylthio	2-Hydroxyethyl	O	6.59
16.	Methyl	3,5-Dichlorophenylthio	2-Hydroxyethyl	O	5.89
17.	Methyl	3,5-Dimethylphenylthio	2-Hydroxyethyl	S	6.66
18.	Methyl	3-Methoxycarbonylphenylthio	2-Hydroxyethyl	O	5.10
19.	Methyl	3-Acetylphenylthio	2-Hydroxyethyl	O	5.14
20.	Methyl	3-Cyanophenylthio	2-Hydroxyethyl	O	5.00
21.	Allyl	Phenylthio	2-Hydroxyethyl	O	5.60
22.	Methyl	Phenylthio	2-Hydroxyethyl	S	6.96
23.	Propyl	Phenylthio	2-Hydroxyethyl	S	5.00
24.	Isopropyl	Phenylthio	2-Hydroxyethyl	S	7.23
25.	Ethyl	3,5-Dimethylphenylthio	2-Hydroxyethyl	S	8.11
26.	Isopropyl	3,5-Dimethylphenylthio	2-Hydroxyethyl	S	8.30
27.	Ethyl	3,5-Dichlorophenylthio	2-Hydroxyethyl	O	7.37
28.	Ethyl	Phenylthio	2-Hydroxyethyl	O	6.92
29.	Propyl	Phenylthio	2-Hydroxyethyl	O	5.47
30.	Isopropyl	Phenylthio	2-Hydroxyethyl	O	7.20
31.	Ethyl	3,5-Dimethylphenylthio	2-Hydroxyethyl	O	7.89
32.	Isopropyl	3,5-Dimethylphenylthio	2-Hydroxyethyl	O	8.57
33.	Ethyl	3,5-Dichlorophenylthio	2-Hydroxyethyl	O	7.85
34.	Methyl	4-Methylphenylthio	2-Hydroxyethyl	O	3.66
35.	Methyl	Phenylthio	2-Hydroxyethyl	O	5.15
36.	Methyl	Phenylthio	2-Hydroxyethyl	S	6.01
37.	Iodo	Phenylthio	2-Hydroxyethyl	O	5.44
38.	Ethenyl	Phenylthio	2-Hydroxyethyl	O	5.69
39.	2-Phenylethenyl	Phenylthio	2-Hydroxyethyl	O	5.22
40.	Benzyl	Phenylthio	2-Hydroxyethyl	O	4.37
41.	Methyl	Phenylthio	2-Methoxyethyl	O	5.06
42.	Methyl	Phenylthio	2-Acetoxyethyl	O	5.17
43.	Methyl	Phenylthio	2-Benzoyloxyethyl	O	5.12
44.	Methyl	Phenylthio	Ethyl	O	6.48
45.	Methyl	Phenylthio	2-Chloroethyl	O	5.82
46.	Methyl	Phenylthio	2-Fluoroethyl	O	5.69
47.	Methyl	Phenylthio	Propyl	O	5.48
48.	Methyl	Phenylthio	Benzyl	O	7.06
49.	Ethyl	Phenylthio	Ethyl	O	7.72
50.	Ethyl	Phenylthio	Ethyl	S	7.58
51.	Ethyl	3,5-Dimethylphenylthio	Ethyl	O	8.24
52.	Ethyl	3,5-Dimethylphenylthio	Ethyl	S	8.30
53.	Ethyl	Phenylthio	Benzyl	O	8.23
54.	Ethyl	3,5-Dimethylphenylthio	Benzyl	O	8.55
55.	Ethyl	Phenylthio	Benzyl	S	8.09
56.	Ethyl	3,5-Dimethylphenylthio	Benzyl	S	8.14
57.	Isopropyl	Phenylthio	Ethyl	O	7.99
58.	Isopropyl	Phenylthio	Benzyl	O	8.51
59.	Isopropyl	Phenylthio	Ethyl	S	7.89
60.	Isopropyl	Phenylthio	Benzyl	S	8.14

Table-1 (contd.)

61.	Methyl	Phenylthio	Methyl	O	5.68
62.	Methyl	Phenylthio	Butyl	O	5.33
63.	Methyl	Phenylthio	Methyl	S	5.66
64.	Methyl	Phenylthio	Propyl	S	5.92
65.	Ethyl	3,5-Dichlorophenylthio	Ethyl	S	7.89
66.	Ethyl	Phenylthio	Isopropyl	S	6.66
67.	Ethyl	Phenylthio	Cyclohexyl	S	5.79
68.	Ethyl	Phenylthio	Cyclohexylmethyl	S	6.45
69.	Ethyl	Phenylthio	4-Methylbenzyl	S	7.11
70.	Ethyl	Phenylthio	4-Chlorobenzyl	S	7.92
71.	Ethyl	Phenylthio	2-Phenylethyl	S	7.04
72.	Ethyl	3,5-Dichlorophenylthio	Ethyl	O	8.13
73.	Ethyl	Phenylthio	Isopropyl	O	6.47
74.	Ethyl	Phenylthio	Cyclohexyl	O	5.40
75.	Ethyl	Phenylthio	Cyclohexylmethyl	O	6.35
76.	Ethyl	Phenylthio	2-Cyclohexylethyl	O	7.02
77.	Cyclopropyl	Phenylthio	Ethyl	S	7.02
78.	Cyclopropyl	Phenylthio	Ethyl	O	7.00

Table 2. The description of physical parameters used in study are given in Table 2.

Functional families of descriptors	Descriptor definition
Constitutional descriptors	Hydrogen acceptor (H_{acc}) and hydrogen donor (H_{don})
Steric descriptors	Molecular refractivity (MR)
Hydrophobic parameter	Substituent constant (F_r)
Electronic parameter	Swain and Lupton field parameter (ζ)

logical activity with these descriptors. Some equations with most acceptable statistical parameter are shown in Table 5. Different combinations of physico-chemical parameters are showing some acceptable correlation with biological activity. The study was carried out using stepwise multiple regression analysis in order to develop QSAR equations.

Table 3. The descriptors values of R_1 , R_2 and R_3 are given in Table 3, are calculated with Hansch well characterized substituent's table. For QSAR study, S is considered as 1 and O as -1

Sl. no.	R_1F_r	R_1MR	$R_1\zeta$	R_2H_{acc}	R_2MR	R_3H_{acc}	BA_{obs}	BA_{cal}	Residual
1.	0.77	5.65	-0.04	0	38.55	1	4.15	4.989	0.839
2.	0.77	5.65	-0.04	1	39.6	1	3.85	4.334	0.484
3.	0.77	5.65	-0.04	1	40.22	1	4.72	4.372	-0.348
4.	0.77	5.65	-0.04	0	38.55	1	5.59	4.989	-0.601
5.	0.77	5.65	-0.04	0	43.19	1	5.57	5.799	0.229
6.	0.77	5.65	-0.04	0	52.51	1	4.92	6.368	1.448
7.	0.77	5.65	-0.04	0	37.91	1	4.35	5.477	1.127
8.	0.77	5.65	-0.04	0	34.45	1	5.48	5.266	-0.214
9.	0.77	5.65	-0.04	0	38.82	1	4.89	5.533	0.643
10.	0.77	5.65	-0.04	0	41.69	1	5.24	5.708	0.468
11.	0.77	5.65	-0.04	0	46.65	1	5	6.01	1.01
12.	0.77	5.65	-0.04	1	39.6	1	4.47	4.334	-0.136
13.	0.77	5.65	-0.04	1	35.44	1	4.09	4.607	0.517
14.	0.77	5.65	-0.04	1	40.22	1	4.06	4.372	0.312
15.	0.77	5.65	-0.04	0	44.19	1	6.59	5.86	-0.73

Table-3 (contd.)

16.	0.77	5.65	-0.04	0	44.75	1	5.89	5.894	0.004
17.	0.77	5.65	-0.04	0	44.19	1	6.66	5.86	-0.8
18.	0.77	5.65	-0.04	2	77.97	1	5.1	4.902	-0.198
19.	0.77	5.65	-0.04	1	43.18	1	5.14	4.553	-0.587
20.	0.77	5.65	-0.04	1	38.28	1	5	4.254	-0.746
21.	1.42	14.41	0.03	0	32.89	1	5.6	6.176	0.576
22.	1.43	10.3	-0.05	0	32.89	1	6.96	6.716	-0.244
23.	1.97	14.96	-0.06	0	32.89	1	5	7.743	2.743
24.	1.84	14.96	-0.05	0	32.89	1	7.23	7.257	0.027
25.	1.43	10.3	-0.05	0	44.19	1	8.11	7.405	-0.705
26.	1.84	14.96	-0.05	0	44.19	1	8.3	7.946	-0.354
27.	1.43	10.3	-0.05	0	44.75	1	7.37	7.44	0.07
28.	1.43	10.3	-0.05	0	32.89	1	6.92	6.716	-0.204
29.	1.97	14.96	-0.06	0	32.89	1	5.47	7.743	2.273
30.	1.84	14.96	-0.05	0	32.89	1	7.2	7.257	0.057
31.	1.43	10.3	-0.05	0	44.19	1	7.89	7.405	-0.485
32.	1.84	14.96	-0.05	0	44.19	1	8.57	7.946	-0.624
33.	1.43	10.3	-0.05	0	44.75	1	7.85	7.44	-0.41
34.	0.77	5.65	-0.04	0	38.55	1	3.66	4.989	1.329
35.	0.77	5.65	-0.04	0	32.89	1	5.15	5.171	0.021
36.	0.77	5.65	-0.04	0	32.89	1	6.01	5.171	-0.839
37.	0.59	13.76	0.4	0	32.89	1	5.44	5.461	0.021
38.	0.88	9.79	0.07	0	32.89	1	5.69	5.356	-0.334
39.	2.08	34.95	0.15	0	32.89	1	5.22	4.5	-0.72
40.	1.97	30.81	0.04	0	32.89	1	4.37	4.315	-0.055
41.	0.77	5.65	-0.04	0	32.89	1	5.06	5.171	0.111
42.	0.77	5.65	-0.04	0	32.89	1	5.17	5.171	0.001
43.	0.77	5.65	-0.04	0	32.89	2	5.12	4.418	-0.702
44.	0.77	5.65	-0.04	0	32.89	0	6.48	5.924	-0.556
45.	0.77	5.65	-0.04	0	32.89	0	5.82	5.924	0.104
46.	0.77	5.65	-0.04	0	32.89	0	5.96	5.924	-0.036
47.	0.77	5.65	-0.04	0	32.89	0	5.48	5.924	0.444
48.	0.77	5.65	-0.04	0	32.89	0	7.06	5.924	-1.136
49.	1.43	10.3	-0.05	0	32.89	0	7.72	7.469	-0.251
50.	1.43	10.3	-0.05	0	32.89	0	7.58	7.469	-0.111
51.	1.43	10.3	-0.05	0	44.19	0	8.24	8.158	-0.082
52.	1.43	10.3	-0.05	0	44.19	0	8.3	8.158	-0.142
53.	1.43	10.3	-0.05	0	32.89	0	8.23	7.469	-0.761
54.	1.43	10.3	-0.05	0	44.19	0	8.55	8.158	-0.392
55.	1.43	10.3	-0.05	0	32.89	0	8.09	7.469	-0.621
56.	1.43	10.3	-0.05	0	44.19	0	8.14	8.158	0.018
57.	1.84	14.96	-0.05	0	32.89	0	7.99	8.01	0.02
58.	1.84	14.96	-0.05	0	32.89	0	8.51	8.01	-0.5
59.	1.84	14.96	-0.05	0	32.89	0	7.89	8.01	0.12
60.	1.84	14.96	-0.05	0	32.89	0	8.14	8.01	-0.13
61.	0.77	5.65	-0.04	0	32.89	0	5.68	5.924	0.244

Table-3 (contd.)

62.	0.77	5.65	-0.04	0	32.89	0	5.33	5.924	0.594
63.	0.77	5.65	-0.04	0	32.89	0	5.66	5.924	0.264
64.	0.77	5.65	-0.04	0	32.89	0	5.92	5.924	0.004
65.	1.43	10.3	-0.05	0	44.75	0	7.89	8.193	0.303
66.	1.43	10.3	-0.05	0	32.89	0	6.66	7.469	0.809
67.	1.43	10.3	-0.05	0	32.89	2	5.79	5.963	0.173
68.	1.43	10.3	-0.05	0	32.89	2	6.45	5.963	-0.487
69.	1.43	10.3	-0.05	0	32.89	0	7.11	7.469	0.359
70.	1.43	10.3	-0.05	0	32.89	0	7.92	7.469	-0.451
71.	1.43	10.3	-0.05	0	32.89	0	7.04	7.469	0.429
72.	1.43	10.3	-0.05	0	44.75	0	8.13	8.193	0.063
73.	1.43	10.3	-0.05	0	32.89	0	6.47	7.469	0.999
74.	1.43	10.3	-0.05	0	32.89	2	5.4	5.963	0.563
75.	1.43	10.3	-0.05	0	32.89	2	6.35	5.963	-0.387
76.	1.43	10.3	-0.05	0	32.89	2	7.02	5.963	-1.057
77.	1.49	13.53	-0.03	0	32.89	0	7.02	7.026	0.006
78.	1.49	13.53	-0.03	0	32.89	0	7	7.026	0.026

Correlation matrix to demonstrate inter correlation among ligand descriptors versus biological activity shown in Table 4.

The correlation between calculated descriptors as an

independent variable and biological activity as response variable was found out by sequential linear regression analysis. Some equations with good statistical results are shown in Table 5.

Table 4

	A_{exp}	R_1F_r	R_1MR	$R_1\zeta$	R_2H_{acc}	R_2MR	R_3H_{acc}
A_{exp}	1						
R_1F_r	0.651	1					
R_1MR	0.304	0.815	1				
$R_1\zeta$	-0.209	-0.09	0.363	1			
R_2H_{acc}	-0.409	-0.314	-0.246	-0.041	1		
R_2MR	-0.001	-0.181	-0.215	-0.125	0.5	1	
R_3H_{acc}	-0.448	-0.127	-0.031	0.108	0.167	0.119	1

Table 5

Sl. no.	Equation	Statistical parameter
1.	$-\log IC_{50} = 4.518 (\pm 0.657) R_1F_r - 0.283 (\pm 0.061) R_1MR + 6.910 (\pm 2.539) R_1\zeta - 1.430 (\pm 0.350) R_2H_{\text{acc}} + 0.590 (\pm 0.442) R_2H_{\text{don}} + 0.058 (\pm 0.018) R_2MR + 1.351 (\pm 0.958) R_3H_{\text{acc}}$	$N = 57, R = 0.845, R^2 = 0.707, SE = 0.791, F = 20.084, \text{Adj } R^2 = 0.672, \text{PRESS} = 43.757, \text{SSY} = 106.777, S_{\text{PRESS}} = 0.876, R^2_{\text{CV}} = 0.590$
2.	$-\log IC_{50} = 0.104 (\pm 0.031) R_1MR - 4.650 (\pm 2.185) R_1\zeta - 1.744 (\pm 0.426) R_2H_{\text{acc}} + 0.085 (\pm 0.544) R_2H_{\text{don}} + 0.073 (\pm 0.023) R_2MR - 0.885 (\pm 0.232) R_3H_{\text{acc}} + 2.500 (\pm 1.189) R_3H_{\text{don}}$	$N = 57, R = 0.743, R^2 = 0.551, SE = 0.979, F = 10.243, \text{Adj } R^2 = 0.498, \text{PRESS} = 171.165, \text{SSY} = 106.777, S_{\text{PRESS}} = 1.733, R^2_{\text{CV}} = -0.603$
3.	$-\log IC_{50} = 1.734 (\pm 0.276) R_1F_r - 0.476 (\pm 1.817) R_1\zeta - 1.431 (\pm 0.358) R_2H_{\text{acc}} + 0.641 (\pm 0.451) R_2H_{\text{don}} + 0.069 (\pm 0.019) R_2MR - 0.822 (\pm 0.192) R_3H_{\text{acc}} + 1.818 (\pm 0.987) R_3H_{\text{don}}$	$N = 57, R = 0.833, R^2 = 0.694, SE = 0.808, F = 18.912, \text{Adj } R^2 = 0.657, \text{PRESS} = 79.980, \text{SSY} = 106.777, S_{\text{PRESS}} = 1.180, R^2_{\text{CV}} = 0.251$

Table-5 (contd.)

4.	$-\log IC_{50} = 2.946 (\pm 0.421) R_1F_r - 0.137 (\pm 0.039) R_1MR$ $- 1.367 (\pm 0.321) R_2H_{acc} + 0.578 (\pm 0.406) R_2H_{don} + 0.061$ $(\pm 0.017) R_2MR - 0.735 (\pm 0.173) R_3H_{acc} + 1.995 (\pm 0.886)$	$N = 57, R = 0.868, R^2 = 0.753, SE = 0.726.$ $F = 25.386, Adj R^2 = 0.723, PRESS = 49.340,$ $SSY = 106.777, S_{PRESS} = 0.930, R^2_{CV} = 0.538$
5.	$-\log IC_{50} = 4.802 (\pm 0.620) R_1F_r - 0.286 (\pm 0.059) R_1MR$ $+ 8.330 (\pm 2.431) R_1\zeta + 0.301 (\pm 0.422) R_2H_{don} + 0.027$ $(\pm 0.015) R_2MR - 0.840 (\pm 0.180) R_3H_{acc} + 2.901 (\pm 0.887)$	$N = 57, R = 0.853, R^2 = 0.727, SE = 0.763.$ $F = 22.244, Adj R^2 = 0.695, PRESS = 44.392,$ $SSY = 106.777, S_{PRESS} = 0.883, R^2_{CV} = 0.584$
6.	$-\log IC_{50} = 4.373 (\pm 0.552) R_1F_r - 0.266 (\pm 0.052) R_2MR$ $+ 7.369 (\pm 2.136) R_1\zeta - 1.185 (\pm 0.294) R_2H_{acc} + 0.052$ $(\pm 0.014) R_2MR - 0.761 (\pm 0.158) R_3H_{acc} + 2.641 (\pm 0.592)$	$N = 57, R = 0.890, R^2 = 0.792, SE = 0.666.$ $F = 31.782, Adj R^2 = 0.767, PRESS = 27.800,$ $SSY = 106.777, S_{PRESS} = 0.698, R^2_{CV} = 0.740$
7.	$-\log IC_{50} = 4.525 (\pm 0.621) R_1F_r - 0.283 (\pm 0.058) R_1MR$ $+ 7.234 (\pm 2.407) R_1\zeta - 0.642 (\pm 0.286) R_2H_{acc} - 0.055$ $(\pm 0.384) R_2H_{don} - 0.730 (\pm 0.178) R_3H_{acc} + 4.529 (\pm 0.498)$	$N = 57, R = 0.858, R^2 = 0.737, SE = 0.737,$ $F = 23.322, Adj R^2 = 0.705, PRESS = 44.786,$ $SSY = 106.777, S_{PRESS} = 0.886, R^2_{CV} = 0.581$
8.	$-\log IC_{50} = 4.297 (\pm 0.549) R_1F_r - 0.262 (\pm 0.051) R_1MR$ $+ 7.246 (\pm 2.116) R_1\zeta - 1.246 (\pm 0.294) R_2H_{acc} + 0.527$ $(\pm 0.368) R_2H_{don} + 0.061 (\pm 0.015) R_2MR - 0.753 (\pm 0.157)$ $R_3H_{acc} + 1.852 (\pm 0.805)$	$N = 57, R = 0.895, R^2 = 0.801, SE = 0.659,$ $F = 28.103, Adj R^2 = 0.772, PRESS = 30.163,$ $SSY = 106.777, S_{PRESS} = 0.727, R^2_{CV} = 0.718$
9.	$Y = 0.827x + 1.051$	$N = 20, R^2 = 0.781$

Conclusion :

Among these equations, the eq. (6) was considered the best model with the correlation coefficient ($r = 0.890$) explaining 79.2% variance in activity. The low standard error of estimate(s), high F value and more than one third value of coefficient suggest that the model is statistically highly significant. The data showed overall statistical significance $>99.9\%$ with $F = 31.782$ against tabulated value for Fischer's test at 99.9% significance ($F_{6,50\alpha 0.001} = 16.31$). The plot for training set are shown in Fig. 2.

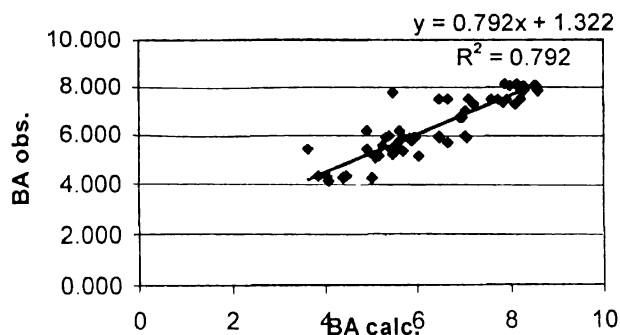


Fig. 2

The above studies indicate

(i) Discussion about R_1 :

Positive F_r indicates much carbon containing chain

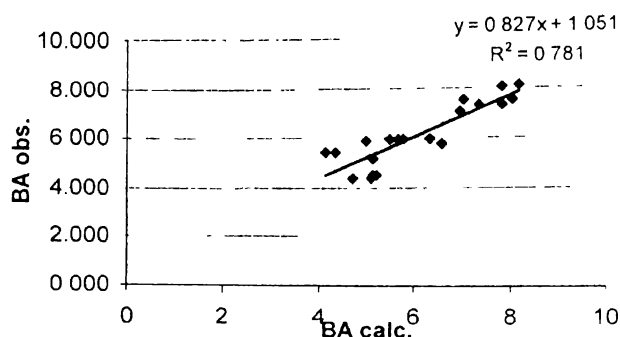


Fig. 3. The above model also predict the well inhibitory activity of the molecules of the test set as shown in Fig. 3, where the comparable correlation coefficient value ($r = 0.890$) was observed.

supports the activity (e.g. compounds 17 and 25), positive ζ indicates that positive field effects supports the activity (e.g. compounds 28, 35 and 40) but negative MR indicates that there should be some branching to reduce the bulkiness of chain on increasing the number of carbon (e.g. compounds 23 and 24).

(ii) Discussion about R_2 :

Negative H_{acc} indicates that increase of H acceptors in the group reduces the activity (e.g. compounds 15 and 16) and positive MR indicates that bulkiness in the group supports the activity (e.g. compounds 49, 51 and 72).

(iii) Discussion about R_3 :

Negative H_{acc} indicates that increase of H acceptors in the group reduces the activity (e.g. compounds 44, 45 and 46).

External validation :

The validation of the best model (eq. (6)) has been done on a test set of 20 compounds, where good correlation ($r^2 = 0.781$) was observed between the predicted and observed activity. The eq. (9) describes correlation between observed (y) and calculated (x) activities of the test set (Fig. 4).

In this study, compound 23 is outlier, other 20 compounds having residual less than 1.0 except this is having 2.784.

Discussion for outlier :

The equation is not suitable for compound 23, may be due to less number of compounds included in the study for R_1 as straight chain. Equation is already supporting branching with R_1 ligand.

Conclusion :

In the lead molecule (Fig. 1), if, R_1 is branched bulky group, R_2 is bulky without H-acceptors and R_3 is without H-acceptors will help to design more potent molecule against HIV-1 for future researcher.

Acknowledgement

The authors are thankful to P. Ayash Kumar, TEVA API, Gr. Noida for his support in this work.

References

1. S. J. O'Brien and M. Dean, Scientific American, September, 1997, 28-35.
2. J. Gold, Education in Chemistry, ACS, 1998, 12.
3. F. Vezinet-brun, *Ann. Pediar.*, 1987, **34**, 560.
4. M. Nasr, J. Craddock and M. Johnston, *Drug News & Perspectives*, 1993, **6**, 338.
5. D. Kkatzmann, E. Champagne, S. Chamaret, J. Gruest, D. Gretard, T. Hercend, J. C. Gluclman and L. Montagnier, *Nature*, 1984, **312**, 767.
6. J. L. Rossio, L. Bess (Jr.), L. E. Henderson, P. Cresswell and L. O. Arthur, *AIDS Res., Hum. Retroviruses*, 1995, **11**, 1433.
7. Micheal W. Adler, *ABC of AIDS*, ed. 5.
8. R. C. Gallo, S. Z. Salahuddin, M. Popovic, G. E. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Readfield, J. Oleske, B. Safai, G. White, P. Foster and P. Markham, *Science*, 1984, **224**, 500.
9. E. De Clercq, Elsevier Publication, 1990, 1-20.
10. C. Duda-Seiman, D. Duda-Seiman, M. V. Putz and Dciubotariu, *Digest Journal of Nanaomaterials and Biostructures*, 2007, **2**, 207.
11. R. Garg, S. P. Gupta, H. Gao, M. S. Babu, A. K. Debnath and C. Hansch, *Chem. Rev.*, 1999, **99**, 3525.
12. H. Tanaka, H. Takashima, M. Ubasawa K. Sekiya, I. Nitta, M. Baba, S. Shigeta, S. Walker, E. De Clercq and T. Miyasaka, *J. Med. Chem.*, 1992, **35**, 4713.
13. H. Tanaka, M. Baba, H. Hayakawa, T. Sakamaki, T. Miyasaka, M. Ubasawa, H. Takashima, K. Sekiya, I. Nitta, S. Shigeta, R. T. Walker, J. Balzarini and E. De Clercq, *J. Med. Chem.*, 1991, **34**, 349.
14. R. Potkins, R. Benhida, A. M. Aubertin, D. S. Grierson and C. Monneret, *J. Med. Chem.*, 1995, **40**, 1845.
15. D. K. Kim, J. Gam, Y. W. Kim, J. Lim and H. T. Kim, *J. Med. Chem.*, 1997, **40**, 2363.
16. J. M. Luco and F. H. Ferretti, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 392.
17. L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", ed. D. Bawden, Research Studies Press Ltd., Letchworth, Hertfordshire, 1986, Chap. 3.
18. M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.
19. L. B. Kier and L. H. Hall, *Quant. Struct. Act. Relat.*, 1991, **10**, 134.
20. C. Hansch and Albert Leo, "Substituents Constant for Correlation Analysis in Chemistry and Biology", Wiley-Interscience, New York, 1979.
21. Systat, SPSS Inc., 444, North Michigan Avenue, Chicago, IL, USA.