Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Bioorganic & Medicinal Chemistry 12 (2004) 745–754
Bioorganic & Medicinal Chemistry

QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners using molecular connectivity and E-state parameters

Kunal Roy* and J. Thomas Leonard

Drug Theoretics & Cheminformatics Lab, Division of Medicinal & Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Received 20 September 2003; revised 12 November 2003; accepted 13 November 2003

Abstract—Anti-HIV-1 activity (assayed in MT-4 cell line) and HIV-1 reverse transcriptase (RT) binding affinity of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners (Chan et al., *J. Med. Chem.*, **2001**, *44*, 1866) have been modeled using E-state index along with molecular connectivity and indicator parameters in an attempt to explore the different fragments of the molecules contributing significantly to the activities. Stepwise multiple regression procedure was adopted to develop the topological models. The models generated were of acceptable statistical quality and predictive potential. The results show that for both the response variables, first order fragmental valence molecular connectivity of the *meta* substituents of the aryl ring plays a significant role: second *meta* substituents show supraadditive action on the activities probably due to enhanced binding (presumably through dispersion interaction) of the ligand with the binding site. Again, presence of sulfone moiety contributes significantly to the activities. Further, presence of *meta*-trifluoromethyl group at the aryl ring is detrimental for both the activity parameters. Additionally, the anti-HIV-1 model shows specific contributions of the E-state values of different atoms and positive contribution of the *ortho*-methoxy group present on the aryl ring.

1. Introduction

Acquired immunodeficiency syndrome (AIDS), characterized by opportunistic infections and opportunistic neoplasms, and caused by the retrovirus *Human immunodeficiency virus type 1* (*HIV-1*), is one of the leading causes of death worldwide. The AIDS epidemic has claimed more than 3.1 million lives in 2002, and an estimated 5 million people have acquired the HIV in 2002, bringing to 42 million people globally living with the virus.[1]

HIV-1 displays selective tropism for mature human helper T lymphocytes, which express the CD4 (or T4) surface protein.[2] Like other retroviruses, the HIV-1 particle is surrounded by a lipid bilayer derived from host-cell membranes.[3] The infectivity of HIV requires surface glycoprotein subunits (gp120) and transmembrane glycoprotein subunits (gp41) of gp160, a viral precursor protein. Each gp120 forms three-dimensional binding sites for the CD4 receptor and a G-protein-coupled chemokine receptor [either CC chemokine receptor 5 (CCR5) or CXC chemokine receptor 4 (CXCR4)].[4]

HIV entry is mediated by an interaction between CD4 and members of the chemokine receptors.[4,5] HIV-1 cell entry inhibitors include chemokine-receptor inhibitor, CD4-receptor inhibitor, membrane-fusion inhibitor and other attachment inhibitors. After fusion HIV releases copies of the RNA genome into the cytoplasm, and the viral reverse-transcriptase enzyme transcribes single-stranded viral RNA into double-stranded DNA that can be integrated into the genetic material of the human host.[3] Reverse transcriptase inhibitors were the first agents approved for the treatment of HIV-1. The viral integrase enzyme is required for the integration of proviral DNA into the host genome before replication. Integrase inhibitors are in clinical trials. When the infected cell synthesizes new protein, integrated proviral DNA is also translated into the protein building blocks of new viral progeny. The viral components then assemble on the cell surface and bud out as immature

* Corresponding author. Tel.: +93-33-2414-6676; fax: +91-33-2414-6677; e-mail: kunalroy_in@yahoo.com

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.11.009

viral particles. The final maturation of newly formed viruses requires the HIV-1 protease to make up an infectious virion. The inhibition of key enzymes, HIV-1 reverse transcriptase and HIV-1 protease, provides the most attractive target for the anti-HIV drug development.

The success of the currently available antiretroviral drugs is limited by the emergence of drug resistant viruses, the requirements of complex regimens and the toxic effects. This necessitates research on exploring novel classes of safe and effective agents with low risk of cross-resistance with other antiretroviral drugs.

Due to pressing necessity to develop potential anti-HIV drugs, Quantitative Structure-Activity relationship (QSAR) studies have been used on various anti-HIV drug candidates acting on different targets. Garg et al.[6] reported Hansch analysis of anti-HIV activity of hydroxyethoxymethylphenylthiothymines (HEPT), tetrahydroimidazobenzodiazepinones (TIBO), tertiary-butyldimethylsilylspiroaminooxathioledioxide pyrimidine nucleosides, dihydropyridodiazepinone (nevirapine) derivatives, pyridinones, α-anilinophenylacetamides, 2′,3′-dideoxynuceloside analogues, cyclic ureas and cycloalkylpyranones. The anti-HIV activity of HEPT derivatives has also been studied by other authors using different techniques like neural networks (NN),[7–11] multiple linear regression (MLR),[10–13] partial least square (PLS)[12] and Hologram QSAR (HQSAR).[14] 3-D-QSAR was performed on HEPT[15] analogues using Comparative molecular field analysis (CoMFA). The anti-HIV activity of TIBO derivatives has also been studied by HQSAR,[14] PLS[16] and by utilizing MLR techniques.[17] CoMFA has been applied in order to explain the structural requirements for HIV-1 reverse transcriptase (HIV-1 RT) inhibitory activity of TIBO derivatives.[18] QSAR was performed on thiazolothiazepines[19] using CoMFA and Comparative molecular similarity indices analysis (CoMSIA). CoMFA was also applied to nevirapine derivatives, which are active against wild-type (WT) and mutant-type (Y181C) HIV-1 reverse transcriptase.[20] CoMFA has been performed on diverse cyclic urea derivatives[21,22] and on porphyrin derivatives.[23] CoMFA, CoMSIA, and HQSAR analyses were carried out on 3-amino-2-hydroxyl-4-phenylbutanoic acids.[24] QSAR of diverse classes[25] of salicylpyrazolinones, dioxepinones, coumarins, quinones, and benzoic hydrazides compounds were studied using CoMSIA. The anti-HIV activity of tetrahydropyrimidine-2-ones was studied by CoMFA and by first order valence molecular connectivity index $^{1}\chi^{v}$.[26] QSAR of anti-HIV activity of flavonoids was done by using MLR techniques.[27] PLS-multivariate regression was used for QSAR study of flavones.[28] Genetic function approximation method was used to perform the QSAR of catechols and noncatechols for the anti-HIV activity.[29] QSAR has been performed concerning the anti-HIV activity of a series of d4T phosphoramidate derivatives by NN approach[30] and by Hansch type analysis.[31] Hansch and Free-Wilson analyses were performed on anti-HIV quinoline derivatives.[32] QSAR of anti-HIV activity of *bis*-tetraazamacrocyclic[33] compounds were analyzed using PLS analysis. Recently, advances in QSAR studies of HIV-1 RT inhibitors have been reviewed by Gupta.[34] Scozzafava et al. have recently reported new approaches for the design of antiviral drugs[35] and also reviewed the advances in the development of antiviral sulfonamides.[36]

Recently, we have modeled[37] anti-HIV-1 activity (assayed in MT-4 cell line) and RT binding affinity data of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners reported by Chan et al.[38] using quantum chemical and physicochemical parameters by multiple regression technique with principal component factor analysis as the data pre-processing step. In the present paper, we have modeled the same data set using E-state index and molecular connectivity parameters along with appropriate indicator variables.

2. Materials and methods

Structural specificity of drug molecules is exhibited at the atomic or fragmental level instead of the whole molecule. In the drug receptor interaction phenomenon, a portion of the molecule (pharmacophore) may play more important role than the other segments. Though basic information for constitution of the topological indices are derived from the atom level (count of atoms, bonds, paths of bonds, etc.), most of the indices are applied to the whole molecule after summing up all components over the whole molecule. Thus, QSAR studies at the atomic or fragmental level are justified in the present context.[39]

The electrotopological state atom (E-state) index developed by Hall and Kier[40] is an atom level descriptor encoding both the electronic character and topological environment of each skeletal atom in a molecule. It is derived from chemical graph theoretic approach and has two basic components: (1) intrinsic topological and electronic state of an atom; (2) effect of the environment influencing the atom, considering differences in the intrinsic topological states of different atoms and topological distance among them which determine the magnitude of the interactions. The formalism of E-state index considers that each atom in a molecule is present in an information field composed of the other atoms of the molecule. Every other atom has an effect on a specific atom depending on the difference in electronegativity or electron-richness between them and also on their relative distance. The electrotopological states of atoms have been shown to bear relationships with relative electron-richness or ionicity of different atoms, their topological states (i.e., relative exposed or buried nature), and relative positions among different substituents or groups.

The intrinsic value has been defined as the ratio of a measure of electronic state (Kier–Hall valence state electronegativity[41]) to the local connectedness. The count of valence electrons which are the most reactive and involved in chemical reactions and bond formations are considered in the expression of I to encode the elec-

tronic feature. To reflect differences in electronegativity among the atoms, principal quantum number is employed in the expression of I. The topological attribute is included by using adjacency count of atom. The intrinsic value of an atom i is defined as

$$I_i = \left[(2/N)^2\delta^v + 1\right]/\delta \quad (1)$$

In eq. (1), N stands for principal quantum number and δ^v and δ indicate the count of valence electrons and sigma electrons associated with the atom i in the hydrogen-suppressed graph. The intrinsic electrotopological state calculated according to eq. (1) produces different values of an atom in different degrees of unsaturation and in different degrees of substitutions (branching). The values are also different for different atoms having differences in electronegativity. The intrinsic values increase with increase in electronegativity or electron-richness and decrease with increase in branching (substitution).

The perturbation factor stands for the influence of the information field on the intrinsic state of an atom. Such interactions decrease with increase in relative distance between two atoms. The perturbation factor for the intrinsic state of atom i is defined as

$$\Delta I_i = \sum_{j\neq i} \frac{I_i - I_j}{r_{ij}^2} \quad (2)$$

In eq. (2), r_{ij} stands for the graph separation factor, i.e., count of skeletal atoms in the shortest path connecting the atoms i and j including both atoms.

Summation of intrinsic state of an atom and influence of the field is called electrotopological state of the atom.

$$S_i = I_i + \sum_{j\neq i} \Delta I_{ij} \quad (3)$$

It is a representation of molecular structure information as it varies with changes in structural features including branching, cyclicity, homologation, heteroatom variation, changes in relative positions of different groups.

The electrotopological state considers both bonded and non-bonded interactions: the bonded component depends simply on difference in electronegativity among the adjacent atoms. The non-bonded interactions may be through inductive effect across the skeleton and is a function of graph separation factor and electronegativity difference. Thus, electrotopological state represents electronic distribution information modified by both local and global topology. The information encoded in the E-state value for an atom is the electronic accessibility at that atom.

The E-state index has been projected as an useful tool in the context of QSAR studies and reported to have power to identify atoms or fragments in the molecules which are important for the biological activity.[42–44] In a recent paper, Rose and Hall[45] have commented that topological models directly give structural information to guide design of new molecules and the topological model developed by them in the paper was statistically better than a previous model[46] based on ab initio quantum mechanical calculations. The present group of authors also have used E-state index to explore QSAR of ligands acting on pharmacologically relevant targets of contemporary interest.[47–49] In continuation of such efforts, the present communication will show here the utility of E-state index in QSAR studies by exploring QSAR of anti-HIV-1 activity and RT binding affinity data (Table 1) of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners[38] using E-state index.

As we see from previous papers[45,50] that use of molecular connectivity along with E-state index helps to develop better models, we have also tried to include first order valence molecular connectivity ($^1\chi^v$) in the relations. The molecular connectivity parameters[51,52] are a group of descriptors developed by Kier and Hall, which have been extensively used in QSAR studies. The first order valence connectivity is defined as

$$^1\chi^v = \sum_{i<j}\left(\delta_i^v\delta_j^v\right)^{-0.5} \quad (4)$$

In eq. (4), δ_i^v stands for valence δ value of atom i as defined in Ref. 52.

From the previous Hansch analysis[37] on the present data set, we have seen that molar refractivity values of the *meta* substituents of the aryl ring contribute significantly to the response variables. Thus, in the present study, we have tried fragmental valence connectivity of the *meta* substituents as predictor variable. Additionally, indicator variables (defined in Table 2) as found important from previous analysis were also tried here to improve the topological models.

The Anti-HIV-1 activity (assayed in MT-4 cell line) and RT binding affinity data [IC_{50} (μm)] of the compounds[38] were converted to the logarithmic scale [pC (mM)] (Table 1) and then used for subsequent QSAR analyses as the response variable. All 68 compounds considered in the present study contain 16 common atoms (excluding hydrogens). The atoms of the molecules were numbered keeping serial numbers of the common atoms same in all the compounds (as shown in Fig. 1). The electrotopological states of the 16 common atoms of the 68 compounds were found out using a GW-BASIC program ELECTRO1 developed by one of the

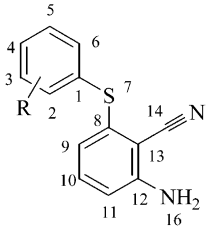
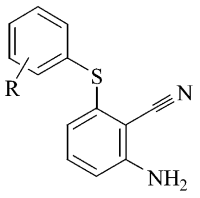
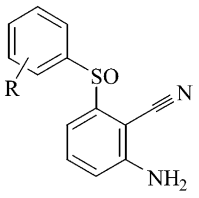


Figure 1. General structure 2-amino-6-arylthiobenzonitriles: the common atoms have been numbered **1**–**16**.

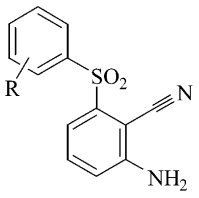
Table 1. Observed, calculated and predicted anti-HIV-1 activity and HIV-RT binding affinity of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners



1-23



24-38



39-68

Sl. No.	R	Anti HIV-1 activity[a]					HIV-1 RT binding affinity[a]				
		Obs.[b]	Calcd.[c]	Res.[c]	Pred.[c]	Pres.[c]	Obs.[b]	Calc.[d]	Res.[d]	Pred.[d]	Pres.[d]
1	H	1.836	1.782	0.054	1.772	0.064	2.061	1.695	0.366	1.664	0.397
2	2-OCH_3	2.367	2.171	0.196	2.062	0.305	2.569	1.695	0.874	1.622	0.947
3	3-OCH_3	2.222	2.058	0.164	2.039	0.183	2.824	2.459	0.365	2.431	0.393
4	2-CH_3	1.796	1.362	0.434	1.303	0.493	—	—	—	—	–
5	3-CH_3	2.215	2.124	0.091	2.116	0.099	3.018	2.348	0.670	2.309	0.709
6	4-CH_3	0.939	1.082	−0.143	1.097	−0.158	2.244	1.695	0.549	1.649	0.595
7	2-Cl	2.387	1.536	0.851	1.454	0.933	2.143	1.695	0.448	1.658	0.485
8	3-Cl	2.131	2.248	−0.117	2.259	−0.128	1.796	2.402	−0.606	2.443	−0.647
9	4-Cl	—	—	—	—	—	1.921	1.695	0.226	1.676	0.245
10	2-Br	1.523	1.424	0.099	1.413	0.110	—	—	—	—	–
11	3-Br	2.292	2.511	−0.219	2.548	−0.256	1.824	2.563	−0.739	2.662	−0.838
12	3-F	2.009	1.924	0.085	1.908	0.101	1.921	1.993	−0.072	1.997	−0.076
13	2-CN	—	—	—	—	—	2.041	1.695	0.346	1.666	0.375
14	3-CN	2.762	2.007	0.756	1.955	0.807	2.959	2.321	0.638	2.286	0.673
15	4-CN	1.359	0.804	0.555	0.650	0.709	—	—	—	—	–
16	3-CF_3	1.893	1.373	0.520	1.216	0.677	2.149	1.440	0.709	1.169	0.980
17	3-NH_2	1.502	1.943	−0.441	1.970	−0.468	—	—	—	—	–
18	2,5-Cl_2	—	—	—	—	—	2.456	2.402	0.054	2.399	0.057
19	3,5-$(CH_3)_2$	3.367	3.165	0.202	3.137	0.230	2.959	3.563	−0.604	3.624	−0.665
20	3,5-Cl_2	—	—	—	—	—	3.921	3.780	0.141	3.763	0.158
21	3-Cl,5-CH_3	2.754	3.289	−0.535	3.365	−0.611	2.770	3.617	−0.847	3.702	−0.932
22	3-OCH_3, 5-CH_3	2.699	3.099	−0.400	3.157	−0.458	3.854	3.674	0.180	3.656	0.198
23	3-OCH_3, 5-CF_3	2.292	2.612	−0.320	2.714	−0.422	1.886	3.022	−1.136	3.387	−1.501
24	2-OCH_3	2.319	2.633	−0.314	2.791	−0.472	1.921	1.695	0.226	1.676	0.245
25	3-OCH_3	1.796	2.517	−0.721	2.583	−0.787	1.721	2.459	−0.738	2.516	−0.795
26	2-CH_3	1.032	1.821	−0.789	1.915	−0.883	—	—	—	—	–
27	3-CH_3	1.534	2.581	−1.047	2.638	−1.104	2.000	2.348	−0.348	2.369	−0.369
28	4-CH_3	1.310	1.533	−0.223	1.548	−0.238	—	—	—	—	–
29	2-Br	1.407	1.883	−0.476	1.927	−0.520	—	—	—	—	–
30	3-Br	4.097	2.969	1.128	2.838	1.259	2.319	2.563	−0.244	2.595	−0.276
31	4-Br	1.694	1.481	0.213	1.462	0.232	—	—	—	—	–
32	2-CN	2.409	1.833	0.576	1.785	0.624	2.004	1.695	0.309	1.669	0.335
33	3-CN	1.848	2.466	−0.618	2.492	−0.644	—	—	—	—	–
34	3-CF_3	1.398	1.838	−0.440	1.957	−0.559	—	—	—	—	–
35	3,5-$(CH_3)_2$	3.469	3.626	−0.157	3.641	−0.172	3.301	3.563	−0.262	3.590	−0.289
36	2,5-Cl_2	2.007	2.465	−0.458	2.498	−0.491	2.208	2.402	−0.194	2.415	−0.207
37	3-Cl, 5-CH_3	3.495	3.750	−0.255	3.773	−0.278	3.284	3.617	−0.333	3.651	−0.367
38	3-OCH_3, 5-CF_3	2.684	3.081	−0.397	3.190	−0.506	3.046	3.022	0.024	3.014	0.032
39	H	2.699	2.704	−0.005	2.704	−0.005	2.161	2.599	−0.438	2.635	−0.474
40	2-OCH_3	3.222	3.104	0.118	3.039	0.183	2.854	2.599	0.255	2.578	0.276
41	3-OCH_3	3.046	2.987	0.059	2.981	0.065	3.222	3.363	−0.141	3.378	−0.156
42	4-OCH_3	1.602	1.666	−0.064	1.678	−0.076	1.886	2.599	−0.713	2.657	−0.771
43	2-CH_3	2.638	2.291	0.347	2.246	0.392	2.347	2.599	−0.252	2.619	−0.272
44	3-CH_3	3.398	3.050	0.348	3.025	0.373	3.699	3.252	0.447	3.217	0.482
45	4-CH_3	2.022	1.996	0.026	1.994	0.028	2.137	2.599	−0.462	2.637	−0.500
46	2-Cl	2.387	2.465	−0.078	2.472	−0.085	2.229	2.599	−0.370	2.629	−0.400
47	3-Cl	3.229	3.174	0.055	3.170	0.059	3.398	3.307	0.091	3.299	0.099
48	4-Cl	2.523	1.848	0.675	1.722	0.801	—	—	—	—	–
49	2-Br	2.301	2.353	−0.052	2.358	−0.057	1.921	2.599	−0.678	2.654	−0.733
50	3-Br	3.268	3.437	−0.169	3.461	−0.193	3.699	3.467	0.232	3.431	0.268
51	4-Br	1.699	1.943	−0.244	1.967	−0.268	—	—	—	—	–
52	2-F	2.523	2.787	−0.264	2.851	−0.328	2.301	2.599	−0.298	2.623	−0.322
53	3-F	2.523	2.850	−0.327	2.906	−0.383	—	—	—	—	–
54	2-CN	2.268	2.304	−0.036	2.307	−0.039	2.222	2.599	−0.377	2.630	−0.408
55	3-CN	2.620	2.936	−0.316	2.955	−0.335	2.745	3.226	−0.481	3.262	−0.517
56	4-CN	1.097	1.721	−0.624	1.858	−0.761	—	—	—	—	–

(*continued on next page*)

Table 1 (*continued*)

Sl. No.	R	Anti HIV-1 activity[a]					HIV-1 RT binding affinity[a]				
		Obs.[b]	Calcd.[c]	Res.[c]	Pred.[c]	Pres.[c]	Obs.[b]	Calc.[d]	Res.[d]	Pred.[d]	Pres.[d]
57	3-CF_3	2.456	2.309	0.147	2.263	0.193	2.276	2.344	−0.068	2.373	–0.097
58	2,5-Cl_2	3.523	2.936	0.587	2.885	0.638	3.523	3.307	0.216	3.288	0.235
59	3,5-Cl_2	4.155	4.458	−0.303	4.530	−0.375	4.523	4.684	−0.161	4.707	–0.184
60	3,5-$(CH_3)_2$	5.000	4.095	0.905	3.996	1.004	5.155	4.467	0.688	4.387	0.768
61	3-Br, 5-CH_3	4.699	4.483	0.216	4.458	0.241	5.523	4.682	0.841	4.572	0.951
62	3-Cl, 5-CH_3	4.523	4.220	0.303	4.186	0.337	5.301	4.522	0.779	4.431	0.870
63	3-OCH_3, 5–CH_3	4.301	4.033	0.268	4.002	0.299	5.000	4.578	0.422	4.529	0.471
64	3-OCH_3, 5-CF_3	4.046	3.556	0.490	3.409	0.637	4.398	3.926	0.472	3.763	0.635
65	3-OH, 5-CH_3	3.367	3.387	−0.020	3.388	−0.021	—	—	—	—	–
66	3-OCH_2CH_3, 5-CH_3	4.222	4.120	0.102	4.105	0.117	—	—	—	—	–
67	3-$O(CH_2)_2CH_3$, 5-CH_3	4.222	3.935	0.287	3.848	0.374	—	—	—	—	–
68	3-$O(CH_2)_2CH_3$, 5-CH_3	3.222	3.510	−0.288	4.440	−1.218	3.398	3.404	−0.006	3.554	–0.156

[a] Obs., observed; Calc., calculated; Res., residual = Obs.–Calc.; Pred., predicted; Pres., predicted residual = Obs.−Pred.
[b] Taken from Ref. 38.
[c] Calculated/Predicted from eq. (5).
[d] Calculated/Predicted from eq. (11).

Table 2. Definitions of indicator parameters

Parameter	Definition
I_p	Indicator variable having value 1 if *para* substituent is present, value 0 otherwise.
I	Indicator variable having value 1 for sullfonyl compounds, value 0 otherwise.
$I_{2m\text{-}Me}$	Indicator variable having value 1 if bi-*meta*-methyl substituents are present, value 0 otherwise.
$I_{o\text{-}OMe}$	Indicator variable having value 1 if *ortho*-methoxy substituent is present, value 0 otherwise.
$I_{m\text{-}CF3}$	Indicator variable having value 1 if *meta*-trifluoromethyl substituent is present, value 0 otherwise.

authors.[53] The program uses, as input, only the connection table in a specific format along with intrinsic state values of different atoms. To the output file thus obtained, the biological activity data were introduced to make it ready for subsequent regression analysis. Stepwise multiple regression technique was adopted for the development of the final topological models using E-state values of the common atoms, first order fragmental valence connectivity of the *meta* substituents and the indicator parameters. The regression analyses were carried out using a GW-BASIC program *RRR98*.[53] The statistical quality of the equations[54] was judged by the parameters like explained variance (R_a^2, i.e., adjusted R^2), correlation coefficient (r or R), standard error of estimate (s), average of absolute values of the residuals (*AVRES*), variance ratio (F) at specified degrees of freedom (df) and 95% confidence intervals of the regression coefficients. PRESS (leave-one-out) statistics[55,56] were calculated using the programs *KRPRES1* and *KRPRES2*,[53] and the reported parameters are cross-validation R^2 (Q^2), predicted residual sum of squares (*PRESS*), standard deviation based on PRESS (S_{PRESS}), standard deviation of error of prediction (*SDEP*) and average absolute predicted residual ($Pres_{av}$). All the accepted equations have regression coefficients and F ratios significant at 95% and 99% levels respectively, if not stated otherwise (marked with *). A compound was considered as an outlier if the residual is more than twice the standard error of estimate for a particular equation. Finally, 'leave–10%–out' was also applied on some selected equations to show robustness and predictive potential of the generated equations.

3. Results and discussion

Stepwise development of topological models for both anti-HIV-1 activity and HIV-1 RT binding affinity using fragmental molecular connectivity index of *meta* substituents ($[^1\chi^v]_m$), E-state parameters and indicator variables (I_p, $I_{2\text{-}m\text{-}Me}$, $I_{o\text{-}OMe}$ and $I_{m\text{-}CF3}$) has been shown in Table 3. The values of fragmental molecular connectivity parameters of *meta* substituents of the aryl ring and selected E-state parameters are shown in Table 4.

3.1. QSAR of anti-HIV-1 activity

Among the parameters stated above, $[^1\chi^v]_m$ emerged as the single best parameter explaining 39.9% of the variance of anti-HIV-1 activity ($r = 0.639$, $s = 0.755$). However, to explore contribution of each *meta* substituent separately, we have defined two terms $[^1\chi^v]_{ma}$ and $[^1\chi^v]_{mb}$. If fragmental $^1\chi^v$ values of the two *meta* substituents are different, then the higher value is called $[^1\chi^v]_{ma}$ and the other is $[^1\chi^v]_{mb}$. In case of same two *meta* substituents, $[^1\chi^v]_{ma}$ and $[^1\chi^v]_{mb}$ are same. $[^1\chi^v]_{mb}$ actually signifies the impact of $^1\chi^v$ values of the second *meta* substituents. On using $[^1\chi^v]_{ma}$ and $[^1\chi^v]_{mb}$ terms instead of $[^1\chi^v]_m$ term, significant increase in statistical quality occurs: the resultant equation shows explained variance of 44.9% ($R = 0.683$) and standard error of estimate drops to about 0.723. Among the rest of the descriptors, S_7 was found as the best additional descriptor, which could increase the value of explained variance to 57.0% ($R = 0.769$, $s = 0.638$). Interestingly, on use of square term of $[^1\chi^v]_{ma}$, the explained variance rises to 60.4%

Table 3. Step wise development of equations for modeling anti-HIV-1 data ($n=64$) and HIV-1 RT binding data ($n=51$)

Key activity	Combination of descriptors	Ref. eq. no.	Statistics			
			R_a^2	R	F (*df*)	s
Anti-HIV-1	$[^1\chi^v]_m$	—	0.399	0.639	42.8 (1,62)	0.755
	$[^1\chi^v]_{ma}$, $[^1\chi^v]_{mb}$	—	0.449	0.683	26.6 (2,61)	0.723
	S_7, $[^1\chi^v]_{ma}$, $[^1\chi^v]_{mb}$	—	0.570	0.769	28.9 (3,60)	0.638
	S_7, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.604	0.793	25.1 (4,59)	0.612
	S_c, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.594	0.787	24.1 (4,59)	0.620
	S_4, S_c, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.726	0.865	34.3 (5,58)	0.510
	S_4, S_c, $I_{m\text{-}CF3}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.739	0.874	30.7 (6,57)	0.498
	S_4, S_c, $I_{m\text{-}CF3}$, $I_{o\text{-}OMe}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.755	0.884	28.7 (7,56)	0.482
	S_4, S_c, S_{15}, $I_{m\text{-}CF3}$, $I_{o\text{-}OMe}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	(5)	0.775	0.897	28.2 (8,55)	0.461
HIV-1 RT binding affinity	$[^1\chi^v]_m$	—	0.316	0.574	24.0 (1,49)	0.827
	$[^1\chi^v]_{mb}$	—	0.411	0.650	35.9 (1,49)	0.787
	S_7, $[^1\chi^v]_{mb}$	—	0.513	0.730	27.3 (2,48)	0.698
	S_7, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.553	0.767	16.5 (4,46)	0.668
	S_7, $I_{m\text{-}CF3}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	—	0.656	0.831	20.1 (5,45)	0.586
	S_c, $I_{m\text{-}CF3}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	(10)	0.660	0.833	20.4 (5,45)	0.583
	I, $I_{m\text{-}CF3}$, $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$, $[^1\chi^v]_{mb}$	(11)	0.730	0.870	28.0 (5,45)	0.520

($R=0.793$, $s=0.612$). Again, as E-state values of atoms 7, 8, 9 and 13 are highly intercorrelated ($r^2>0.9$), a new term S_C was defined as sum of the E-state values of those atoms. On using S_C instead of S_7 along with $[^1\chi^v]_{ma}$, $[^1\chi^v]^2_{ma}$ and $[^1\chi^v]_{mb}$, an equation of comparable quality ($R_a^2=0.594$, $R=0.787$, s = 0.620) was obtained. The next best descriptor was S_4 and the resultant relation predicted 68.2% and explained 72.6% of the variance of anti-HIV-1 data. The standard error of estimate of the equation was 0.510 while standard error of prediction was 0.545. On using $I_{m\text{-}CF3}$ as the next additional descriptor, the predicted variance increased to 69.4% and standard error of prediction dropped to 0.535. When $I_{o\text{-}OMe}$ was used as the next descriptor, an equation with 71.4% predicted variance and 75.5% explained variance was obtained.

The final equation was obtained on using an additional descriptor, S_{15}.

$$\begin{aligned} pC_1 = {} & 0.573(\pm0.377)S_4 - 0.129(\pm0.043)S_C - 6.281 \\ & \times(\pm5.091)S_{15} + 0.992(\pm0.643)I_{o-OMe} \\ & - 0.921(\pm0.589)I_{m-CF_3} + 1.546(\pm0.698) \\ & \times\left[^1\chi^v\right]_{ma} + 2.680(\pm0.813)\left[^1\chi^v\right]_{mb} - 0.530 \\ & \times(\pm0.408)\left[^1\chi^v\right]^2_{ma} + 58.024(\pm46.265) \end{aligned} \quad (5)$$

$$\begin{aligned} & n = 64,\ Q^2 = 0.724,\ R_a^2 = 0.775,\ R^2 = 0.804, \\ & R = 0.897,\ F = 28.2(df\,8, 55),\ s = 0.461, \\ & AVRES = 0.339,\ SDEP = 0.508,\ S_{PRESS} = 0.548, \\ & PRESS = 16.5,\ Pres_{av} = 0.404 \end{aligned}$$

The 95% confidence intervals of the regression coefficients are given within parentheses. Eq. (5) shows 72.4% predicted variance and 77.5% explained variance. The positive coefficient of the variable S_4 indicates that the anti-HIV-1 activity increases with increase in the E-state value of atom 4 while the negative coefficients of S_C and S_{15} indicate that the activity decreases with increase in E-state values of atoms 7, 8, 9, 13 and 15. As S_4 is considerably correlated with I_p ($r^2=0.460$), it actually implies the impact of the *para* substituents. Presence of the *para* substituents causes a steric clash with adjacent residues in the binding site, particularly Pro95 that leads to displacement of the ligand away from Tyr181, interaction with which is a necessary requirement for the activity.[38] Again, S_{15} indicates the importance of the nitrile functionality for anti-HIV-1 activity. Chan et al.[38] observed van der Waals contact of the nitrile group of the ligand with the aliphatic portion of the side chain of Lys103. As E-state of an atom is considered as a measure of its electronic accessibility, negative coefficient of S_C indicates that the activity increases as the benzene ring becomes less electron-rich. This is also in agreement with that the activity increases in case of the sulfonyl congeners: presence of electron withdrawing sulfonyl moiety makes the benzene ring less electron-rich. The coefficients of $I_{o\text{-}OMe}$ and $I_{m\text{-}CF3}$ indicate that an *ortho*-methoxy group on the aryl ring will be favorable for the activity while a *meta*-trifluoromethyl group will be detrimental. The $^1\chi^v$ terms indicate the importance of the *meta* substituents: the $^1\chi^v$ value of the first *meta* substituent shows a parabolic relation while the impact of the second *meta* substituent is more as evidenced from the larger coefficient. This suggests that the size of the *meta* substituents, especially the second one (in a bi-*meta* substituted compound) is of critical importance for the anti-HIV-1 activity. All these observations are in accordance with the previous Hansch analysis[37] on the data set.

Eq. (5) has two outliers: compounds **27** and **30**. On deleting compound **30**, the following relation was obtained:

$$\begin{aligned} pC_1 = {} & 0.572(\pm0.356)S_4 - 0.132(\pm0.040)S_C - 5.656 \\ & \times(\pm4.841)S_{15} + 0.942(\pm0.610)I_{o-OMe} \\ & - 0.840(\pm0.561)I_{m-CF_3} + 1.354(\pm0.675) \\ & \times\left[^1\chi^v\right]_{ma} + 2.767(\pm0.772)\left[^1\chi^v\right]_{mb} - 0.492 \\ & \times(\pm0.387)\left[^1\chi^v\right]^2_{ma} + 52.376(\pm43.989) \end{aligned} \quad (6)$$

Table 4. Selected E-state parameters and fragmental molecular connectivity values

Compd. no.	S_4	S_7	S_8	S_9	S_{13}	S_{15}	$[^1\chi^v]_{ma}$[a]	$[^1\chi^v]_{mb}$[a]
1	1.967	1.540	0.905	1.908	0.559	9.033	—	–
2	1.910	1.466	0.843	1.880	0.515	9.128	—	–
3	1.882	1.489	0.861	1.887	0.526	9.109	0.651	–
4	2.003	1.566	0.912	1.921	0.563	9.102	—	–
5	2.049	1.554	0.909	1.917	0.562	9.088	0.500	–
6	1.232	1.549	0.907	1.915	0.561	9.077	—	–
7	1.871	1.434	0.828	1.863	0.505	9.070	—	–
8	1.815	1.470	0.850	1.874	0.519	9.062	0.567	–
9	0.704	1.490	0.864	1.882	0.528	9.056	—	–
10	1.956	1.519	0.882	1.900	0.543	9.091	—	–
11	1.966	1.524	0.888	1.902	0.546	9.079	0.982	–
12	1.382	1.314	0.742	1.795	0.439	9.014	0.189	–
13	1.812	1.368	0.772	1.826	0.461	9.095	—	–
14	1.744	1.418	0.807	1.845	0.484	9.082	0.470	–
15	0.626	1.450	0.830	1.858	0.500	9.072	—	–
16	1.007	1.070	0.542	1.651	0.280	9.037	0.817	–
17	1.827	1.474	0.853	1.876	0.521	9.063	0.290	–
18	1.719	1.364	0.773	1.829	0.464	9.098	0.567	–
19	2.132	1.568	0.913	1.926	0.564	9.142	0.500	0.500
20	1.663	1.399	0.796	1.840	0.478	9.090	0.567	0.567
21	1.897	1.483	0.854	1.883	0.521	9.116	0.567	0.500
22	1.965	1.503	0.865	1.896	0.528	9.164	0.651	0.500
23	0.923	1.019	0.498	1.630	0.247	9.113	0.817	0.651
24	1.795	−1.516	0.404	1.635	0.250	9.125	—	–
25	1.768	−1.471	0.421	1.642	0.261	9.106	0.651	–
26	1.889	−1.391	0.473	1.677	0.298	9.099	—	–
27	1.935	−1.390	0.469	1.673	0.296	9.084	0.500	–
28	1.108	−1.388	0.468	1.670	0.296	9.074	—	–
29	1.842	−1.438	0.443	1.656	0.277	9.087	—	–
30	1.851	−1.420	0.448	1.657	0.281	9.075	0.982	–
31	0.921	−1.409	0.452	1.658	0.284	9.066	—	–
32	1.698	−1.614	0.332	1.581	0.196	9.091	—	–
33	1.629	−1.543	0.367	1.600	0.219	9.078	0.470	–
34	0.893	−1.925	0.103	1.407	0.015	9.034	0.817	–
35	2.017	−1.400	0.474	1.682	0.299	9.139	0.500	0.500
36	1.605	−1.618	0.334	1.584	0.199	9.095	0.567	–
37	1.783	−1.485	0.415	1.639	0.256	9.113	0.567	0.500
38	0.809	−2.017	0.058	1.385	−0.018	9.110	0.817	0.651
39	1.725	−3.699	−0.052	1.384	−0.006	9.014	—	–
40	1.668	−3.867	−0.114	1.356	−0.050	9.109	—	–
41	1.641	−3.812	−0.096	1.363	−0.039	9.090	0.651	–
42	0.560	−3.780	−0.085	1.368	−0.032	9.076	—	–
43	1.762	−3.730	−0.045	1.397	−0.002	9.083	—	–
44	1.808	−3.722	−0.048	1.393	−0.003	9.069	0.500	–
45	0.972	−3.716	−0.050	1.391	−0.004	9.058	—	–
46	1.630	−3.862	−0.129	1.339	−0.060	9.050	—	–
47	1.573	−3.806	−0.107	1.350	−0.047	9.043	0.567	–
48	0.444	−3.775	−0.093	1.358	−0.037	9.037	—	–
49	1.715	−3.777	−0.075	1.377	−0.023	9.072	—	–
50	1.724	−3.752	−0.069	1.378	−0.019	9.059	0.982	–
51	0.784	−3.737	−0.065	1.379	−0.016	9.051	—	–
52	1.387	−4.105	−0.285	1.231	−0.169	8.990	—	–
53	1.141	−3.962	−0.215	1.271	−0.126	8.995	0.189	–
54	1.571	−3.965	−0.185	1.302	−0.104	9.076	—	–
55	1.502	−3.883	−0.150	1.321	−0.081	9.062	0.470	–
56	0.366	−3.834	−0.128	1.334	−0.066	9.053	—	–
57	0.766	−4.282	−0.415	1.127	−0.285	9.018	0.817	–
58	1.478	−3.969	−0.184	1.305	−0.101	9.079	0.567	–
59	1.421	−3.913	−0.161	1.316	−0.087	9.071	0.567	0.567
60	1.890	−3.744	−0.044	1.403	−0.001	9.123	0.500	0.500
61	1.807	−3.774	−0.065	1.387	−0.016	9.114	0.982	0.500
62	1.656	−3.829	−0.103	1.359	−0.044	9.097	0.567	0.500
63	1.724	−3.834	−0.092	1.372	−0.037	9.145	0.651	0.500
64	0.682	−4.394	−0.459	1.106	−0.318	9.094	0.817	0.651
65	1.724	−3.804	−0.086	1.372	−0.032	9.105	0.500	0.220
66	1.755	−3.843	−0.087	1.380	−0.033	9.183	1.200	0.500
67	1.775	−3.849	−0.084	1.386	−0.030	9.215	1.700	0.500
68	1.789	−3.855	−0.081	1.391	−0.028	9.242	2.200	0.500

[a] See text for definition.

$n = 63$, $Q^2 = 0.750$, $R_a^2 = 0.794$, $R^2 = 0.821$,
$R = 0.906$, $F = 30.9(df 8, 54)$, $s = 0.437$,
$AVRES = 0.326$, $SDEP = 0.477$, $S_{\mathrm{PRESS}} = 0.515$,
$PRESS = 14.3$, $\mathrm{Pre}s_{\mathrm{av}} = 0.388$

When another outlier (**27**) was deleted, the following relation was obtained:

$$\begin{aligned} pC_1 = {} & 0.605(\pm 0.342)S_4 - 0.134(\pm 0.039)S_C - 5.350 \\ & \times (\pm 4.644)S_{15} + 0.906(\pm 0.585)I_{\mathrm{o-OMe}} \\ & - 0.825(\pm 0.538)I_{\mathrm{m-CF_3}} + 1.431(\pm 0.650) \\ & \times \left[{}^1\chi^v\right]_{\mathrm{ma}} + 2.651(\pm 0.746)\left[{}^1\chi^v\right]_{\mathrm{mb}} - 0.539 \\ & \times (\pm 0.373)\left[{}^1\chi^v\right]_{\mathrm{ma}}^2 + 49.566(\pm 42.201) \end{aligned} \quad (7)$$

$n = 62$, $Q^2 = 0.770$, $R_{\mathrm{a}}^2 = 0.811$, $R^2 = 0.836$,
$R = 0.914$, $F = 33.7(df 8, 53)$, $s = 0.418$,
$AVRES = 0.314$, $SDEP = 0.457$, $S_{\mathrm{PRESS}} = 0.495$,
$PRESS = 13.0$, $\mathrm{Pre}s_{\mathrm{av}} = 0.374$

The calculated and predicted anti-HIV-1 activity values according to eq. (5) are shown in Table 1.

Again, indicator variable I (denoting presence or absence of the sulfonyl moiety) shows high correlation with S_C ($r^2 = 0.778$). On using I as predictor variable instead of S_C and omitting the insignificant terms, the following relation was obtained:

$$\begin{aligned} pC_1 = {} & 0.898(\pm 0.127)S_4 + 1.156(\pm 0.215)I \\ & + 0.642(\pm 0.567)I_{\mathrm{o-OMe}} + 1.515(\pm 0.600)\left[{}^1\chi^v\right]_{\mathrm{ma}} \\ & + 1.862(\pm 0.592)\left[{}^1\chi^v\right]_{\mathrm{mb}} - 0.751(\pm 0.338\left[{}^1\chi^v\right]_{\mathrm{ma}}^2 \end{aligned} \quad (8)$$

$n = 64$, $Q^2 = 0.761$, $R_{\mathrm{a}}^2 = 0.781$, $R^2 = 0.799$,
$R = 0.894$, $F = 382.9(df 6, 58)$, $s = 0.455$,
$AVRES = 0.311$, $SDEP = 0.473$, $S_{\mathrm{PRESS}} = 0.501$,
$PRESS = 14.3$, $\mathrm{Pre}s_{\mathrm{av}} = 0.342$

The intercept term of eq. (8) was insignificant and thus has been set to zero. The correlation coefficient ($r = 0.894$) of eq. (8) is slightly inferior to that ($r = 0.897$) of eq. (5) though the former shows higher predicted variance (Q^2 value 0.761 versus 0.724). When a significant outlier (**30**) for eq. (8) was deleted, the following relation was obtained:

$$\begin{aligned} pC_1 = {} & 0.681(\pm 0.228)S_4 + 1.109(\pm 0.206)I \\ & + 0.635(\pm 0.479)I_{\mathrm{o-OMe}} + 1.193(\pm 0.524)\left[{}^1\chi^v\right]_{\mathrm{ma}} \\ & + 2.102(\pm 0.510)\left[{}^1\chi^v\right]_{\mathrm{mb}} - 0.639(\pm 0.292)\left[{}^1\chi^v\right]_{\mathrm{ma}}^2 \\ & + 0.411(\pm 0.457) \end{aligned} \quad (9)$$

$n = 63, Q^2 = 0.823, R_a^2 = 0.840, R^2 = 0.856,$
$R = 0.925, F = 55.4(df6, 56), s = 0.385,$
$AVRES = 0.275, SDEP = 0.402, S_{PRESS} = 0.426,$
$PRESS = 10.2, Pres_{av} = 0.308$

The variable I in eqs. (8) and (9) suggest that the presence of sulfonyl moiety is conducive to the anti-HIV-1 activity. Intercorrelation (r^2) among important predictor variables for anti-HIV-1 data modeling is shown in Table 5.

3.2. QSAR of HIV-1 RT binding affinity

Among E-state parameters, fragmental molecular connectivity index of *meta* substituents ($[^1\chi^v]_m$), and indicator variables (I_p, $I_{2\text{-m-Me}}$, $I_{o\text{-OMe}}$ and $I_{m\text{-CF3}}$), $[^1\chi^v]_m$ emerged as the single best parameter explaining 31.6% of the variance of HIV-1 RT binding affinity ($r = 0.574$, $s = 0.827$). However, $[^1\chi^v]_{mb}$ was found to be better descriptor than $[^1\chi^v]_m$ as it could explain 41.1% of the variance ($r = 0.650$, $s = 0.787$). The best additional descriptor was S_7 which could increase explained variance to 51.3% ($r = 0.730$, $s = 0.698$). Again, on using $[^1\chi^v]_{ma}$ and $[^1\chi^v]^2_{ma}$ as additional descriptors, an equation with 55.3% explained variance ($r = 0.767$, $s = 0.668$) was obtained. When $I_{m\text{-CF3}}$ was used as an additional descriptor, the resultant equation could predict 59.9% and explain 65.6% of the variance (R = 0.831). The standard error of estimate and prediction were 0.586 and 0.627 respectively. As S_7 is highly intercorrelated with S_8, S_9 and S_{13} ($r^2 > 0.9$), a new variable S_C as defined previously was introduced and the following relation was obtained:

$$\begin{aligned} pC_2 = {} & -0.111(\pm0.050)S_C - 1.231(\pm0.604)I_{m-CF_3} \\ & + 1.615(\pm0.863)[^1\chi^v]_{ma} \\ & + 2.431(\pm0.769)[^1\chi^v]_{mb} - 0.796(\pm0.480)[^1\chi^v]^2_{ma} \\ & + 2.209(\pm0.490) \end{aligned} \quad (10)$$

$n = 51, Q^2 = 0.603, R_a^2 = 0.660, R^2 = 0.694,$
$R = 0.833, F = 20.4(df5, 45), s = 0.583,$
$AVRES = 0.470, SDEP = 0.624, S_{PRESS} = 0.664,$
$PRESS = 19.9, Pres_{av} = 0.535$

Table 5. Intercorrelation (r^2) matrix for E-state parameters, molecular connectivity parameters and indicator variables for anti-HIV-1 data modeling ($n = 64$)

	S_4	S_{15}	S_C	$[^1\chi^v]_{ma}$	$[^1\chi^v]_{mb}$	$[^1\chi^v]^2_{ma}$	$I_{m\text{-CF3}}$	$I_{o\text{-OMe}}$	I
S_4	1.000	0.164	0.112	0.024	0.005	0.012	0.241	0.016	0.062
S_{15}		1.000	0.006	0.304	0.384	0.373	0.017	0.030	0.005
S_c			1.000	0.020	0.016	0.032	0.004	0.002	0.778
$[^1\chi^v]_{ma}$				1.000	0.296	0.773	0.080	0.045	0.013
$[^1\chi^v]_{mb}$					1.000	0.181	0.069	0.017	0.013
$[^1\chi^v]^2_{ma}$						1.000	0.016	0.014	0.033
$I_{m\text{-CF3}}$							1.000	0.005	0.008
$I_{o\text{-OMe}}$								1.000	0.004
I									1.000

The negative coefficient of S_C in eq. (10) indicates that increase in E-state values of the atoms 7, 8, 9 and 13 decrease RT binding affinity. Considering E-state of an atom is as a measure of its electronic accessibility, the negative coefficient of S_C indicates that the activity increases as the benzene ring becomes less electron-rich, e.g., in case of the sulfonyl congeners. The negative coefficient of $I_{m\text{-CF3}}$ indicates the detrimental effect of *meta*-trifluoromethyl substituent on the aryl nucleus for the RT binding affinity. The first order valence molecular connectivity of the *meta* substituents show significant contributions: the value of the first *meta* substituent shows a parabolic relation while that for the second one (in a di-*meta* substituted compound) shows enhanced impact as evidenced from larger regression coefficient.

Again, the variable I is highly correlated with S_C ($r^2 = 0.817$). Hence, I was used instead of S_C and the following relation with enhanced statistics was obtained:

$$\begin{aligned} pC_2 = {} & 0.904(\pm0.300)I - 1.095(\pm0.536)I_{m-CF_3} \\ & + 1.747(\pm0.772)[^1\chi^v]_{ma} + 2.430(\pm0.685)[^1\chi^v]_{mb} \\ & - 0.879(\pm0.430)[^1\chi^v]^2_{ma} + 1.695(\pm0.458) \end{aligned} \quad (11)$$

$n = 51, Q^2 = 0.685, R_a^2 = 0.730, R^2 = 0.757,$
$R = 0.870, F = 28.0(df5, 45), s = 0.520,$
$AVRES = 0.415, SDEP = 0.556, S_{PRESS} = 0.591,$
$PRESS = 15.7, Pres_{av} = 0.467$

Eq. (11) could predict and explain 68.5% and 73.0% respectively of the variance of data set. On deleting an outlier (**23**), the following relation was obtained:

$$\begin{aligned} pC_2 = {} & 0.853(\pm0.284)I - 0.819(\pm0.544)I_{m-CF_3} \\ & + 1.669(\pm0.727)[^1\chi^v]_{ma} + 2.601(\pm0.655)[^1\chi^v]_{mb} \\ & - 0.852(\pm0.404)[^1\chi^v]^2_{ma} + 1.719(\pm0.425) \end{aligned} \quad (12)$$

$n = 50, Q^2 = 0.731, R_a^2 = 0.763, R^2 = 0.787,$
$R = 0.887, F = 32.5(df5, 44), s = 0.487,$
$AVRES = 0.396, SDEP = 0.514, S_{PRESS} = 0.548,$
$PRESS = 13.2, Pres_{av} = 0.452$

The calculated and predicted HIV-1 RT binding affinity according to eq. (11) are shown in Table 1. The variable I in eqs. (11) and (12) suggests the importance of the presence of sulfonyl moiety for the binding affinity. Intercorrelation (r^2) among important predictor variables for HIV-1 RT binding affinity data modeling is shown in Table 6.

3.3. Overview of QSAR

The above equations show that for both the response variables, first order valence molecular connectivity of the *meta* substituents of the aryl ring plays a significant

Table 6. Intercorrelation (r^2) matrix for E-state parameters, molecular connectivity parameters and indicator variables for HIV-1 RT data modeling ($n = 51$)

	S_7	S_C	$[^1\chi^v]_{ma}$	$[^1\chi^v]_{mb}$	$[^1\chi^v]^2_{ma}$	$I_{m\text{-}CF3}$	I
S_7	1.000	0.996	0.006	0.003	0.019	0.001	0.812
S_C		1.000	0.007	0.004	0.020	0.008	0.817
$[^1\chi^v]_{ma}$			1.000	0.209	0.733	0.083	0.001
$[^1\chi^v]_{mb}$				1.000	0.101	0.096	0.000
$[^1\chi^v]^2_{ma}$					1.000	0.019	0.016
$I_{m\text{-}CF3}$						1.000	0.001
I							1.000

Table 7. Results of leave–10%–out cross-validation applied on Eqs. (5) and (11) Model equation, $pC = \Sigma\beta_i x_i + \alpha$

Eq. no.	Number of cycles	Average regression coefficients (standard deviations)	Statistics Q^2 (Average *Pres*)
(5)	11[a]	0.575 (0.080) S_4 −0.129 (0.007) S_C −6.336 (0.705) S_{15} + 0.993 (0.093) $I_{o\text{-}OMe}$ −0.922 (0.111) $I_{m\text{-}CF3}$ + 1.531 (0.163) $[^1\chi^v]_{ma}$ + 2.686 (0.129) $[^1\chi^v]_{mb}$ −0.509 (0.115) $[^1\chi^v]^2_{ma}$ + 58.522 (6.406)	0.727 (0.404)
(11)	10[b]	0.902 (0.066) I −1.107 (0.148) $I_{m\text{-}CF3}$ + 1.746 (0.122) $[^1\chi^v]_{ma}$ + 2.448 (0.145) $[^1\chi^v]_{mb}$ −0.887 (0.075) $[^1\chi^v]^2_{ma}$ + 1.695 (0.045)	0.679 (0.482)

Q^2 denotes cross-validated R^2. Average *Pres* means average of absolute values of predicted residuals.

[a] Compounds were deleted in 11 cycles in the following manner: each compound in each cycle is followed by 11th next compound.

[b] Compounds were deleted in 10 cycles in the following manner: each compound in each cycle is followed by 10th next compound.

role: second *meta* substituents show supraadditive effect on the activities probably due to enhanced binding (presumably through dispersion interaction) of the ligand with the binding site. This is further corroborated by the X-ray crystal structures of the complex of HIV-1 RT with non-nucleoside inhibitors,[38] which clearly shows that the binding site can favorably accommodate the *meta* substituents. Again, presence of sulfone moiety contributes significantly to the activities, which is in accordance with the report[38] that sulfonyl oxygens maintain Tyr181 side chain of the binding site of RT in the proper position for optimum interaction with the ligand. The sulfonyl group also makes the benzene ring less electron-rich, which may be a requirement for optimum activity. Again, presence of *meta*-trifluoromethyl group at the aryl ring is detrimental for both the activity parameters. Additionally, the anti-HIV-1 model shows negative contribution of the E-state values of atom 15 (nitrile nitrogen) and positive contributions of E-state value of atom 4 and presence of an *ortho*-methoxy group at the aryl nucleus.

The final equations are of acceptable statistical quality and predictive potential as evidenced from 'leave–10%–out' cross-validation applied on eqs. (5) and (11), results of which are summarized in Table 7. The final topological models generated in this study are comparable to the best LFER models based on the physicochemical parameters,[37] and are better than the best models based on electrostatic potential point charges as reported previously.[37] This suggests that E-state index along with molecular connectivity provides an useful tool for modeling studies with direct physicochemical significance.

Acknowledgements

A financial grant from J.U. Research Fund is thankfully acknowledged. One of the authors (JTL) thanks AICTE, New Delhi, India for QIP fellowship.

References and notes

1. http://www.avert.org/worldstats.htm.
2. Klatzmann, D.; Champagne, E.; Chamaret, S.; Gruest, J.; Guetard, D.; Hercend, T.; Gluckman, J. C.; Montagnier, L. *Nature* **1984**, *312*, 767.
3. Kilby, J. M.; Eron, J. J. *N. Eng. J. Med.* **2003**, *348*, 2228.
4. Clapham, P. R.; McKnight, A. *J. Gen. Virol.* **2002**, *83*, 1809.
5. Hoxie, J. A.; LaBranche, C. C.; Endres, M. J.; Turner, J. D.; Berson, J. F.; Doms, R. W.; Matthews, T. J. *J. Reprod. Immunol.* **1998**, *41*, 197.
6. Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K.; Hansch, C. *Chem. Rev.* **1999**, *99*, 3525.
7. Douali, L.; Villemin, D.; Cherqaoui, D. *Curr. Pharm. Des.* **2003**, *9*, 1817.
8. Mager, P. P. *Curr. Med. Chem.* **2003**, *10*, 1643.
9. Douali, L.; Villemin, D.; Cherqaoui, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1200.
10. Bazoui, H.; Zahouily, M.; Boulajaaj, S.; Sebti, S.; Zakarya, D. *SAR. QSAR. Environ. Res.* **2002**, *13*, 567.
11. Jalali-Heravi, M.; Parastar, F. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 147.
12. Luco, J. M.; Ferretti, F. H. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 392.
13. Hannongbua, S.; Lawtrakul, L.; Limtrakul, J. *J. Comput. Aided Mol. Des.* **1996**, *10*, 145.
14. Pungpo, P.; Hannongbua, S.; Wolschann, P. *Curr. Med. Chem.* **2003**, *10*, 1661.
15. Hannongbua, S.; Nivesanond, K.; Lawtrakul, L.; Pungpo, P.; Wolschann, P. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 848.
16. Xie, H. P.; Jiang, J. H.; Cui, H.; Shen, G. L.; Yu, R. Q. *Comput. Chem.* **2002**, *26*, 591.
17. Huuskonen, J. J. *Chem. Inf. Comput. Sci.* **2001**, *41*, 425.
18. Hannongbua, S.; Pungpo, P.; Limtrakul, J.; Wolschann, P. *J. Comput. Aided Mol. Des* **1999**, *13*, 563.
19. Makhija, M. T.; Kulkarni, V. M. *J. Comput. Aided Mol. Des.* **2002**, *16*, 181.
20. Pungpo, P.; Hannongbua, S. *J. Mol. Graph. Model* **2000**, *18*, 581.
21. Debnath, A. K. *J. Med. Chem.* **1999**, *42*, 249.
22. Debnath, A. K. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 761.
23. Debnath, A. K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P.; Neurath, A. R. *J. Med. Chem.* **1994**, *37*, 1099.

24. Huang, X.; Xu, L.; Luo, X.; Fan, K.; Ji, R.; Pei, G.; Chen, K.; Jiang, H. *J. Med. Chem.* **2002**, *45*, 333.
25. Makhija, M. T.; Kulkarni, V. M. *J. Comput. Aided Mol. Des.* **2001**, *15*, 961.
26. Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S. P. *Bioorg. Med. Chem.* **2001**, *9*, 3059.
27. Olivero-Verbel, J.; Pacheco-Londono, L. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1241.
28. Buolamwini, J. K.; Raghavan, K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. *Pharm. Res.* **1996**, *13*, 1892.
29. Makhija, M. T.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2002**, *10*, 1483.
30. Knaggs, M. H.; McGuigan, C.; Harris, S. A.; Heshmati, P.; Cahard, D.; Gilbert, I. H.; Balzarini, J. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2075.
31. Siddiqui, A. Q.; McGuigan, C.; Ballatore, C.; Zuccotto, F.; Gilbert, I. H.; De Clercq, E.; Balzarini, J. *J. Med. Chem.* **1999**, *42*, 4122.
32. Font, M.; Monge, A.; Ruiz, I.; Heras, B. *Drug Des. Discov.* **1997**, *14*, 259.
33. Joao, H. C.; De Vreese, K.; Pauwels, R.; De Clercq, E.; Henson, G. W.; Bridger, G. J. *J. Med. Chem.* **1995**, *38*, 3865.
34. Gupta, S. P. *Prog. Drug Res.* **2002**, *58*, 223.
35. Scozzafava, A.; Casini, A.; Supuran, C. T. *Curr. Med. Chem.* **2002**, *9*, 1167.
36. Scozzafava, A.; Owa, T.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Med. Chem.* **2003**, *10*, 925.
37. Leonard, J. T., Roy, K. *QSAR Comb. Sci.* **2003**, in press.
38. Chan, J. H.; Hong, J. S.; Hunter, R. N., III; Orr, G. F.; Cowan, J. L.; Sherman, D. L.; Sparks, S. M.; Reitter, B. E.; Andrews, C. W., III; Hazen, R. J.; St Clair, M.; Boone, L. R.; Ferris, R. G.; Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Ott, R. J.; Ren, J.; Hopkins, A.; Stuart, D. I.; Stammers, D. K. *J. Med. Chem.* **2001**, *44*, 1866.
39. Hall, L. H.; Mohney, B.; Kier, L. B. *Quant. Struct-Act. Relat.* **1991**, *10*, 43.
40. Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
41. Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1981**, *70*, 583.
42. Hall, L. H.; Mohney, B.; Kier, L. B. *Quant. Struct-Act. Relat.* **1993**, *12*, 44.
43. Kier, L. B.; and Hall, L. H. In *Advances in Drug Research*; Testa, B., Ed.; Academic Press: New York, 1992; Vol. 22, p 1.
44. Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: San Diego, 1999.
45. Rose, K.; Hall, L. H. *SAR QSAR Environ. Res.* **2003**, *14*, 113.
46. Robert, D.; Carbo-Dorca, R. *SAR QSAR Environ. Res.* **1999**, *10*, 401.
47. Roy, K.; Pal, D. K.; Sengupta, C. *Drug Des. Discov.* **2001**, *17*, 207.
48. Roy, K.; De, A. U.; Sengupta, C. *Drug Des. Discov.* **2002**, *18*, 33.
49. Roy, K.; Chakraborty, S.; Saha, A. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3753.
50. Maw, H. H.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 290.
51. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
52. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, 1986.
53. The GW-BASIC programs *ELECTRO1*, *RRR98, KRPRES1* and *KRPRES2* were developed by Kunal Roy and standardized on known data sets.
54. Snedecor, G. W.; Cochran, W. G. In *Statistical Methods*; Oxford and IBH Publishing Co. Pvt. Ltd: New Delhi, 1967; p 381.
55. Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; pp 312.
56. Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghose, A. K.; Viswanadhan, V. N., Eds.; Marcel Dekker, Inc: New York, 2001; p 73.