

Molecular modeling of viral nucleocapsid protein Zn fingers modulators

Shweta Sharma¹, Sarvesh Paliwal¹, Smrita Singh² & Mymoon Akhter*³

¹Department of Pharmacy, Banasthali University, Banasthali-304 022, Rajasthan, India

²Bioinformatics Infrastructure Facility; ³Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Jamia Hamdard University-110 062, New Delhi, India

Received 10 April 2015; revised 23 May 2015

The relationship between the biological activity and the physicochemical properties of agents is well documented. Using QSAR methodology, three different methods *viz.* multiple linear regression, partial least square and neural regression analysis were conducted on a series of 71 (sulfonyl) benzene analogs as viral nucleocapsid protein zinc finger modulators for HIV. Neural regression analysis was found to relate the two best. The best nonlinear regression model showed good correlative and predictive ability, r^2 (training) =0.97 and r^2 (test) =0.892 values. The results obtained from this study indicate the importance of Verloop, Ipso atom E-state index, electrostatic parameters such as dipole moment and electronic parameters such as VAMP polarization and Vamp dipole in determining the activity of viral nucleocapsid protein zinc finger inhibitors.

Keywords: Ipso atom E-state index, QSAR, Verloop, Viral Nucleocapsid Protein Zinc Finger Inhibitors

Acquired immunodeficiency syndrome (AIDS), caused by human immunodeficiency virus (HIV) infection, is still one of the most important challenges for the chemotherapy of the early 21st century¹. The status of the HIV/AIDS epidemic has evolved with time and continued to grow in 2010, reaching an estimate of 33.4 million population^{2,3}. Current AIDS therapies involve use of inhibitors of reverse transcriptase (RT) and protease (PR), against human immunodeficiency virus (HIV) when used in double or triple therapies which leads to long-term decrease in viral load⁴. Virus survival of Highly Active Anti Retroviral Therapy (HAART) is due to (i) outgrowth of viral strains with lower sensitivity to inhibitors, (ii) latency in the form of integrated and nonintegrated DNA in lymphocytes and monocytes, (iii) very high levels of replication in lymphoid tissues and (iv) replication in some organs have reduced accessibility to inhibitors⁵⁻⁹. At the molecular level, replication features of HIV with high levels of mutation and recombination probably account for the rapid selection and propagation of inhibitor-resistant viral

species¹⁰. There are various targets available like reverse transcriptase, integrase, Chemokine receptors (CK) *etc* with antiretroviral agents but these are unable to fulfill the therapeutic requirement. Thus, there is a high need for HIV potential target and novel inhibitors.

In the context of potential target and inhibitor design, Nucleocapsid7 (NCp7) is targeted as it contains one or two copies of a Cys (X)2 Cys (X)4 His (X)4 Cys (CCHC) Zn finger motif. Structural studies indicate that the bound zinc ion serves to stabilize the three-dimensional fold of the ZBD by participating in the formation of local secondary structure elements^{11,12}. Zinc finger inhibitors block replication of virus by disrupting the shape of nucleocapsid (NC). It also has mutationally non permissive nature. NCp7 mainly controls (a) Reverse Transcription and Integration, both these processes requires not only reverse transcriptase but also the (NC) protein which functions as a nucleic acid chaperone by which it facilitates the rearrangement of nucleic acids into conformations that are thermodynamically more stable than the original structure, (b) Packaging of RNA into assembling virus particles relying on the ability of the NC domain of the Gag precursor to bind specific regions of the RNA, forming a nucleation site for the multimerization of Gag upon the RNA.

*Correspondence:

Fax: +9101126059663

Email: mymoonakhter@gmail.com

Abbreviations: FFNN; Feed Forward Neural Network, HIV; Human Immunodeficiency Virus, MLR; Multiple Linear Regression, PLS; Partial Least Square.

Till date, a large number of antiretroviral agents have been developed and many of them faced the limitation of resistance (due to mutation of HIV virus). Therefore it makes a promising target for HIV^{13,14}. The purpose of this study is to find the potential descriptors modulating the activity of viral nucleocapsid Zn finger protein which would retrieve important structural information about the target for new improved drug candidates.

Materials and methods

Data set generation

The EC50 values for 71 (sulfonyl) benzene analogs were obtained from published literature¹⁴ as described in Table 1. The effective concentration of inhibitory activity of compounds in the series are expressed as the negative logarithm of EC50 in order to reduce the skewness of data set, where EC50 refers to experimentally determined molar concentration of the compound required to inhibit 50% of HIV-1 activity.

Molecular Representation

Chemical structures of all the compounds were sketched with the help of Accelrys (Discovery Studio version 2.0)¹⁵ and were imported into the work sheet of TSAR 3.3 software as .mol files. The series had two major substitutions, which were defined using “define substituent” in the Tsar worksheet’s tool bar to study the impact of varying substitutions. Three-dimensional structures of all imported structures were generated and optimized.

Descriptor Calculation

At first, more than 200 structural, geometrical, electronic and hydrophobic descriptors were calculated for both whole molecule and substituent. The descriptors with the same values for all the compounds were discarded. The correlation matrix was generated to study the data patterns and to reduce data redundancy. The correlation terms involved in correlation matrix indicates extend of co-linearity. This process was repeated for each and every set of two parameters and final descriptors that highly correlated to biological activity were retained Table 2.

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis

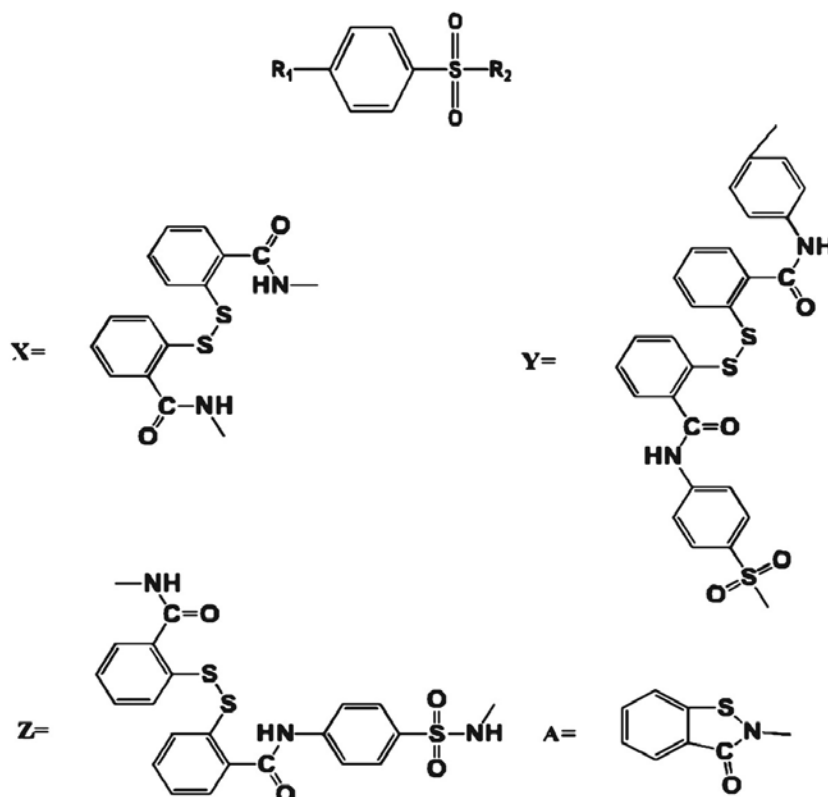
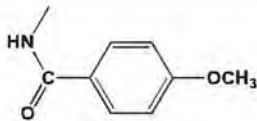

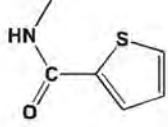
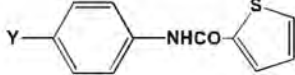
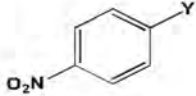
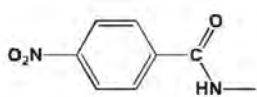
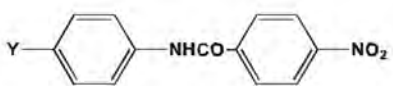
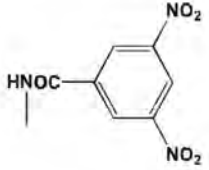
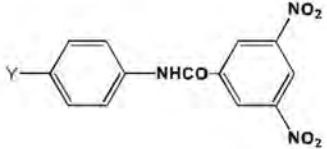
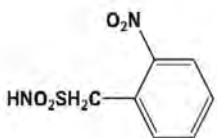
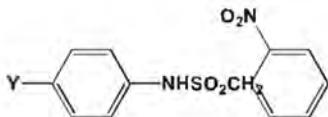
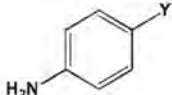
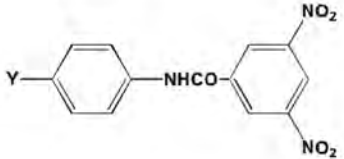
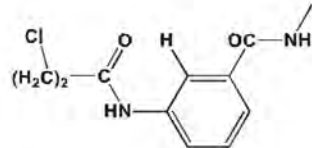
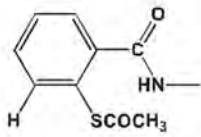


Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

COMP	R ₁	R ₂	EC ₅₀
1			0.33
2			0.41
3			1.6
4			1.9
5			2.5
6			111
7			2.3
8			4.3
9			1.5
10			1.0

(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

11			3.8
12			12.2
13	—NO_2		12.2
14			78
15			240
16			12.3
17	—NH_2		1
18	—NHCOCH_3		0.62
19		—NHCOCH_3	38
20		—NHCOCH_3	2.8

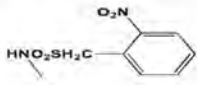
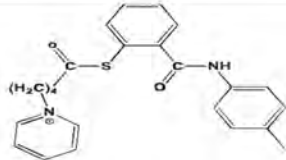
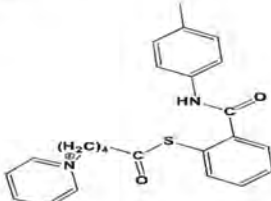
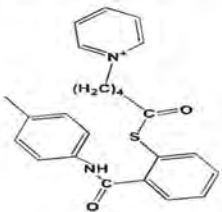
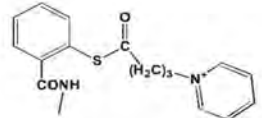
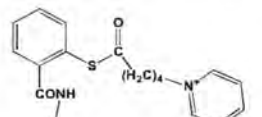
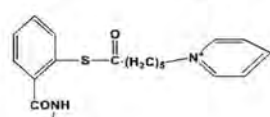
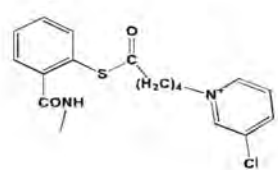
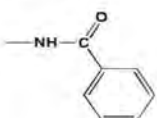
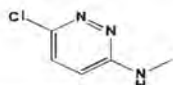
(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

21		—NHCOCH ₃	2.6
22		—NHCOCH ₃	2.8
23		—NHCOCH ₃	4.2
24		—NHCOCH ₃	3.8
25		—NHCOCH ₃	6.2
26		—NHCOCH ₃	3.8
27		—NHCOCH ₃	6.2
28			1.6
29			5.5
30		—NH ₂	1.1

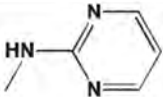
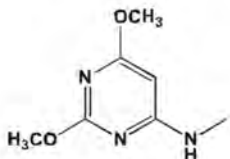
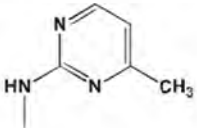
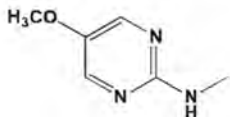
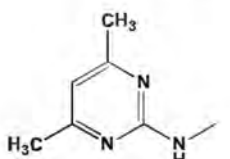
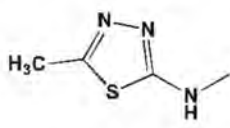
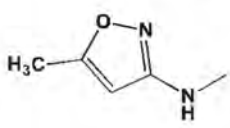
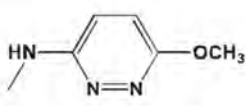
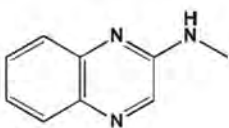
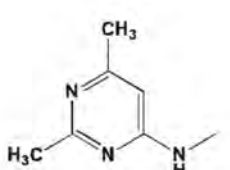
(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

31			2.9
32	—NHCOCH_3		4.6
33	—NHCOCH_3		4.9
34		—NHCOCH_3	10.1
35		—NHCOCH_3	6.2
36		—NHCOCH_3	10.2
37		—NHCOCH_3	54
38	A	—NHCOCH_3	12.6
39	A		9.3
40	A		7.1

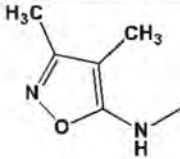
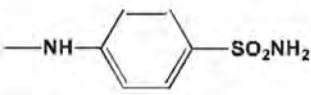
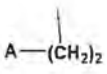




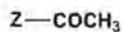
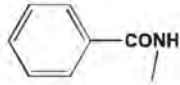
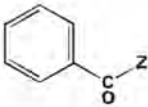
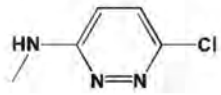
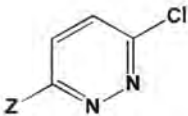
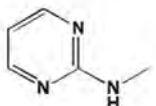
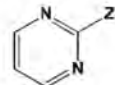
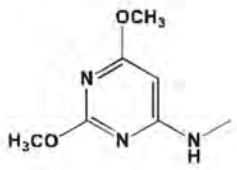
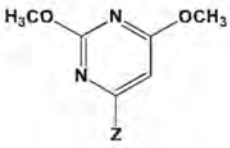
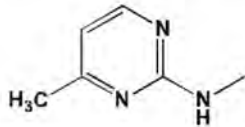
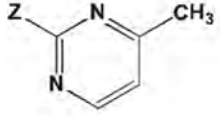
(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (Contd.)

41	A		6.5
42	A		7.6
43	A		8.8
44	A		17.8
45	A		6.9
46	A		7.8
47	A		8.7
48	A		22.6
49	A		9.2
50	A		6.5

(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

51	A		12.8
52	A		6.2
53			2
			
54			0.85
			
55			1.5
			
56			3.8
57			2.9
58			2
59			2.1
60			4.2

(Contd.)

Table 1—Structure and biological activity data of (sulfonyl) benzene derivatives used in QSAR analysis (*Contd.*)

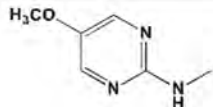
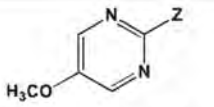
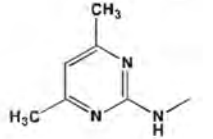
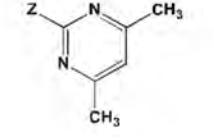
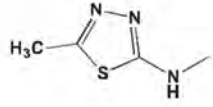
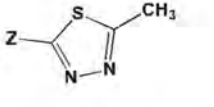
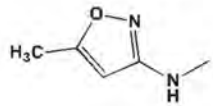
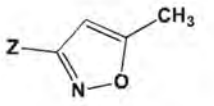
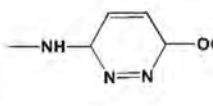
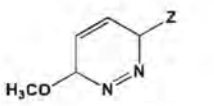
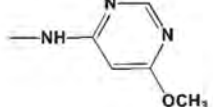
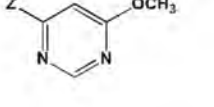
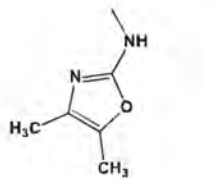
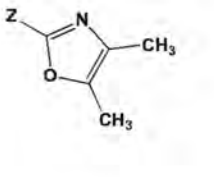
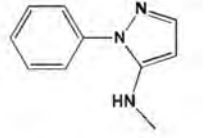
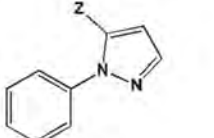
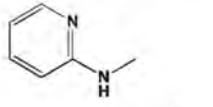
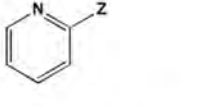
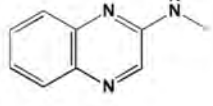
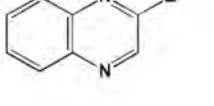
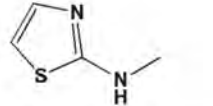
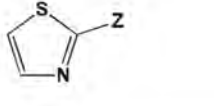
61			2.1
62			1.6
63			3.4
64			1.9
65			4.2
66			1.6
67			2.4
68			0.33
69			1.1
70			0.72
71			1

Table 2—Statistical tests and their values obtained after performing MLR and PLS analysis

Statistical tests	MLR	PLS
S value	0.251	0.878
F value	31.114	
F probability	8.706e-015	
Regression coefficient, r	0.905	
r ² (training)	0.819	0.819
r ² (test)	0.711	0.703
Cross Validation, r ² (CV)	0.678	0.698
Residual Sum of Squares	2.584	10.199
Predictive Sum of Squares	4.613	14.171
E-static		0.525
Fraction of variance explained		0.783

Statistical analysis

The compounds were divided into training set and test set. The main objective is to generate two sets with similar molecular diversity in order to be reciprocally representative and to cover all the main structural and physicochemical characteristics of the global data set¹⁶. Out of total 71 compounds, 49 compounds were taken in training set and 16 were included in test set. The reduced data set of five descriptors were subjected to regression analysis (statistical technique that identifies the relationship between two or more quantitative variables) technique *i.e.*, MLR and PLS (linear) and neural (non linear) analysis to establish correlation between most statistically significant descriptors and biological activity. Multiple linear regression (MLR) analysis was used for modeling quantitative relationships between a y-variable (dependent variable) and a block of x-variables (independent variables). The main restriction of MLR analysis is the case of large descriptors to compounds ratio or multicollinear descriptors in general. This makes the problem ill-trained and makes the results unstable¹⁷.

PLS is a (linear) method suitable for overcoming the problems in MLR related to multicollinear or over-abundant descriptors¹⁸⁻²⁰ and recommended as an alternative approach to enlarge the information contained in each model and avoids the danger of over fitting^{21,22}. The models generated (both MLR & PLS) were cross validated using leave out one row cross validation where one compound is removed from the dataset while the rest of the data comprises

the training set to estimate the coefficients of the QSAR model²³. The activity of the test compound is then predicted using the model based on the training set compounds. In addition to internal validation, the developed models were also validated using external set of components (test set). The drawback of this method is that the time to perform the calculation increases with the square of the size of the training set²³. Non Linear Regression analysis (neural) models extend the structure-activity relationships to non-linear functions of input descriptors. Such models may become more accurate, especially for large and diverse datasets. However, usually, they are harder to interpret.

Outlier (often encountered) is any observation in a dataset that is inconsistent with the remainder of the observations in that data set. The outlier is inconsistent in the sense that it is not indicative of possible future behavior of data sets coming from the same source^{24,25}.

Results and Discussion

The data set of 71 compounds was subjected to analysis by three different QSAR methods successfully. Initially, the data set containing 71 compounds with r² value of 0.561 and r² (CV) value of 0.344 was improved with the help of applicability domain software by identifying two outliers (compound number 6 and 38) which were detected and removed for improvement of statistical data of model from the graph. After deleting the two outlier compound 6 and 38 the value of r² (CV) improved to 0.409 and 0.462 respectively. Remaining compounds were divided into two sets, training set with 49 compounds and test set with 16 compounds. Method of regression model showed very poor value of both r² = 0.799 and r² (CV) = 0.545, so it envisaged the need for further refinement of model (with improved statistical quality of model).

Consequently, model generated was improved by discarding 4 outliers using applicability domain Software graph (technique used to validate the outliers deleted from data set and according to OECD principles it is restricted for all models²⁶). In this series total six outliers compound no. (6, 38, 59, 31, 16 and 33) were encountered which were deleted to improve the quality of the model. Outliers with high leverage and standard value were screened between standard and leverage. Threshold leverage = 3(k + 1)/n = 3(6+1)/49 = 0.428

Where k-number of descriptors left n- Number of compounds in set

The model developed after deletion of these potential outliers satisfied all the statistical criteria of a robust model. So the best model generated using MLR analysis for this data set resulted in r^2 value of 0.819 and r^2 (CV) of 0.678 values given in Table 2. Value of r^2 closer to 1.0 indicates good correlation while the value of r^2 (CV) above 0.6 indicates the good internal predictive capability of the developed model. Moreover the small difference between r^2 and r^2 (CV) values also augments the high prognostic quality of the model. Many other statistical tests were also performed on the training set to assure that the model formed is sound. For example, high value of F-test (31.114) reveals that the model constructed is statistically significant. The standard deviation (SD) of estimate in the model measures the accuracy of predictions made with regression. It shows how far the activity values are spread about their average. Its lower value (0.251) indicates soundness of the model. The efficacy of the model was also evaluated by checking its statistical stability using predictive and residual sum of squares.

The developed model is represented by equation 1 . Equation for MLR Analysis

$$Y = -0.147 \times X1 - 0.0253 \times X2 + 6.532 \times X3 + 0.153 \times X4 + 0.057 \times X5 - 0.0597 \times X6 + 0.225 \dots (1)$$

Equation for PLS Analysis

$$Y = -0.143 \times X1 - 0.0257 \times X2 + 6.460 \times X3 + 0.155 \times X4 + 0.057 \times X5 - 0.061 \times X6 + 0.232 \dots (2)$$

Where, X1 = Verloop B3 (Subst. 2), X2 = Dipole Moment X Component (Subst. 1), X3 = Kier ChiV6 (ring) index, X4 = Ipso Atom E-State, X5 = VAMP

Polarization XY Component (Whole molecule), X6=VAMP Dipole Y Component (Whole molecule).

PLS analysis (dimension 2) was also performed on the same data set for validation of the MLR model. The resulted r^2 (CV) value of 0.8198 clearly demonstrates the high predictive ability of the developed PLS model (Eq. 2). The other statistical test values of this model are given in Table 2. The best non-linear regression (neural) model generated showed good correlative and predictive ability with statistical significance, $r^2=0.974$ value with exclusion value 1 and hidden nodes 30. Both MLR and PLS, generated comparable results, with r^2 (CV) values of 0.8199 and 0.8198 respectively. The graphs showing MLR, PLS and Neural analysis for test and training is shown in Fig. 1-3. The actual and predicted biological activity values of MLR and PLS analysis for training and test are given in Table 3. Plot Dependency graphs interpret that the independent variables are showing relationship with dependent variables.

Analysis of Descriptors Entered

Descriptors entered the model are VerloopB3 (Subst. 2), Dipole Moment X component (Subst.1), Kier ChiV6 (ring) index, Ipso atom E-state index, VAMP Dipole (Whole Molecule), and VAMP polarization (Whole Molecule).

The Verloop B1-B5 parameters describe the width of the substituent in the direction perpendicular to the length of the substituent²⁷. As Verloop B3 is negatively correlated so, decrease in value of verloop results in increase in biological activity as confirmed by FFNN plot dependency, for example, substituting benzyl group with methyl results in increase in activity. Dipole Moment X component (Subst.1) is an electrostatic descriptor which explains the charge

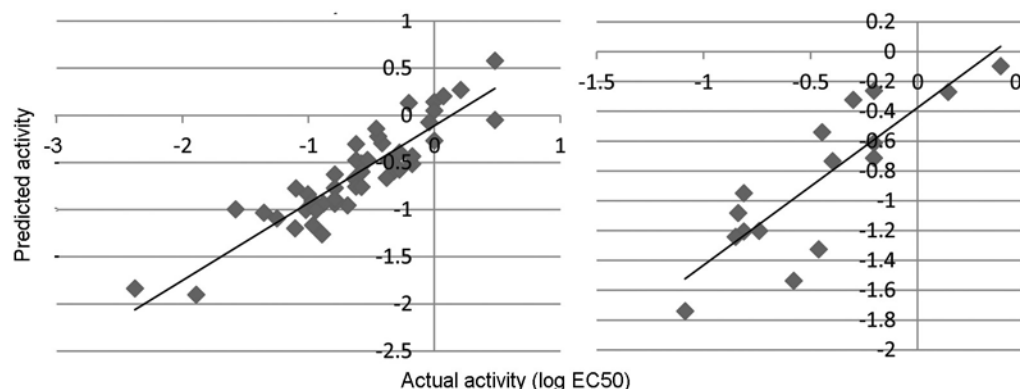


Fig. 1 — Plot of Actual vs. Predicted EC50 values for training and test set (MLR analysis)

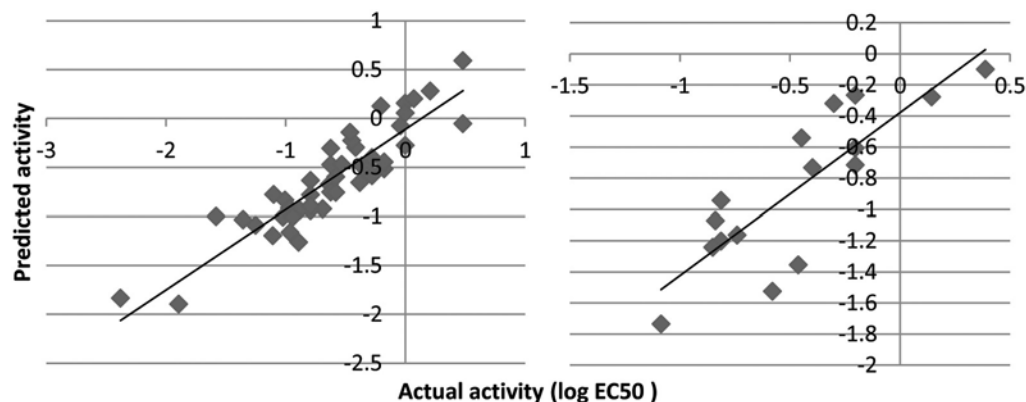


Fig. 2 — Plot of Actual vs. Predicted EC50 values training and test set (PLS analysis)

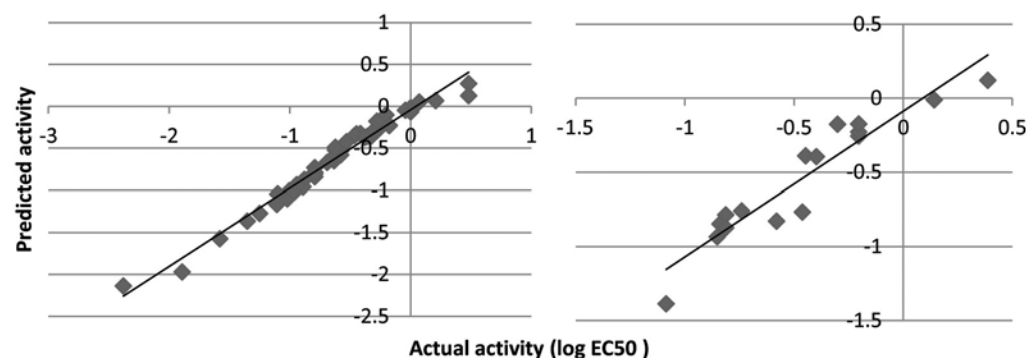


Fig. 3 — Plot of Actual vs. Predicted EC50 values for training and test set (Feed Forward neural network analysis)

distribution in a molecule provides valuable information on structure and polarity of molecule. As in the regression equation dipole moment is negatively correlated with the biological activity, this indicates that decreasing the polarity of the molecule or lead compound by substituting less/non polar groups for example alkyl group substituted by alcoholic group decrease the polarity of the molecule as a whole will account for an increase in the biological activity²⁸ as confirmed by FFNN plot dependency. Kier ChiV6 (ring) indices are the measure of the atomic arrangement in the compound which explained hydrogen bond interaction with receptor. It encodes information about the degree of branching of the atom within the molecule. The result show that when the kier chi V6 value is positively correlated as confirmed by FFNN plot dependency hence increasing its value, biological activity also increases and *vice versa*. Another descriptor that entered the model was Ipso atom E-state index. The E-state variable encodes the intrinsic electronic state of the atom as influenced by the electronic environment of all other atoms within the topological framework of the molecule²⁸. Ipso atom is a C-atom at

the juncture of two aromatic rings as in the 9 and 10 positions in naphthalene and an aromatic carbon which is bonded to a substituent. So ipso atom E-state index gives three very important aspects of structure information *i.e.* the electron accessibility associated with ipso atom type, characteristic of the E-state index, an indication of the presence of ipso atom type, and the count of the number of atoms of an ipso atom²⁹

An ipso atom E-state index is positively correlated with biological activity in the regression confirmed by FFNN plot dependency indicating that increase in Ipso atom index value will definitely lead to an increase in biological activity of lead molecule. VAMP is a molecular orbital package in TSAR 3.3, calculate the electrostatic properties such as total energy, electronic energy, nuclear repulsion energy, HOMO and LUMO ionization potential of molecules. Vamp polarization XY is positively correlated to the biological activity as shown by the regression equation and confirmed by FFNN plot dependency. VAMP Dipole Y Component (Whole molecule) is negatively correlated in regression equation further confirmed by FFNN plot dependency. Therefore,

Table 3 — Showing the actual and predicted values of biological activity obtained after MLR, PLS and FFNN analysis of Training and Test set

Compound	Actual	Predicted Activity		
		MLR	PLS	FFNN
ame	activity			
1	0.481	0.579	0.563	0.272
3	-0.204	0.131	0.225	-0.100
4	-0.278	-0.394	-0.338	-0.291
7	-0.361	-0.484	-0.440	-0.361
8	-0.633	-0.671	-0.585	-0.648
9	-0.176	-0.514	-0.483	-0.225
10	0	0.143	0.356	-0.017
14	-1.892	-1.902	-1.900	-1.970
15	-2.380	-1.834	-1.869	-2.136
17	0	0.0499	0.1611	-0.044
18	0.207	0.270	0.521	0.068
19	-1.579	-0.996	-1.032	-1.571
22	-0.414	-0.295	-0.279	-0.324
23	-0.447	-0.223	-0.190	-0.328
24	-0.623	-0.754	-0.730	-0.578
25	-0.579	-0.599	-0.617	-0.508
26	-0.792	-0.627	-0.673	-0.729
27	-0.579	-0.758	-0.789	-0.534
28	-0.792	-0.772	-0.950	-0.785
34	-0.690	-0.953	-1.023	-0.667
35	-1.004	-0.837	-0.704	-1.007
36	-0.792	-0.936	-0.867	-0.839
37	-1.021	-1.001	-0.825	-1.097
39	-1.100	-0.775	-0.871	-1.041
40	-0.968	-0.908	-1.028	-0.984
43	-0.880	-0.935	-1.018	-0.873
44	-0.944	-1.000	-1.065	-0.933
45	-1.250	-1.093	-1.198	-1.271
47	-0.892	-1.262	-1.336	-0.957
48	-0.939	-1.00	-1.102	-0.937
49	-1.354	-1.032	-1.165	-1.361
50	-0.963	-1.164	-1.255	-1.018
52	-1.107	-1.198	-1.396	-1.165
53	-0.792	-0.890	-0.937	-0.804
55	0.070	0.202	0.235	0.050
56	-0.176	-0.433	-0.343	-0.228
57	-0.579	-0.509	-0.451	-0.579
58	-0.462	-0.143	-0.066	-0.348
60	-0.322	-0.516	-0.508	-0.351
61	-0.623	-0.477	-0.453	-0.522
62	-0.322	-0.592	-0.539	-0.314
64	-0.531	-0.474	-0.446	-0.423
65	-0.278	-0.577	-0.604	-0.178
66	-0.623	-0.305	-0.253	-0.487
68	-0.380	-0.662	-0.627	-0.369
69	0.481	-0.048	-0.0001	0.126
70	-0.041	-0.074	-0.051	-0.047

(Contd.)

Table 3 — Showing the actual and predicted values of biological activity obtained after MLR, PLS and FFNN analysis of Training and Test set (*Contd.*)

Compound	Actual	Predicted Activity		
		MLR	PLS	FFNN
ame	activity			
72	0	-0.269	-0.194	-0.065
*2	0.387	0.0676	0.204	0.120
*11	-0.579	-0.821	-0.662	-0.826
*29	-0.204	-0.387	-0.372	-0.255
*30	-0.740	-0.990	-0.823	-0.758
*32	-0.462	-0.359	-0.553	-0.765
*42	-0.812	-0.654	-0.748	-0.784
*46	-0.838	-0.821	-0.793	-0.846
*51	-0.812	-0.811	-0.981	-0.871
*54	-0.301	-0.357	-0.258	-0.175
*63	-0.204	-0.251	-0.166	-0.227
*67	-0.204	-0.011	0.127	-0.174
*71	0.142	0.035	0.235	-0.010
*12	-1.086	-0.552	-1.068	-1.386
*21	-0.447	-0.388	-0.355	-0.389
*41	-0.851	-0.763	0.893	-0.933
*5	-0.397	-0.250	-0.419	-0.395

there will be an increase in the biological activity as the value of Vamp parameters increased.

Conclusion

In the current scenario of the drug design program, in the synthesis of potent and safer drug candidates against HIV, necessary care has to be taken in addressing the emergence of optimized drug candidates using molecular modeling approaches. In order to understand various physicochemical and structural properties of (sulfonyl) benzene derivatives for receptor binding, we reported herein our extensive 2D-QSAR investigation. In the present study, we evaluated a series of (sulfonyl) benzene analogs in order to determine the better structural characteristics that were able to improve the anti-HIV-1 activity of viral nucleocapsid protein zinc finger inhibitors. A comparison was made between the models derived from PLS, MLR and nonlinear (neural) regression analysis using conventional QSAR descriptors on the 49 (sulfonyl) benzene analogs in the training and 16 analogs in the test as anti-HIV agents. The validation procedures employed in this work illustrates the accuracy and robustness of the generated QSAR model. All the result discussed indicates that by using MLR, PLS and neural analysis with Verloop, Ipso atom E-state index, electrostatic parameters like dipole moment and VAMP Polarization and Vamp Dipole (electronic parameters) a very robust model has been derived that also possesses very powerful

predictive ability. Thus, the model reported in the present study will be helpful in development of new compounds with improved efficacy.

Acknowledgment

The authors sincerely thank Banasthali University, India for facilities and financial support provided.

References

- 1 Camarasa MJ, Velazquez S, San-Felix A, Perez-Perez MJ & Gago F, Dimerization of inhibitors of HIV-1 reverse transcriptase, protease & integrase: a single mode of inhibition for the three References HIV enzymes? *Antiviral Res*, 71 (2006) 260.
- 2 Fader LD, Bethell R, Bonneau P, Bös M, Bousquet Y, Cordingley MG, Coulombe R, Deroy P, Faucher AM, Gagnon A, Goudreau N, Grand-Maitre C, Guse I, Hucke O, Kawai SH, Lacoste JE, Landry S, Lemke CT, Malenfant E, Mason S, Morin S, O'Meara J, Simoneau B, Titolo S & Yoakim C, Discovery of a 1,5-dihydrobenzo[b][1,4]diazepine-2,4-dione series of inhibitors of HIV-1 capsid assembly. *Bioorg Med Chem Lett*, 21 (2011) 398.
- 3 *AIDS Epidemic Update*: UNAIDS/WHO. <http://www.unaids.org> (November, 2009).
- 4 Günthard HF, Wong JK, Ignacio CC, Guatelli JC, Riggs NL, Havlir DV & Richman DD, Human immunodeficiency virus replication and genotypic resistance in blood and lymph nodes after a year of potent antiretroviral therapy. *J Virol*, 72 (1998) 2422.
- 5 Korber BT, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R & Wolinsky SM, Genetic differences between blood & brain derived viral sequence from human immune deficiency virus type-1 infected patients: evidence of conserved elements in the V region of the envelope protein of brain derived sequences. *J. Virol*, 68 (1994) 7467.

- 6 Larder BA, Viral resistance and the selection of antiretroviral combinations. *J AIDS Human Retroviruses*, 10 (1995) 28.
- 7 Mercure L, Phaneuf D & Wainerg MA, Detection of unintegrated human immunodeficiency virus type-1 DNA in persistently infected CD8⁺ cells. *J Gen Virol*, 74 (1993) 2077.
- 8 Pantaleo G, Graziosi C, Demarest JF, Butini L, Montroni M, Fox CH, Oreinstein JM, Kotler DP & Fauci A, HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature*, 362 (1993) 355.
- 9 Zhang YM, Imamichi H, Imamichi T, Clifford Lane H, Falloon J, Vasudevachari MB & Salzman NP, Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol*, 71 (1997) 6662.
- 10 Finzi D & Siliciano RF, Viral dynamics in HIV-1 infection. *Cell Biol*, 93 (1998) 665.
- 11 Berg JM, Potential metal binding domains in nucleic acid binding protein. *Science*, 232 (1986) 485.
- 12 Covey SN, Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res*, 14 (1986) 623.
- 13 Negroni M & Buc H, Mechanisms of retroviral recombination. *Annu Rev Genet*, 35 (2001) 275.
- 14 Turpin JA, Song Y, Inman JK, Huang M, Wallqvist A, Maynard A, Covell DG, Rice WG & Appella E, Synthesis and biological properties of novel pyridinoalkanoil thiolesters (PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J Med Chem*, 12 (1999) 67.
- 15 <http://accelrys.com/products/accord/desktop/tsar.html>: TSAR 3.3, Oxford Molecular Limited
- 16 Rameshwar N, Krishna K, Ashok KB & Parthasarathy T, QSAR studies of N1-(5-chloro-2- pyridyl)-2-[[4-(alkyl methyl)] amino]-5-chlorobenzamide analoges. *Bioorg Med Chem*, 14 (2006) 319.
- 17 Benfenati E, *Quantitative structure activity relationships [QSAR] for pesticide regulatory purposes*. 1st edn (Elsevier, Philadelphia) 2007, 186.
- 18 Kim KH, Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? *J Computer Aided Mol Des*, 21 (2007) 421.
- 19 Wold S, Ruhe A, Wold H & Dunn WJ, The Collinearity Problem in Linear regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J Sci Stat Comput*, (5) (1984) 735.
- 20 Wold S, Sjostrom M & Eriksson L, Personal Memories of the early PLS development. *Chemom Intell Lab Syst*, 58 (2001) 109.
- 21 Phatak A & de Jong S, The geometry of partial least squares. *J Chemom*, 11 (1997) 311.
- 22 Kubinyi H, Evolutionary variable selection in regression and PLS analysis. *J Chemometr*, 10 (1996) 119.
- 23 Paliwal S, Narayan A & Paliwal S, Quantitative Structure Activity Relationship Analysis of Dicationic Diphenylisoxazole as Potent Anti-Trypanosomal Agents. *QSAR Comb Sci*, 28 (2009) 1367.
- 24 Dudek AZ, Arodzb T & Gálvezc J, Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Comb Chem High Throughput Screen*, 9 (2006) 213.
- 25 Balakrishnan N, Childs A, *Encyclopaedia of Mathematics*, (Kluwer Academic Publishers, ISBN 978-1556080104) 2001.
- 26 OECD, The report from the expert group on (Quantitative) Structure-Activity Relationships [(Q) SARs] on the principles for the validation of (Q) SARs. *J Series and Testing Assessment*, 49 (2004) 206.
- 27 Paliwal S, Seth D, Yadav D, Yadav R. & Paliwal S, *In silico* structure based drug design approach to develop novel pharmacophore model of human peroxisome proliferator-activated receptor γ agonist. *J Enzym Inhib Med Chem*, 26 (2011) 129.
- 28 Sapre NS, Pancholi N, Gupta S & Sapre N, Computational modeling of tetrahydroimidazo-[4,5,1-jk] [1,4] benzodiazepine derivatives An atomistic drug design approach using kier-hall electrotopological state (E-state) indices. *J Comput Chem*, 29 (2008) 1699.
- 29 Karleson M, *Wiley. Interscience*, New York February 2000.