Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Bioorganic & Medicinal Chemistry 12 (2004) 1493–1503
Bioorganic & Medicinal Chemistry

QSAR study on some anti-HIV HEPT analogues using physicochemical and topological parameters

Shovanlal Gayen, Bikash Debnath, Soma Samanta and Tarun Jha*

Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, PO Box No 17020, Jadavpur University, Kolkata-700 032, India

Received 23 September 2003; accepted 18 December 2003

Abstract—A Quantitative Structure–Activity Relationship (QSAR) study has been carried out using topological indices, physicochemical and indicator parameters on a series of HEPT analogues for their HIV reverse transcriptase inhibitory activity. Correlation analysis and multiple linear regression (MLR) method were used to find out the best QSAR model. The results clearly explained that decreased hydrophobicity of substituents at R_1 and R_2 positions are favorable for the activity and presence of di-substitution at phenyl ring as well as *i*-Pr at R_1 position have detrimental effect but presence of OH group at R_2 position increases the activity. The atoms numbered as 1, 2, 3, 4, 5, 6 and 11 may constitute pharmacophore moiety of the HEPT analogues for their anti-HIV activity. Leave one out (LOO–) cross validation method was used to judge the predictive power of final equations. From the study one can propose that atom or fragmental level descriptors are more useful to interpret drug–receptor interactions in these analogues. The potentiality of ETSA index to recognize these atoms (Pharmacophoric atoms) was established.

1. Introduction

The pandemic form of sexually transmitted human immunodeficiency virus (HIV) that cause the acquired immunodeficiency syndrome (AIDS) has lead to the development of various non-nucleoside reverse transcriptase inhibitors (NNRTIs), which are less toxic and chemically more stable, having slower metabolizing rate and slowly emitted from the human body than nucleoside reverse transcriptase inhibitors (NRTIs). Among the non-nucleosides analogues, 1-[(2-hydroxy ethoxy) methyl]-6-(phenyl thio)-thymine (HEPT) is a potent inhibitor of HIV-1 reverse transcriptase (RT) which is necessary for the catalytic formation of proviral DNA from viral RNA.[1]

These highly specific non-competitive inhibitors (NNRTIs) have the capacity to bind non-covalently at a common allosteric site different from the site of polymerase activity. Several crystal structures of reverse transcriptase with different bound NNRTIs and comparison of binding pockets as well as the geometry of bound inhibitors have revealed that these (NNRTIs) bind in the same region of the enzyme. There is also a striking similarity in their binding mode of different structurally diverse NNRTIs. This similarity was also confirmed by a hologram QSAR study of NNRTIs.[2]

In general, aromatic portions of NNRTIs bind with the hydrophobic type of residues in the binding pocket such as Tyr 181, Tyr 188, Trp 229 and the polar portions have the characteristic to form hydrogen bonds with residues (Lys 101, Lys 103 of P66 and Glu 138 of P51) located at the entrance of the binding pocket. For rationale and effective design of inhibitors these characteristic features have to be considered. Another point must be kept in mind is that the emergence of drug resistant viral strains has limited the therapeutic efficiency of NNRTIs.[3]

In the present work, a quantitative structure–activity relationship (QSAR) study has been carried out using electrotopological state atom (ETSA) indices and hydrophobic as well as indicator parameters on a series of HEPT analogues. The general structure of that is shown in Figure 1. ETSA index is an atom/sub-molecular descriptor encoding both electronic and topological information. Increasing use of this topological index

Keywords: HEPT analogues; Reverse transcriptase inhibition; QSAR; Hydrophobicity; ETSA index.
* Corresponding author. Tel./fax: +91-33-24146677; e-mail: tjjupharm@yahoo.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.12.031

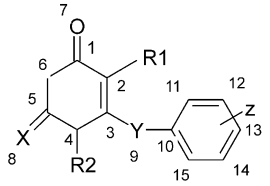

Figure 1. General structure of HEPT analogues.

in QSAR studies has demonstrated its importance to uncover a functional region of molecule with potential as pharmacophore. ETSA indices are calculated by a computer programme 'mouse', written in C^{++} language. The software was developed by us and can run in Windows system to calculate ETSA indices only. The structural details and the anti-HIV activity are collected from published work[4] to investigate the pharmacophoric and physicochemical requirements of substituents of HEPT analogues for their interesting reverse transcriptase inhibitory activity as a part of our composite programme of rational drug design.[5]

The aim of this work is two-fold: to perform a study to show that atom/fragmental level descriptors can give better statistically significant results than the whole molecular descriptors and to determine or recognize the pharmacophoric atoms relevant to the activity desired.

2. Results and discussion

In order to get the linear relationship with independent variables, logarithms of the biological activity (Log EC_{50}) data of 84HEPT analogues, collected from published work[4], were used and are shown in Table 1. Calculated ETSA indices, the hydrophobic parameters and some indicator parameters are presented in Table 2. Correlation analysis was performed among the dependant variable (Log EC_{50}) and useful independent variables and the result is shown in Table 3. From the correlation study it was found that S_1, S_2, S_3, S_4, S_5 and S_6 (ETSA indices of atom number 1, 2, 3, 4, 5 and 6 respectively) have higher individual correlation to biological activity. As they are highly auto correlated with each other we first tried to use separately S_1, S_2, S_3, S_4, S_5 and S_6 with other independent parameters to get statistically good QSAR models.

When S_1 was used with other descriptors, three eqs 1–3 were obtained given below:

$$\text{Log } EC_{50} = 4.805\ (\pm 0.804) + 3.457\ (\pm 0.532)\ S_1 + 1.218\ (\pm 0.377)\ S_{11} + 1.224\ (\pm 0.199)\ I_1 + 1.224\ (\pm 0.226)\ I_2 \quad (1)$$

$n = 84$; R = 0.864; %EV = 74.688; $R^2_A = 0.734$;
$F_{(4,79)} = 58.277$; $p < 0.000$; S.E.E. = 0.673
PRESS = 40.593; SSY = 141.256; $R^2_{CV} = 0.713$;
$S_{PRESS} = 0.717$; P.S.E = 0.695

$$\text{Log } EC_{50} = 5.049\ (\pm 0.775) + 2.655\ (\pm 0.583)\ S_1 + 1.004\ (\pm 0.369)\ S_{11} + 1.372\ (\pm 0.198)\ I_1 + 1.278\ (\pm 0.217)\ I_2 + 0.331\ (\pm 0.117)\ \pi R_2 \quad (2)$$

$n = 84$; R = 0.878; %EV = 77.056; $R^2_A = 0.756$;
$F_{(5,78)} = 52.392$; $p < 0.000$; S.E.E. = 0.645
PRESS = 37.403; SSY = 141.256; $R^2_{CV} = 0.735$;
$S_{PRESS} = 0.692$; P.S.E = 0.667

$$\text{Log } EC_{50} = 5.449\ (\pm 0.730) + 2.257\ (\pm 0.541)\ S_1 + 0.857\ (\pm 0.345)\ S_{11} + 1.384\ (\pm 0.181)\ I_1 + 1.240\ (\pm 0.201)\ I_2 - 0.757\ (\pm 0.163)\ I_{OH} \quad (3)$$

$n = 84$; R = 0.895; %EV = 80.150; $R^2_A = 0.789$;
$F_{(5,78)} = 62.990$; $p < 0.000$; S.E.E. = 0.600
PRESS = 32.203; SSY = 141.256; $R^2_{CV} = 0.772$;
$S_{PRESS} = 0.638$; P.S.E = 0.619

where n is the number of data points, R is correlation coefficient. The other statistical parameters considered are %EV (percentage of explained variance), R^2_A (adjusted R^2), F (ratio between the variances of observed and calculated activities), p (probability factor), S.E.E (standard error of estimate), PRESS (predicted residual sum of squares), SSY (total sum of squares), R^2_{CV} (cross validated R^2), S_{PRESS} (standard error of PRESS) and P.S.E. (uncertainty factor).[6] The values within the parenthesis are confidence intervals of corresponding parameters.

Eqs 1–3 explain the variation of anti-HIV activity of HEPT analogues up to 74.688%, 77.056% and 80.150% respectively. All the three equations show the importance of atoms 1 and 11. The biological activity will increase with the decrease of S_1 and S_{11} (ETSA indices of atoms 1 and 11 respectively) values as evidenced by their positive regression coefficients in these equations. I_1 represents the di-substitution at phenyl ring and have detrimental effect to the activity. Positive coefficient of I_2 signifies that absence of isopropyl group at R_1 favour the activity. The QSAR model (2) explains that increased hydrophobicity of R_2 substituents decreases the activity, though it has little effect on the activity as indicated by the lower value of the regression coefficients.

The eqs 4–6 are obtained when S_2 (ETSA index of atom 2) was taken as variable to get QSAR models. Among these, eq 6 is the best model giving higher correlation coefficient (R) 0.900 and lower value of standard error of estimate (S.E.E). Activity increases with the decrease in S_2 value and change of pattern of activity

Table 1. Biological activity data of HEPT analogues

Compd[a]	X	R_1	Y	Z	R_2	Log EC_{50}[b]
1	O	Me	CH_2	H	$CH_2OCH_2CH_2OH$	4.637
2	O	Me	C≡C	H	$CH_2OCH_2CH_2OH$	4.853
3	O	I	S	H	$CH_2OCH_2CH_2OH$	5.443
4	O	SPh	S	H	$CH_2OCH_2CH_2OH$	4.677
5	O	$COCHMe_2$	S	H	$CH_2OCH_2CH_2OH$	4.920
6	O	COPh	S	H	$CH_2OCH_2CH_2OH$	4.885
7	O	CH2Ph	S	H	$CH_2OCH_2CH_2OH$	4.637
8	O	$CH=CPh_2$	S	H	$CH_2OCH_2CH_2OH$	6.075
9	O	C≡CMe	S	H	$CH_2OCH_2CH_2OH$	4.720
10	O	C≡CPh	S	H	$CH_2OCH_2CH_2OH$	5.468
11	O	CH=CHPh	S	H	$CH_2OCH_2CH_2OH$	5.221
12	O	$CH=CH_2$	S	H	$CH_2OCH_2CH_2OH$	5.958
13	O	Me	S	2-Me	$CH_2OCH_2CH_2OH$	4.148
14	O	Me	S	2-OMe	$CH_2OCH_2CH_2OH$	4.720
15	O	Me	S	3-Me	$CH_2OCH_2CH_2OH$	5.584
16	O	Me	S	3-Et	$CH_2OCH_2CH_2OH$	5.568
17	O	Me	S	3-*t*-Bu	$CH_2OCH_2CH_2OH$	4.920
18	O	Me	S	3-Cl	$CH_2OCH_2CH_2OH$	4.885
19	O	Me	S	3-Br	$CH_2OCH_2CH_2OH$	5.243
20	O	Me	S	3-I	$CH_2OCH_2CH_2OH$	4.999
21	O	Me	S	$3\text{-}NO_2$	$CH_2OCH_2CH_2OH$	4.468
22	O	Me	S	3-OH	$CH_2OCH_2CH_2OH$	4.085
23	O	Me	S	3-OMe	$CH_2OCH_2CH_2OH$	4.657
24	O	Me	S	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	6.584
25	O	Me	S	$3,5\text{-}Cl_2$	$CH_2OCH_2CH_2OH$	5.885
26	S	Me	S	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	6.656
27	O	Me	S	3-COOMe	$CH_2OCH_2CH_2OH$	5.102
28	O	Me	S	3-COMe	$CH_2OCH_2CH_2OH$	5.136
29	O	Me	S	3-CN	$CH_2OCH_2CH_2OH$	4.999
30	O	$CH_2CH=CH_2$	S	H	$CH_2OCH_2CH_2OH$	5.601
31	O	COOMe	S	H	$CH_2OCH_2CH_2OH$	5.180
32	O	CONHPh	S	H	$CH_2OCH_2CH_2OH$	4.744
33	S	Et	S	H	$CH_2OCH_2CH_2OH$	6.957
34	S	Et	S	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	8.106
35	S	*i*-Pr	S	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	8.300
36	S	Et	S	$3,5\text{-}Cl_2$	$CH_2OCH_2CH_2OH$	7.365
37	O	Pr	S	H	$CH_2OCH_2CH_2OH$	5.468
38	O	Me	S	H	CH_2OMe	5.677
39	O	Me	S	H	CH_2OPr	5.443
40	O	Me	S	H	CH_2OBu	5.327
41	O	Me	S	H	CH_2OCH_2Ph	7.054
42	S	Et	S	$3,5\text{-}Me_2$	CH_2OEt	8.355
43	S	Et	S	$3,5\text{-}Cl_2$	CH_2OEt	7.885
44	S	Et	S	H	CH_2-*i*-Pr	6.656
45	S	Et	S	H	CH_2OCH_2-*c*-Hex	6.455
46	S	Et	S	H	CH_2OCH_2Ph	8.106
47	S	Et	S	$3,5\text{-}Me_2$	CH_2OCH_2Ph	8.160
48	S	Et	S	H	$CH_2OCH_2Ph(4\text{-}Me)$	7.107
49	S	Et	S	H	$CH_2OCH_2Ph(4\text{-}Cl)$	7.919
50	S	Et	S	H	$CH_2OCH_2{\cdot}CH_2Ph$	7.040
51	S	*i*-Pr	S	H	CH_2OEt	7.852
52	S	*i*-Pr	S	H	CH_2OCH_2Ph	8.179
53	S	*c*-Pr	S	H	CH_2OEt	7.021
54	O	Et	S	H	$CH_2O\text{-}I\text{-}Pr$	6.467
55	O	Et	S	H	$CH_2O\text{-}c\text{-}Hex$	5.397
56	O	Et	S	H	$CH_2OCH_2\text{-}c\text{-}Hex$	6.346
57	O	Et	S	H	$CH_2OCH_2CH_2Ph$	7.016
58	O	Me	S	H	Et	5.657
59	O	Me	S	H	Bu	5.920
60	O	Et	CH_2	H	$CH_2OCH_2CH_2OH$	6.455
61	O	Et	CH_2	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	7.885
62	O	Et	CH_2	H	CH_2OEt	7.386
63	O	*i*-Pr	CH_2	H	$CH_2OCH_2CH_2OH$	7.199
64	O	*i*-Pr	CH_2	$3,5\text{-}Me_2$	$CH_2OCH_2CH_2OH$	8.567
65	O	*i*-Pr	CH_2	H	CH_2OEt	8.375
66	O	*i*-Pr	CH_2	$3,5\text{-}Me_2$	CH_2OEt	9.220
67	O	*i*-Pr	CH_2	H	Bu	7.374
68	O	Et	CH_2	H	CH_2CH_2OMe	6.601
69	O	*i*-Pr	CH_2	H	CH_2CH_2OMe	7.282

(*continued on next page*)

Table 1 (*continued*)

Compd[a]	X	R_1	Y	Z	R_2	Log EC_{50}[b]
70	O	Me	S	H	$CH_2OCH_2CH_2OH$	5.154
71	O	C≡CH	S	H	$CH_2OCH_2CH_2OH$	4.744
72	O	Me	S	3-F	$CH_2OCH_2CH_2OH$	5.481
73	O	i-Pr	S	H	$CH_2OCH_2CH_2OH$	7.228
74	O	Me	S	H	$CH_2OCH_2CH_2OMe$	5.060
75	O	Me	S	H	CH_2OEt	6.480
76	S	Et	S	H	CH_2OEt	7.584
77	O	Et	S	H	CH_2OEt	7.720
78	O	Et	S	3,5-Me_2	CH_2OEt	8.266
79	O	Et	S	3,5-Me_2	CH_2OCH_2Ph	8.493
80	O	*i*-Pr	S	H	CH_2OEt	7.919
81	O	*c*-Pr	S	H	CH_2OEt	6.999
82	O	Et	CH_2	3,5-Me_2	CH_2OEt	7.199
83	O	Et	CH_2	H	Bu	6.677
84	S	Me	S	H	$CH_2OCH_2CH_2OH$	6.008

[a] Compd = compound number.
[b] Taken from published work.[4]

with other variable was discussed previously in case of equations of S_1.

$$\text{Log } EC_{50} = 2.075\ (\pm 0.664) + 2.150\ (\pm 0.345)\ S_2 + 1.460\ (\pm 0.370)\ S_{11} + 1.250\ (\pm 0.202)\ I_1 + 1.170\ (\pm 0.230)\ I_2 \quad (4)$$

$n = 84$; $R = 0.860$; %EV = 73.995; $R^2_A = 0.727$;
$F_{(4,79)} = 56.197$; $p < 0.000$; S.E.E. = 0.682
PRESS = 42.269; SSY = 141.256; $R^2_{CV} = 0.701$;
$S_{PRESS} = 0.731$; P.S.E = 0.709

$$\text{Log } EC_{50} = 3.670\ (\pm 0.694) + 1.330\ (\pm 0.359)\ S_2 + 1.033\ (\pm 0.345)\ S_{11} + 1.408\ (\pm 0.184)\ I_1 + 1.213\ (\pm 0.207)\ I_2 - 0.764\ (\pm 0.170)\ I_{OH} \quad (5)$$

$n = 84$; $R = 0.890$; %EV = 79.348; $R^2_A = 0.780$;
$F_{(5,78)} = 59.938$; $p < 0.000$; S.E.E. = 0.612
PRESS = 33.572; SSY = 141.256; $R^2_{CV} = 0.762$;
$S_{PRESS} = 0.656$; P.S.E = 0.632

$$\text{Log } EC_{50} = 3.810\ (\pm 0.669) + 1.288\ (\pm 0.345)\ S_2 + 0.861\ (\pm 0.338)\ S_{11} + 1.398\ (\pm 0.177)\ I_1 + 0.984\ (\pm 0.216)\ I_2 - 0.757\ (\pm 0.163)\ I_{OH} + 0.528\ (\pm 0.195)\ \pi R_1 \quad (6)$$

$n = 84$; $R = 0.900$; %EV = 81.142; $R^2_A = 0.797$;
$F_{(6,77)} = 55.222$; $p < 0.000$; S.E.E. = 0.588
PRESS = 31.881; SSY = 141.256; $R^2_{CV} = 0.774$;
$S_{PRESS} = 0.643$; P.S.E = 0.616

The eqs 4–6 explain up to 73.995%, 79.348%, 81.142% of the variation in activity respectively and the cross-validated R^2 of these equations have significant values to judge the predictive ability of the model.

Among ETSA indices, S_3 (ETSA index of atom number 3) has higher individual correlation to the biological activity as revealed by correlation matrix. As indicator parameter I_{OH} is highly correlated with S_3 they cannot be used together otherwise almost similar type of models (eqs 7–9) obtained incase of S_3 and each model indicates that increase biological activity corresponds to lower value of S_3 and influence of other parameters are also same as before.

$$\text{Log } EC_{50} = 3.577\ (\pm 0.266) + 3.800\ (\pm 0.437)\ S_3 + 1.352\ (\pm 0.198)\ I_1 + 1.148\ (\pm 0.231)\ I_2 \quad (7)$$

$n = 84$; $R = 0.859$; %EV = 73.767; $R^2_A = 0.728$;
$F_{(3,80)} = 74.985$; $p < 0.000$; S.E.E. = 0.681
PRESS = 40.403; SSY = 141.256; $R^2_{CV} = 0.714$;
$S_{PRESS} = 0.711$; P.S.E = 0.694

$$\text{Log } EC_{50} = 2.137\ (\pm 0.644) + 3.263\ (\pm 0.477)\ S_3 + 0.940\ (\pm 0.385)\ S_{11} + 1.286\ (\pm 0.194)\ I_1 + 1.115\ (\pm 0.224)\ I_2 \quad (8)$$

$n = 84$; $R = 0.870$; %EV = 75.607; $R^2_A = 0.744$;
$F_{(4,79)} = 61.216$; $p < 0.000$; S.E.E. = 0.660
PRESS = 38.649; SSY = 141.256; $R^2_{CV} = 0.726$;
$S_{PRESS} = 0.699$; P.S.E = 0.678

$$\text{Log } EC_{50} = 2.808\ (\pm 0.709) + 2.569\ (\pm 0.574)\ S_3 + 0.857\ (\pm 0.379)\ S_{11} + 1.395\ (\pm 0.197)\ I_1 + 1.188\ (\pm 0.223)\ I_2 + 0.262\ (\pm 0.126)\ \pi R_2 \quad (9)$$

Table 2. ETSA indices, physicochemical (hydrophobic) and indicator parameters of HEPT analogues

Compd	S_1[a]	S_2[a]	S_3[a]	S_4[a]	S_5[a]	S_6[a]	S_{11}[a]	S_{av}[b]	πR_2	πR_1	I_1	I_2	I_{OH}
1	−0.3842	0.4924	0.6492	1.3880	−0.5044	2.2693	1.9344	0.6517	−1.020	0.1500	0.00	0.00	1.00
2	−0.4708	0.3439	0.3022	1.2382	−0.5910	2.2139	1.8334	0.5061	−1.020	0.1500	0.00	0.00	1.00
3	−0.4183	0.4254	0.5226	1.3435	−0.5300	2.2484	1.8941	0.5986	−1.020	0.5100	0.00	0.00	0.00
4	−0.4384	0.4196	0.5026	1.3748	−0.5466	2.3557	1.9088	0.6113	−1.020	−0.1600	0.00	0.00	1.00
5	−0.7070	−0.0584	0.2339	1.1985	−0.6643	2.1712	1.8182	0.3623	−1.020	1.1200	0.00	0.00	1.00
6	−0.7533	−0.1314	0.1877	1.1852	−0.6881	2.1861	1.8114	0.3310	−1.020	1.0800	0.00	0.00	1.00
7	−0.4014	−0.5029	0.5396	1.3956	−0.5333	2.3765	1.9181	0.4790	−1.020	1.0000	0.00	0.00	1.00
8	−0.4796	0.3515	0.4613	1.3829	−0.5691	2.4200	1.9146	0.5945	−1.020	1.7400	0.00	0.00	1.00
9	−0.5260	0.2170	0.4150	1.2998	−0.5751	2.2386	1.8733	0.5933	−1.020	0.4500	0.00	0.00	1.00
10	−0.5499	0.1859	0.3910	1.3090	−0.5903	2.2900	1.8785	0.5877	−1.020	0.9100	0.00	0.00	1.00
11	−0.4631	0.3665	0.4778	1.3603	−0.5564	2.3413	1.9026	0.5877	−1.020	1.1400	0.00	0.00	1.00
12	−0.4883	0.3073	0.4526	1.3110	−0.5622	2.2359	1.8770	0.5427	−1.020	0.3300	0.00	0.00	1.00
13	−0.4026	0.4595	0.5434	1.3644	−0.5227	2.2706	1.0586	0.6188	−1.020	0.1500	0.00	0.00	1.00
14	−0.4384	0.4110	0.4742	1.3227	−0.5585	2.2533	0.6466	0.5774	−1.020	0.1500	0.00	0.00	1.00
15	−0.4034	0.4579	0.5400	1.3603	−0.5236	2.2673	1.9807	0.6164	−1.020	0.1500	0.00	0.00	1.00
16	−0.3997	0.4629	0.5471	1.3705	−0.5199	2.2785	2.0362	0.6232	−1.020	0.1500	0.00	0.00	1.00
17	−0.4040	0.4576	0.5404	1.3756	−0.5242	2.2916	2.0640	0.6228	−1.020	0.1500	0.00	0.00	1.00
18	−0.4364	0.4148	0.4814	1.3172	−0.5566	2.2413	1.7461	0.5770	−1.020	0.1500	0.00	0.00	1.00
19	−0.4151	0.4425	0.5192	1.3440	−0.5353	2.2581	1.8973	0.6022	−1.020	0.1500	0.00	0.00	1.00
20	−0.4053	0.4554	0.5367	1.3578	−0.5255	2.2658	1.9673	0.6142	−1.020	0.1500	0.00	0.00	1.00
21	−0.5351	0.2912	0.3223	1.2040	−0.6553	2.1773	1.3557	0.4674	−1.020	0.1500	0.00	0.00	1.00
22	−0.4659	0.3762	0.4289	1.2787	−0.5861	2.2179	1.5362	0.5416	−1.020	0.1500	0.00	0.00	1.00
23	−0.4310	0.4220	0.4915	1.3297	−0.5511	2.2538	1.8140	0.5858	−1.020	0.1500	0.00	0.00	1.00
24	−0.4018	0.4603	0.5441	1.3695	−0.5220	2.2787	2.0015	0.6215	−1.020	0.1500	1.00	0.00	1.00
25	−0.4678	0.3741	0.4268	1.2834	−0.5880	2.2266	1.6825	0.5425	−1.020	0.1500	1.00	0.00	1.00
26	−0.1935	0.5936	0.7524	1.7399	0.3113	2.6490	2.0695	0.9754	−1.020	0.1500	1.00	0.00	1.00
27	−0.4900	0.3486	0.3978	1.2655	−0.6102	2.2206	1.6114	0.5220	−1.020	0.1500	0.00	0.00	1.00
28	−0.4682	0.3761	0.4336	1.2890	−0.5883	2.2314	1.7052	0.5456	−1.020	0.1500	0.00	0.00	1.00
29	−0.4647	0.3799	0.4377	1.2876	−0.5849	2.2261	1.6751	0.5469	−1.020	0.1500	0.00	0.00	1.00
30	−0.4236	0.4601	0.5174	1.3545	−0.5416	2.2933	1.8957	0.6100	−1.020	0.7300	0.00	0.00	1.00
31	−0.8292	−0.2784	0.1117	1.1119	−0.7177	2.0706	1.7740	0.2448	−1.020	−0.5900	0.00	0.00	1.00
32	−0.7907	−0.1984	0.1503	1.1611	−0.7051	2.1621	1.7996	0.2966	−1.020	−0.7400	0.00	0.00	1.00
33	−0.1620	0.6767	0.7790	1.7526	0.3188	2.6775	1.9793	1.0071	−1.020	0.5500	0.00	0.00	1.00
34	−0.1588	0.6816	0.7871	1.7711	0.3220	2.7003	2.0834	1.0172	−1.020	0.5500	1.00	0.00	1.00
35	−0.1610	0.6864	0.7849	1.7816	0.3194	2.7307	2.0880	1.0237	−1.020	0.8800	1.00	1.00	1.00
36	−0.2247	0.5955	0.6698	1.6850	0.2560	2.6482	1.7644	0.9383	−1.020	0.5500	1.00	0.00	1.00
37	−0.3523	0.5781	0.5886	1.4023	−0.5073	2.3411	1.9215	0.6751	−1.020	0.9400	0.00	0.00	1.00
38	−0.3643	0.5058	0.5997	1.4109	−0.4614	2.2763	1.9234	0.6612	−0.5700	0.1500	0.00	0.00	0.00
39	−0.3556	0.5179	0.6176	1.4622	0.4435	2.3184	1.9374	0.8340	0.2400	0.1500	0.00	0.00	0.00
40	−0.3528	0.5216	0.6226	1.4760	−0.4385	2.3324	1.9424	0.6936	0.6300	0.1500	0.00	0.00	0.00
41	−0.3718	0.4977	0.5916	1.4555	−0.4695	2.3402	1.9354	0.6739	0.4400	0.1500	0.00	0.00	0.00
42	−0.1131	0.7391	0.8617	1.8622	0.3966	2.7448	2.1172	1.0819	−0.2300	0.5500	1.00	0.00	0.00
43	−0.1791	0.6530	0.7444	1.7761	0.3306	2.6926	1.7982	1.0029	−0.2300	0.5500	1.00	0.00	0.00
44	−0.0679	0.8046	0.9649	2.0533	0.5047	2.7985	2.0496	1.1764	1.270	0.5500	0.00	0.00	0.00
45	−0.0970	0.7588	0.8862	1.9264	0.4260	2.8016	2.0431	1.1170	1.330	0.5500	0.00	0.00	0.00
46	−0.1287	0.7191	0.8347	1.8571	0.3745	2.7618	2.0173	1.0698	0.4400	0.5500	0.00	0.00	0.00
47	−0.1255	0.7240	0.8428	1.8756	0.3777	2.7845	2.1213	1.0799	0.4400	0.5500	1.00	0.00	0.00
48	−0.1283	0.7196	0.8354	1.8623	0.3753	2.7692	2.0196	1.0723	0.5900	0.5500	0.00	0.00	0.00
49	−0.1429	0.7022	0.8143	1.8362	0.3542	2.7518	2.0071	1.0526	0.2200	0.5500	0.00	0.00	0.00
50	−0.1229	0.7268	0.8455	1.8730	0.3853	2.7717	2.0227	1.0799	0.6900	0.5500	0.00	0.00	0.00
51	−0.1185	0.7390	0.8513	1.8542	0.3908	2.7525	2.0178	1.0782	−0.2300	0.8800	0.00	1.00	0.00
52	−0.1310	0.7238	0.8324	1.8675	0.3719	2.7922	2.0219	1.0761	0.4400	0.8800	0.00	1.00	0.00
53	−0.0560	0.8501	0.9138	1.8942	0.4186	2.7925	2.0382	1.1355	−0.2300	0.3700	0.00	0.00	0.00
54	−0.3264	0.5987	0.6426	1.4838	−0.4424	2.3665	1.9477	0.7205	0.1800	0.5500	0.00	0.00	0.00
55	−0.3050	0.6264	0.6798	1.5618	−0.4052	2.4305	1.9742	0.7647	0.4300	0.5500	0.00	0.00	0.00
56	−0.3053	0.6255	0.6778	1.5561	−0.4073	2.4312	1.9751	0.7630	1.330	0.5500	0.00	0.00	0.00
57	−0.3312	0.5935	0.6371	1.5027	−0.4479	2.4014	1.9546	0.7259	0.6900	0.5500	0.00	0.00	0.00
58	−0.3159	0.5751	0.7068	1.5776	−0.3543	2.3224	1.9540	0.7519	0.5500	0.1500	0.00	0.00	0.00
59	−0.3038	0.5929	0.7355	1.6644	−0.3256	2.3804	1.9721	0.7906	1.340	0.1500	0.00	0.00	0.00
60	−0.3495	0.5804	0.6539	1.4193	−0.4936	2.3205	1.9483	0.6885	−1.020	0.5500	0.00	0.00	1.00
61	−0.3463	0.5853	0.6621	1.4378	−0.4904	2.3433	2.0524	0.6986	−1.020	0.5500	1.00	0.00	1.00
62	−0.3038	0.6379	0.7285	1.5104	−0.4191	2.3650	1.9821	0.7532	−0.2300	0.5500	0.00	0.00	0.00
63	−0.3517	0.5851	0.6517	1.4297	−0.4962	2.3510	1.9529	0.6949	−1.020	0.8800	0.00	1.00	1.00
64	−0.3485	0.5901	0.6598	1.4482	−0.4930	2.3737	2.0570	0.7051	−1.020	0.8800	1.00	1.00	1.00
65	−0.3060	0.6426	0.7262	1.5208	−0.4216	2.3954	1.9868	0.7596	−0.2300	0.8800	0.00	1.00	0.00
66	−0.3028	0.6476	0.7343	1.5393	−0.4185	2.4182	2.0908	0.7697	−0.2300	0.8800	1.00	1.00	0.00
67	−0.2505	0.7226	0.8512	1.7431	−0.2966	2.4754	2.0276	0.8742	1.340	0.8800	0.00	1.00	0.00
68	−0.2890	0.6623	0.7735	1.6076	−0.3741	2.3894	1.9917	0.7949	−0.3200	0.5500	0.00	0.00	0.00
69	−0.2913	0.6671	0.7712	1.6181	−0.3766	2.4199	1.9963	0.8014	−0.3200	0.8800	0.00	1.00	0.00

(*continued on next page*)

Table 2 (*continued*)

Compd	S_1[a]	S_2[a]	S_3[a]	S_4[a]	S_5[a]	S_6[a]	S_{11}[a]	S_{av}[b]	πR_2	πR_1	I_1	I_2	I_{OH}
70	−0.4050	0.4554	0.5359	1.3510	−0.5252	2.2559	1.8974	0.6113	−1.020	0.1500	0.00	0.00	1.00
71	−0.6064	0.0712	0.3346	1.2398	−0.6100	2.1647	1.8427	0.4323	−1.020	−0.1300	0.00	0.00	1.00
72	−0.4972	0.3354	0.3733	1.2378	−0.6173	2.1932	1.3140	0.5042	−1.020	0.1500	0.00	0.00	1.00
73	−0.1642	0.6815	0.7767	1.7631	0.3163	2.7080	1.9840	1.0136	−1.020	0.8800	0.00	1.00	1.00
74	−0.3775	0.4904	0.5817	1.4205	−0.4794	2.3012	1.9224	0.6561	−0.7400	0.1500	0.00	0.00	0.00
75	−0.3593	0.5129	0.6105	1.4422	−0.4506	2.3004	1.9312	0.6760	−0.2300	0.1500	0.00	0.00	0.00
76	−0.1163	0.7342	0.8535	1.8437	0.3934	2.7220	2.0131	1.0718	−0.2300	0.5500	0.00	0.00	0.00
77	−0.3246	0.6009	0.6452	1.4734	−0.4399	2.3517	1.9451	0.7178	−0.2300	0.5500	0.00	0.00	0.00
78	−0.3214	0.6058	0.6533	1.4919	−0.4367	2.3744	2.0492	0.7279	−0.2300	0.5500	1.00	0.00	0.00
79	−0.3338	0.5906	0.6345	1.5052	−0.4556	2.4142	2.0533	0.7258	0.4400	0.5500	1.00	0.00	0.00
80	−0.3268	0.6056	0.6430	1.4838	−0.4424	2.3821	1.9498	0.7242	−0.2300	0.8800	0.00	1.00	0.00
81	−0.2643	0.7168	0.7055	1.5238	−0.4147	2.4221	1.9702	0.7815	−0.2300	0.3700	0.00	0.00	0.00
82	−0.3214	0.6058	0.6534	1.4919	−0.4367	2.3744	2.0492	0.7279	−0.2300	0.5500	1.00	0.00	0.00
83	−0.2482	0.7179	0.8535	1.7326	−0.2941	2.4450	2.0229	0.8678	1.340	0.5500	0.00	0.00	0.00
84	−0.1967	0.5887	0.7442	1.7214	0.3081	2.6263	1.9654	0.9653	−1.020	0.1500	0.00	0.00	1.00

[a] S_1, S_2, S_3, S_4, S_5, S_6, S_{11} are ETSA indices of atoms 1, 2, 3, 4, 5, 6, 11 respectively.
[b] S_{av} = average of ETSA indices of atoms 1, 2, 3, 4, 5 and 6.

Table 3. Correlation matrix of the ETSA indices, hydrophobic and indicator parameters and biological activity

	$logEC_{50}$	S_1	S_2	S_3	S_4	S_5	S_6	S_{11}	I_{OH}	S_{av}	I_1	I_2	πR_2	πR_1
$LogEC_{50}$	1.00	0.67	0.64	0.70	0.68	0.56	0.67	0.54	−0.55	0.68	0.54	0.50	0.42	0.47
S_1		1.00	0.86	0.97	0.94	0.83	0.89	0.41	−0.55	0.97	0.24	0.24	0.52	0.31
S_2			1.00	0.86	0.75	0.61	0.67	0.33	−0.56	0.82	0.21	0.26	0.51	0.23
S_3				1.00	0.93	0.76	0.84	0.46	−0.62	0.94	0.21	0.30	0.60	0.34
S_4					1.00	0.91	0.96	0.46	−0.58	0.98	0.21	0.24	0.62	0.33
S_5						1.00	0.94	0.36	−0.39	0.93	0.23	0.15	0.41	0.20
S_6							1.00	0.43	−0.45	0.96	0.26	0.24	0.47	0.38
S_{11}								1.00	−0.40	0.43	0.23	0.21	0.35	0.31
I_{OH}									1.00	−0.54	−0.01	−0.14	−0.83	−0.18
S_{av}										1.00	0.24	0.24	0.54	0.31
I_1											1.00	0.10	−0.07	0.11
I_2												1.00	0.07	0.45
πR_2													1.00	0.14
πR_1														1.00

$n = 84$; R = 0.877; %EV = 76.892; $R^2_A = 0.754$;
$F_{(5,78)} = 51.910$; $p < 0.000$; S.E.E. = 0.647
PRESS = 37.568; SSY = 141.256; $R^2_{CV} = 0.734$;
$S_{PRESS} = 0.694$; P.S.E = 0.669

Though the variance ratio (F) has lower value in case of eq 9 comparing with eqs 7 and 8, it has higher cross-validated R^2 values and higher correlation coefficient, which in turn increases the explained variance (%EV) shown above and has decreased values of PRESS and S_{PRESS}. Thus, eq 9 is the best model using S_3 as independent parameter, which explains 76.892% of the variation in activity.

Along with atoms 1–3, atom 4 has also important contribution to the structure–activity relationship of these type of compounds. The only distinct feature is that S_{11} (ETSA index of atom number 11) cannot be used with the combination of parameters used because the resultant models give associated probability value (p) in case of S_{11} greater than 0.05. The models (10, 11 and 12) obtained by considering all possible combinations of S_4 and other independent parameters with biological activity, are given below. The eqs 10 and 11 explain up to 72.996%, 79.070% of the variation in activity respectively. Among the models, the best model (eq 12) explains 80.60% of the variation in activity.

$$\text{Log } EC_{50} = 1.283\ (\pm 0.539) + 3.062\ (\pm 0.363)\ S_4 + 1.344\ (\pm 0.202)\ I_1 + 1.279\ (\pm 0.230)\ I_2 \quad (10)$$

$n = 84$; R = 0.854; %EV = 72.996; $R^2_A = 0.720$;
$F_{(3,80)} = 72.083$; $p < 0.000$; S.E.E. = 0.691
PRESS = 42.005; SSY = 141.256; $R^2_{CV} = 0.703$;
$S_{PRESS} = 0.725$; P.S.E = 0.707

$$\text{Log } EC_{50} = 3.353\ (\pm 0.644) + 1.954\ (\pm 0.396)\ S_4 + 1.476\ (\pm 0.181)\ I_1 + 1.281\ (\pm 0.204)\ I_2 - 0.797\ (\pm 0.166)\ I_{OH} \quad (11)$$

$n = 84$; R = 0.889; %EV = 79.070; $R^2_A = 0.780$;
$F_{(4,79)} = 74.612$; $p < 0.000$; S.E.E. = 0.612
PRESS = 33.236; SSY = 141.256; $R^2_{CV} = 0.765$;
$S_{PRESS} = 0.648$; P.S.E = 0.629

$$\text{Log } EC_{50} = 3.472\ (\pm 0.626) + 1.752\ (\pm 0.392)\ S_4 + 1.465\ (\pm 0.175)\ I_1 + 1.069\ (\pm 0.215)\ I_2 - 0.804\ (\pm 0.161)\ I_{OH} + 0.489\ (\pm 0.196)\ \pi R_1 \quad (12)$$

$n = 84$; R = 0.898; %EV = 80.606; $R^2_A = 0.794$;
$F_{(5,78)} = 64.838$; $p < 0.000$; S.E.E. = 0.593
PRESS = 31.346; SSY = 141.256; $R^2_{CV} = 0.778$;
$S_{PRESS} = 0.634$; P.S.E = 0.611

The best model obtained in case of S_5 and S_6 (ETSA indices of atom number 5 and 6 respectively) taken as independent parameter separately are given below (eqs 13 and 14 respectively) explaining 76.69% and 73.93% of variation in activity. The predictive ability was also good as evidenced by R^2_{CV} values (greater than 0.70).

$$\log EC_{50} = 6.374\ (\pm 0.142) + 0.761\ (\pm 0.193)\ S_5 + 1.488\ (\pm 0.179)\ I_1 + 1.121\ (\pm 0.220)\ I_2 - 0.999\ (\pm 0.146)\ I_{OH} + 0.596\ (\pm 0.198)\ \pi R_1 \quad (13)$$

$n = 84$; R = 0.893; %EV = 79.690; $R^2_A = 0.784$;
$F_{(5,78)} = 61.210$; $p < 0.000$; S.E.E. = 0.606
PRESS = 32.871; SSY = 141.256; $R^2_{CV} = 0.767$;
$S_{PRESS} = 0.649$; P.S.E = 0.626

$$\text{Log } EC_{50} = -2.894\ (\pm 1.001) + 2.684\ (\pm 0.432)\ S_6 + 1.237\ (\pm 0.383)\ S_{11} + 1.203\ (\pm 0.203)\ I_1 + 1.242\ (\pm 0.229)\ I_2 \quad (14)$$

$n = 84$; R = 0.860; %EV = 73.927; $R^2_A = 0.726$;
$F_{(4,79)} = 55.999$; $p < 0.000$; S.E.E. = 0.683
PRESS = 42.212; SSY = 141.256; $R^2_{CV} = 0.701$;
$S_{PRESS} = 0.731$; P.S.E = 0.709

Omission of one of these atoms (atom no 1, 2, 3, 4, 5 and 6) from the model may loose some information regarding the Pharmacophoric requirements for the biological activity. To accommodate these atoms in a single QSAR equation the average of ETSA values (S_{av}) of these atoms is taken and considered as single best variable, which is shown in eqs 15 and 16.

$$\text{Log } EC_{50} = 2.699\ (\pm 0.705) + 2.068\ (\pm 0.441)\ S_{av} + 0.954\ (\pm 0.0.368)\ S_{11} + 1.361\ (\pm 0.197)\ I_1 + 1.271\ (\pm 0.216)\ I_2 + 0.310\ (\pm 0.118)\ \pi R^2 \quad (15)$$

$n = 84$; R = 0.880; %EV = 77.357; $R^2_A = 0.759$;
$F_{(5,78)} = 53.295$; $p < 0.000$; S.E.E. = 0.640
PRESS = 37.043; SSY = 141.256; $R^2_{CV} = 0.738$;
$S_{PRESS} = 0.689$; P.S.E = 0.664

$$\text{Log } EC_{50} = 4.778\ (\pm 0.328) + 1.833\ (\pm 0.387)\ S_{av} + 1.434\ (\pm 0.174)\ I_1 + 1.060\ (\pm 0.213)\ I_2 - 0.826\ (\pm 0.154)\ I_{OH} + 0.499\ (\pm 0.194)\ \pi R_1 \quad (16)$$

$n = 84$; R = 0.900; %EV = 81.090; $R^2_A = 0.799$;
$F_{(5,78)} = 66.894$; $p < 0.000$; S.E.E. = 0.585
PRESS = 30.622; SSY = 141.256; $R^2_{CV} = 0.783$;
$S_{PRESS} = 0.627$; P.S.E = 0.604

All the equations obtained are considered statistically significant because the associated probability values are <0.05 and therefore F is statistically significant. The student t values and associated probability values p of all the equations are given in Table 4. The eq 15 is statistically significant as the variance ratio (F) is more than 95% confidence interval, which is the measure of the level of statistical significance of a regression model. The eq 16 is considered to be the best model where the multiple correlation coefficient is 0.900 and this explains 81.09% of the variation of the biological activity data. Compared to eq 15, eq 16 gives lower values of PRESS and S_{PRESS}, which ensure that optimum number of components are taken and there is no over prediction in the model.

Useful informations are obtained at molecular level to the structural and physicochemical requirements of the HEPT analogues for their anti-HIV activity from the above two equations (eqs 15 and 16). The predictive power of the equations were confirmed by Leave-One-Out (LOO–) cross validation method,[6] where one compound is deleted at once and prediction of the activity of the deleted compound is made based on the QSAR model. The process is repeated after elimination of another compound until all of the compounds have been deleted at once. The observed, calculated, residual, LOO-predicted (pred) and predicted residual (pres) values of the equations are shown in Table 5. These equations (eqs 15 and 16) show that lower value of S_{av} is conducive to the activity. The significances of I_1, I_2 and I_{OH} indicator parameters to the activity are same here as described previously in case of eqs 1–3. Taking into account of the hydrophobic parameters (πR_2 and πR_1 in the eqs 15 and 16 respectively), the interpretation is that greater hydrophobicity of the substituents at R_1 and R_2 positions corresponds to the lesser activity of the HEPT analogues. Another important feature revealed is that the atoms 1, 2, 3, 4, 5, 6 and 11 may constitute the pharmacophore moiety of the HIV-reverse transcriptase inhibitory HEPT analogues as ETSA indices associated with these atoms are useful to generate the significant QSAR models.

3. Conclusions

In order to identify essential chemical structural or substitutional features responsible for the activity

Table 4. *t*- and *p*-values of equations

Eq	Intercept/parameter	*t*-value	*p*-Value	Eq	Intercept/parameter	*t*-value	*p*-Value
1	Intercept	5.975	0.000	9	Intercept	3.964	0.000
	S_1	6.492	0.000		S_3	4.474	0.000
	S_{11}	3.233	0.002		S_{11}	2.259	0.0267
	I_1	6.143	0.000		I_1	7.064	0.000
	I_2	5.428	0.000		I_2	5.337	0.000
2	Intercept	6.512	0.000		πR_2	2.083	0.040
	S_1	4.552	0.000	10	Intercept	2.382	0.019
	S_{11}	2.722	0.008		S_3	8.442	0.000
	I_1	6.931	0.000		I_1	6.668	0.000
	I_2	5.890	0.000		I_2	5.550	0.000
	πR_2	2.837	0.006	11	Intercept	5.207	0.000
3	Intercept	7.464	0.000		S_4	4.934	0.000
	S_1	4.177	0.000		I_{OH}	−4.788	0.000
	S_{11}	2.488	0.015		I_1	8.169	0.000
	I_{OH}	−4.63	0.000		I_2	6.275	0.000
	I_1	7.653	0.000	12	Intercept	5.548	0.000
	I_2	6.163	0.000		S_4	4.468	0.000
4	Intercept	3.123	0.002		I_{OH}	−4.984	0.000
	S_2	6.238	0.000		I_1	8.369	0.000
	S_{11}	3.946	0.000		I_2	4.962	0.000
	I_1	6.199	0.000		πR_1	2.485	0.015
	I_2	5.078	0.000	13	Intercept	44.896	0.000
5	Intercept	5.292	0.000		S_5	3.943	0.000
	S_2	3.707	0.000		I_{OH}	−6.843	0.000
	S_{11}	2.993	0.003		I_1	8.321	0.000
	I_{OH}	−4.49	0.000		I_2	5.098	0.000
	I_1	7.646	0.000		πR_1	3.007	0.003
	I_2	5.864	0.000	14	Intercept	−2.890	0.005
6	S_2	5.695	0.000		S_6	6.213	0.000
	S_{11}	3.727	0.000		S_{11}	3.232	0.002
	I_{OH}	2.548	0.013		I_1	5.927	0.000
	I_1	−4.632	0.000		I_2	5.428	0.000
	I_2	7.892	0.000	15	Intercept	3.826	0.000
	S_2	4.548	0.000		S_{11}	2.589	0.011
	πR_1	2.707	0.008		S_{av}	4.694	0.000
7	Intercept	13.420	0.000		I_1	6.912	0.000
	S_3	8.701	0.000		I_2	5.891	0.000
	I_1	6.814	0.000		πR_2	2.636	0.010
	I_2	4.973	0.000	16	Intercept	14.56	0.000
8	Intercept	3.317	0.001		I_{OH}	−5.35	0.000
	S_3	6.834	0.000		S_{av}	4.740	0.000
	S_{11}	2.441	0.017		I_1	8.239	0.000
	I_1	6.615	0.000		I_2	4.981	0.000
	I_2	4.969	0.000		πR_1	2.575	0.012

QSAR studies on HEPT analogues were performed using atom/ fragmental level descriptors. Thakur et al.[4] used whole molecular descriptors (like Wiener Index, Randic Connectivity Index Log P, Molar Volume, Surface Tension, Molar Refractivity, etc.) for the QSAR study of the same series of HEPT analogues and in which the correlation coefficients (R) are less than 0.85. But the use of ETSA indices and hydrophobicity (π) gave better correlation coefficient ($R > 0.90$) and comparably higher variance ratio (F). A possible explanation of this difference of QSAR models may be that all atoms in a molecule may not be responsible for the activity, rather a part of the structure or some specific atoms are required to elicit a desired effect. These atoms responsible for the activity are called as Pharmacophoric atoms, which may be identified through ETSA indices, shown in Figure 2. So one can propose that atom/fragmental level descriptors are better than whole molecular descriptors to explain structure–activity relationship quantitatively for these HEPT analogues.

4. Experimental

4.1. Data set and parameters

For the development of 2D QSAR model on HEPT analogues for their reverse transcriptase inhibitory activity the electrotopological state atom (ETSA) indices[7] and hydrophobic parameters (π) were used. Furthermore, in order to find out the role of specific substituent/substituent pattern at specific positions towards the activity, some indicator parameters were also used. E-state indices atomwise were calculated using the computer programme 'mouse'.[8] In the program, molecular connection table in specified format and the intrinsic values of different atoms are given as input and output result is the ETSA indices of atoms. Before the calculation the atoms of the molecules were numbered arbitrarily keeping the serial number of atoms same in all molecules. ETSA indices are atom level descriptor and give important pieces of information at molecular level and depend on the value of elec-

Table 5. Observed, Calculated, Residual, LOO- Predicted (Pred), Predicted residual (Pres) values of of eqs 15 and 16

Compd	Obs. value	Equation 15				Equation 16			
		Calc	Res	Pred	Pres	Calc	Res	Pred	Pres
1	4.637	5.576	−0.939	5.612	−1.245	5.221	−0.584	5.249	−0.882
2	4.853	5.178	−0.325	5.188	−0.335	4.954	−0.101	4.957	−0.104
3	5.443	5.428	0.015	5.427	0.016	6.129	−0.686	6.169	−0.726
4	4.677	5.468	−0.791	5.489	−0.813	4.992	−0.315	5.010	−0.333
5	4.920	4.866	0.053	4.864	0.056	5.175	−0.254	5.213	−0.293
6	4.885	4.795	0.090	4.790	0.095	5.097	−0.212	5.130	−0.245
7	4.637	5.203	−0.566	5.225	−0.588	5.329	−0.692	5.394	−0.757
8	6.075	5.439	0.636	5.421	0.654	5.910	0.165	5.857	0.218
9	4.720	5.397	−0.677	5.414	−0.694	5.264	−0.544	5.279	−0.559
10	5.468	5.390	0.078	5.388	0.080	5.483	−0.015	5.484	−0.016
11	5.221	5.413	−0.192	5.418	−0.197	5.598	−0.377	5.638	−0.417
12	5.958	5.296	0.662	5.277	0.681	5.111	0.847	5.088	0.870
13	4.148	4.672	−0.524	4.810	−0.662	5.161	−1.013	5.191	−1.043
14	4.720	4.193	0.527	3.747	0.973	5.085	−0.365	5.096	−0.376
15	5.584	5.547	0.037	5.546	0.038	5.156	0.428	5.143	0.440
16	5.568	5.614	−0.046	5.616	−0.048	5.169	0.399	5.157	0.411
17	4.920	5.639	−0.720	5.674	−0.754	5.168	−0.248	5.176	−0.256
18	4.885	5.242	−0.357	5.250	−0.365	5.084	−0.199	5.090	−0.205
19	5.243	5.438	−0.195	5.443	−0.200	5.130	0.116	5.127	0.116
20	4.999	5.529	−0.530	5.547	−0.548	5.152	−0.153	5.157	−0.158
21	4.468	4.643	−0.175	4.659	−0.191	4.883	−0.415	4.898	−0.430
22	4.085	4.968	−0.883	5.010	−0.925	5.019	−0.934	5.048	−0.963
23	4.657	5.325	−0.668	5.341	−0.684	5.100	−0.443	5.113	−0.456
24	6.584	6.938	−0.354	6.971	−0.387	6.599	−0.016	6.601	−0.017
25	5.885	6.470	−0.585	6.540	−0.655	6.455	−0.570	6.516	−0.631
26	6.656	7.735	−1.079	7.854	−1.198	7.249	−0.593	7.330	−0.674
27	5.102	4.999	0.102	4.996	0.106	4.983	0.118	4.980	0.122
28	5.136	5.138	−0.002	5.138	−0.002	5.027	0.109	5.023	0.113
29	4.999	5.112	−0.113	5.115	−0.116	5.029	−0.030	5.030	−0.031
30	5.601	5.453	0.148	5.449	0.152	5.434	0.167	5.427	0.174
31	5.180	4.581	0.599	4.528	0.652	4.106	1.073	3.923	1.257
32	4.744	4.713	0.031	4.710	0.033	4.126	0.618	4.005	0.739
33	6.957	6.354	0.603	6.289	0.668	6.072	0.885	5.977	0.980
34	8.106	7.835	0.271	7.802	0.304	7.525	0.581	7.448	0.658
35	8.300	9.123	−0.823	9.272	−0.972	8.761	−0.461	8.858	−0.558
36	7.365	7.367	−0.002	7.367	−0.002	7.380	−0.015	7.382	−0.017
37	5.468	5.612	−0.144	5.616	−0.148	5.658	−0.190	5.671	−0.203
38	5.677	5.724	−0.047	5.725	−0.048	6.064	−0.387	6.083	−0.406
39	5.443	6.346	−0.903	6.371	−0.928	6.381	−0.938	6.420	−0.977
40	5.327	6.182	−0.855	6.230	−0.903	6.124	−0.797	6.159	−0.832
41	7.050	6.076	0.978	6.030	1.024	6.088	0.966	6.042	1.012
42	8.355	8.246	0.109	8.234	0.121	8.469	−0.114	8.481	−0.126
43	7.885	7.778	0.107	7.766	0.119	8.324	−0.439	8.367	−0.482
44	6.656	7.481	−0.825	7.574	−0.918	7.208	−0.552	7.259	−0.603
45	6.455	7.371	−0.916	7.466	−1.011	7.099	−0.644	7.147	−0.692
46	8.106	6.972	1.134	6.903	1.203	7.013	1.093	6.945	1.161
47	8.160	8.453	−0.293	8.485	−0.325	8.465	−0.305	8.497	−0.337
48	7.107	7.026	0.081	7.021	0.086	7.017	0.089	7.012	0.095
49	7.919	6.859	1.060	6.798	1.121	6.981	0.938	6.927	0.992
50	7.040	7.076	−0.036	7.078	−0.038	7.031	0.009	7.031	0.009
51	7.852	8.053	−0.201	8.081	−0.229	8.253	−0.401	8.305	−0.453
52	8.179	8.261	−0.082	8.271	−0.093	8.249	−0.070	8.258	−0.079
53	7.021	6.920	0.100	6.910	0.111	7.043	−0.022	7.045	−0.024
54	6.467	6.103	0.364	6.093	0.374	6.372	0.094	6.369	0.098
55	5.397	6.297	−0.900	6.330	−0.933	6.454	−1.057	6.492	−1.094
56	6.346	6.574	−0.228	6.600	−0.254	6.450	−0.104	6.454	−0.108
57	7.016	6.279	0.737	6.237	0.778	6.382	0.633	6.358	0.658
58	5.657	6.288	−0.632	6.316	−0.659	6.231	−0.573	6.254	−0.597
59	5.920	6.631	−0.711	6.710	−0.790	6.302	−0.382	6.317	−0.397
60	6.455	5.665	0.789	5.640	0.815	5.488	0.967	5.455	0.999
61	7.885	7.146	0.739	7.081	0.804	6.941	0.944	6.856	1.029
62	7.386	6.076	1.310	6.049	1.336	6.432	0.954	6.397	0.989
63	7.199	6.954	0.245	6.923	0.276	6.725	0.474	6.662	0.537
64	8.567	8.435	0.132	8.411	0.156	8.178	0.389	8.104	0.463
65	8.375	7.365	1.010	7.252	1.123	7.669	0.706	7.579	0.796
66	9.220	8.846	0.374	8.776	0.443	9.122	0.098	9.101	0.118
67	7.374	8.128	−0.754	8.288	−0.914	7.879	−0.505	7.938	−0.564
68	6.601	6.144	0.457	6.134	0.467	6.509	0.092	6.506	0.095

(*continued on next page*)

Table 5 (*continued*)

Compd	Obs. value	Equation 15				Equation 16			
		Calc	Res	Pred	Pres	Calc	Res	Pred	Pres
69	7.282	7.432	−0.150	7.448	−0.167	7.746	−0.464	7.802	−0.520
70	5.154	5.457	−0.303	5.465	−0.311	5.147	0.007	5.147	0.007
71	4.744	5.034	−0.291	5.046	−0.302	4.679	0.065	4.675	0.068
72	5.481	4.679	0.802	4.593	0.888	4.951	0.530	4.934	0.547
73	7.228	7.642	−0.414	7.714	−0.486	7.309	−0.081	7.324	−0.096
74	5.060	5.660	−0.600	5.673	−0.613	6.055	−0.995	6.104	−1.044
75	6.480	5.868	0.612	5.856	0.624	6.091	0.389	6.073	0.407
76	7.584	6.765	0.819	6.703	0.881	7.016	0.568	6.981	0.603
77	7.720	5.968	1.752	5.935	1.785	6.368	1.352	6.313	1.406
78	8.266	7.448	0.817	7.373	0.893	7.820	0.446	7.766	0.500
79	8.493	7.656	0.837	7.539	0.954	7.816	0.676	7.734	0.759
80	7.919	7.256	0.663	7.180	0.738	7.604	0.315	7.562	0.357
81	6.999	6.123	0.875	6.106	0.893	6.394	0.604	6.374	0.624
82	7.199	7.449	−0.250	7.472	−0.273	7.820	−0.621	7.896	−0.697
83	6.677	6.839	−0.162	6.855	−0.178	6.642	0.034	6.641	0.036
84	6.008	6.254	−0.246	6.276	−0.268	5.796	0.212	5.774	0.234

Obs = observed; Calc = calculated; Res = residual; Pred = LOO-predicted; Pres = predicted residual value.

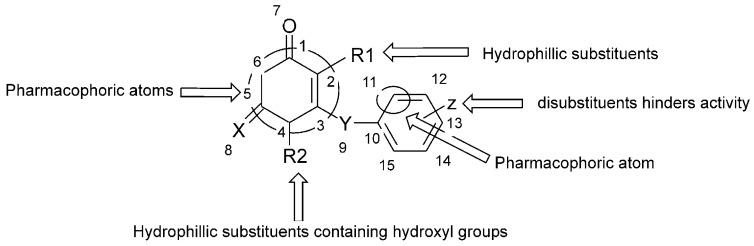

Figure 2. Pharmacophoric atoms and required physicochemical properties of substituents at R_1, R_2 and Z positions of HEPT analogues for their reverse transcriptase inhibition activity.

tronegativity distributed over an atom according to its bonding degree of non-hydrogen atoms. The hydrophobic parameters of the substituents were calculated by HyperChem professional[9] by subtracting the respective compound from that of parent structure due to the additive nature of π.

4.2. Correlation analysis

Correlation analysis of ETSA indices, hydrophobic parameters and indicator parameters was carried out. Autocorrelated parameters were eliminated stepwise depending on their individual correlation with the biological activity. All possible combinations of parameters were considered for multiple regression analysis.

4.3. Multiple regression analysis

Multiple regression analysis[10] was carried out by 'Multi Regress' a programme developed in our laboratory. The statistical quality of the regression equation were justified by parameters like correlation coefficient (R), percentage of explained variance (%EV), adjusted R^2 (R^2_A), variance ratio (F), standard error estimate (S.E.E) All the final equations have significant regression coefficients, intercepts and variance ratio (F) and that are more than 95% level. Use of more than one variable in the multivariate equation was justified by autocorrelation study. The predictive powers of the equation are validated by Leave-One-Out (LOO–) cross validation method. Predicted residual sum of square (PRESS), total sum of squares (SSY), cross-validated R^2 (R^2_{CV}), standard error of PRESS (S_{PRESS}) and predictive standard error or uncertainty factor (P.S.E) for the final equations are considered for the validation of the models.

Acknowledgements

Authors are thankful to the University authority for providing financial support to this work through a Minor Research Project from Unassigned Grants of University Grants Commission (U.G.C), New Delhi, India.

References and notes

1. (a) Font, M.; Monge, A.; Ruiz, I.; Heras, B. *Drug Design Discov*. **1997**, *14*, 259. (b) Kireev, D. B.; Chretien, J. R.; Grierson, D. S.; Monneret, C. *J. Med. Chem*. **1997**, *40*, 4257. (c) Hannongbua, S.; Pungpo, P.; Limitrakul, J.; Wolschann, P. *J. Comput. Aid. Mol. Des*. **1999**, *13*, 563.
2. (a) Pungpo, P.; Hannongbua, S.; wolschann, P. *Curr. Med. Chem*. **2003**, *10*, 1661. (b) Ren, J.; Esnouf, R.; Gar-

man, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; jones, Y.; Stuart, D.; Stammers, D. *Nat. Struc. Biol.* **1995**, *2*, 293. (c) Smerdon, S. J.; Jagger, J.; Wang, J.; Kohlstaedt, L. A.; Chirino, A. J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. *Proc. Natl. Acad. Sci.* **1994**, *91*, 3911.

3. (a) Nanni, R. G.; Ding, J.; Jacobo-Molina, A.; Hughes, S. H.; Arnold, E. A. *Perspect. in Drug Discov. And Des.* **1999**, *1*, 129. (b) Hannongbua, S.; Nivesanond, K.; Lawtrakul, L.; Pungpo, P.; Wolschann, P. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 848.
4. Thakur, A.; Thakur, M. *Bioinform. India* **2003**, *1*, 63.
5. (a) Purkayastha, S. K.; Jha, T.; Pal, D. K.; De, A. U. *Anticancer Drug Des.* **1993**, *8*, 95. (b) Srkanth, K.; Kumar, C. A.; Goswami, D.; De, A. U.; Jha, T. *Ind. J. Biochem. Biophys* **2001**, *38*, 12. (c) Srikanth, K.; Debnath, B.; Jha, T. *Bioorg. Med. Chem.* **2002**, *10*, 1841. (d) Srikanth, K.; Debnath, B.; Jha, T. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 899. (e) Srikanth, K.; Kumar, C. A.; Ghosh, B.; Jha, T. *Bioorg. Med. Chem.* **2002**, *10*, 2119. (f) Debnath, B.; Srikanth, K.; Banerjee, S.; Jha, T. *Internet Electron. J. Mol. Des.* **2002**, *1*, 488. (g) Debnath, B.; Vishnoi, S. P.; Sa, B.; Jha, T. *Internet Electron. J. Mol. Des.* **2003**, *1*, 128. (h) Debnath, B.; Samanta, S.; Roy, K.; Jha, T. *Bioorg. Med. Chem.* **2003**, *11*, 1615. (i) Debnath, B.; Gayen, S.; Naskar, S.K.; Roy, K.; Jha, T. *Drug Des. Discov.* **2003**, *18*, 81. (j) Debnath, B.; Samanta, S.; Naskar, S. K.; Roy, K.; Jha, T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2837. (k) Jha, T.; Debnath, B.; Samanta, S.; De, A. U. *Internet Electron. J. Mol. Des* **2003**, *2*, 539. (l) Debnath, B.; Gayen, S.; Samanta, S.; Jha, T. *Bioorg. Med. Chem.* **2003**, *11*, 5493.
6. (a) Schaper, K. J. *Quant. Struct. Act. Relat* **1999**, *18*, 354. (b) Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407.
7. (a) Hall, L. H.; Mohney, B.; Kier, L. B. *Quant. Struct. Act. Relat* **1991**, *10*, 43. (b) de Gregorio, C.; Kier, L. B.; Hall, L. H. *J. Comput. Aided. Mol. Des.* **1998**, *12*, 557. (c) Rose, K.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651. (d) Roy, K.; De, A. U.; Sengupta, C. *Drug Des. Discov.* **2002**, *18*, 33. (e) Roy, K.; Pal, D. K.; Sengupta, C. *Drug Des. Discov.* **2001**, *17*, 207. (f) Roy, K.; Pal, D. K.; Sengupta, C. *Drug Des. Discov.* **2000**, *17*, 183.
8. Mouse, a computer program written in C^{++} language by S. K. Naskar of Computer science and Engineering department, Jadavpur University and B. Debnath, one of the authors of this article.
9. HyperChem Professional of Hypercube, Inc.; http://www.hyper.com.
10. (a) Snedecor, G. W.; Cochran, W. G. *Statistical Methods*; Oxford & IBH: New Delhi, 1967; p 381. (b) Rawlings, J. O. *Applied Regression Analysis*; Wardsworth: CA: USA, 1988; p 327. (c) Darper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, 1981; p 294. (d) Lien, E. J. In *Quantitative Correlations Using Multiple Regression and Substituent Constants;* Ariens, E. J., Drug Design; Academic: New York, 1975; Vol. V, pp 84. (e) Montgomery, D. C., Ed. *Design and Analysis of Experiments*, 4th ed.; John Wiley and Sons: New York, 1997.