Similarity Analysis studies on (Sulfonyl) Benzene Derivatives as Anti HIV Agents

Shweta Sharma[1], Smrita Singh[2], Mymoona Akhter[3,*], Sarvesh Paliwal[4]

[1,3]Dept. of Pharmacy, Banasthali University, Tonk, Rajasthan, [2]Dept. of Bioinformatics Infrastructure Facility, [3]Dept. of Pharmaceutical Chemistry, Faculty of Pharmacy, Jamia Hamdard, New Delhi

***Corresponding Author:**
Email: mymoonaakhter@gmail.com

Abstract

Introduction: The present investigation was undertaken to understand the overlaying problems of mutation in HIV virus leading to failure to combat AIDS.

Materials and Methods: *In present study* Multiple Linear Regression (MLR) analysis was carried on a series of 71(sulfonyl) benzene analogs reported as viral nucleocapsid protein zinc finger modulators for HIV.

Results: The MLR model obtained using carbo method (N*N) similarity showing good predictive ability, r^2 (training) = 0.655, r^2(test) = 0.605.

Conclusion: The results indicated that Refractivity Similarity (polarizability and volume) and Shape Similarity are important parameters in predicting the activity of viral nucleocapsid protein zinc finger inhibitors.

Keywords: Refractivity Similarity, Shape Simliarity, Carbo, MLR, Viral Nucleocapsid Protein Zinc Finger, HIV.

Access this article online

Website:
www.innovativepublication.com

DOI:
10.5958/2394-2797.2016.00025.3

Introduction

Acquired immunodeficiency syndrome (AIDS), one of the most important challenges for the chemotherapy of the early 21st century[1]. Although, therapeutic interventions such as Highly Active Antireteroviral Therapy is existing but constant challenge is being featured due to the mutation of the virus. Current AIDS therapies are not sufficient in overcoming the disease so there is a high call for new improved drug candidates.

The global scenario of AIDS is alarming and number of infected patients is regularly increasing and continued to grow in 2013, reaching an estimate of 34 million population[2,3]. The rapid mutation of the virus and hybridization of various subtypes, AIDS with TB is another important issue blocks the global efforts for remedy. *Pneumocystis jirovecii* (formerly known as *Pneumocystis carinii*) has been one of the hallmarks of late-stage HIV disease but is now less common because of ART and primary prophylaxis[4].

Protein Zinc finger, a validated target provide evidence to overcome problem of resistance due to non-permissive nature of mutation. It inhibits mainly by disrupting the shape of the Nucleocapsid (NC) which enables maturation of virus thereby controlling these process i.e. reverse transcription, integration and packaging in viral replication cycle.

Molecular similarity was introduced as a concept by Carbo et al.[5] but its use as a 3D QSAR tool was introduced by Good et al. and has been widely used by several other groups since then[6]. Similarity indices represent a quantitative measure of the similarity between two molecules on the basis of their size, shape, electronic distribution, lipid solubility, water solubility, or chemical reactivity[7]. The most widely used form of similarity index applied for calculation of 3D molecular similarity was proposed by Carbo:

$$R_{AB} = \int P_A.P_B\, dv / (\int P^2_A dv)^{1/2} (\int P^2_B\, dv)^{1/2} \quad (1)$$

The numerator in the equation 1 measures property overlap while denominator normalizes similarity result. The difference between equations for Carbo and Petke (Hodgkin) index is only in the denominator part with respect to Carbo index, it is less sensitive to shape of the property but more sensitive to its magnitude[8,9] Richards and Hodgkin 1988]. It is defined as:

$$H_{st} = \frac{2.\sum_{k=1}^{N} Psk.\ Ptk}{.\sum_{k=1}^{N} P2\ sk + \sum_{k=1}^{N} P2tk} = 1 - \frac{d2}{\sum_{k=1}^{N} P2\ sk + \sum_{k=1}^{N} P2tk} \quad (2)$$

Mathematical graphs and matrices represent, characterize and analyze biological sequences, chemical structure features and gene expression analysis. Discovery of useful knowledge from databases,[10] Insights data mining that opens new door for optimization of compounds in terms of discovering new therapeutic compound for HIV. Data mining in 2D QSAR involves collection, selection, transformation, visualization and evaluation of the extracted knowledge (Descriptors). There are number of algorithms based on the nature of the data as well as the desired knowledge that try to fit a model to the data[11].

In this perspective Similarity analysis of a series of compounds was carried out to find the possible descriptors which unfolds important structural information about the viral nucleocapsid Zn finger

protein target for new optimized drug candidates from known inhibitors.

Material and Method

Data Set Generation: The EC_{50} values for 71 reported compounds obtained from published literature[12] is given in (Table 1). Chemical structures of all the compounds were modeled, optimized and imported into the work sheet of TSAR 3.3 software of the compounds were divided into training set and test set with the objective to generate two sets with similar molecular diversity in order to be reciprocally representative and to cover all the main structural characteristics of the global data set[13]. Out of total 71 compounds, 44 compounds were taken in training set and 19 in test set. The molecular similarity indices were computed with the ASP similarity program in the TSAR software. Among the two approaches available (a grid-based method and Gaussian approximation), the Gaussian approximation was used for calculating similarity indices because it closely mirrors that of the grid-based calculations but is much faster. A Gaussian approximation based N x N similarity matrix was constructed and subjected to data reduction techniques.

Descriptor Calculation: At first, more than 300 descriptor were calculated for series for by Carbo index (NxN) similarity matrix. 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 by Pairwise correlation.

Statistical analysis

The reduced data set was subjected to regression analysis i.e. Multiple linear regression (MLR) analysis for modeling quantitative relationships between a y-variable (dependent variable) and a block of x-variables (independent variables).

Results and Discussion

Similarity analysis with dataset of 71 compounds were primary screening found to be with r2 value of 0.67 and $r^2_{(CV)}$ value of 0.634. For improvement of model, with the help of Applicability domain software two outliers (compound number 6 and 11) were detected and removed for improvement of statistical data of model.

Value of $r^2_{(CV)}$ after deleting outlier

Compound Number	Cross validation $r^2_{(CV)}$ value
6	Cross Validation, $r^2_{(CV)}$: 0.3009
11	Cross Validation, $r^2_{(CV)}$:0.3125

Remaining compounds were divided into two sets, training set with 44 compounds and test set with 19 compounds. This model showed very poor value of both $r^2 = 0.4567$ and $r^2_{(CV)} = 0.4589$, so it envisaged the need for further refinement of model (with improved statistical quality of model). Consequently, model generated was improved by discarding 6 outliers on the basis of high leverage and standard value using Applicability Domain Software. In this series total eight outliers were encountered which were compound no. (2, 6, 11, 12, 13, 19, 38 and 39). Figure 1 and Table II showing actual vs. predicted EC_{50} for training and tests set and statistical data respectively.

Threshold leverage=3(k+1)/n=3(2+1)/44=0.2045.

Where, k-number of descriptors left

n- Number of compounds in set

Y= 28.085*X1 + 1.347*X2-28.182 (Equation 3)

Analysis of Descriptors Entered

The results of refractivity similarity and shape similarity are presented in Table III. Refractivity similarity gives information about steric properties of a molecule and is dependent on both the volume and polarizability of the molecule. Refractivity similarity of molecule no 55 with respect to other compounds showed better correlation (more than -0.5 is desirable) with biological activity. As indicated by the equation (3) refractivity similarity is positively correlated so increase in volume shall result in increase in biological activity for example 2,2'-disulfanediylbis(N-methylbenzamide) linkage directly attached to benzene ring viz. (compound no. 1, 18, 54, 70) showed better activity as compared to compounds with same linkage attached through sulphonyl group. Shape similarity describes the quantitation of molecular shape in contrast to the absolute quantitation of size. Shape Similarity vs Molecule 68 showed that molecule 68 values permit a rational prediction of which molecules have high degree of shape similarity for the biological activity. As shape similarity is positively correlated (indicated by equation (3)) so increasing its value resulting in increased biological activity. It is also an important criteria in predicting potent compounds as confirmed by EC_{50} value of the reported compounds in the series.

Conclusion

In present study using (N x N) Matrix using Carbo index linear model has been generated for 71 reported compounds. The validation procedures employed in this work (LOO and activity prediction using test set of compounds) illustrates the accuracy and robustness of the generated QSAR model. The result discussed indicates that, with the Anti-HIV activity can be effectively modeled using Refractivity similarity and Shape similarity. Thus, the model reported in the present study would be helpful in development of new compounds with improved efficacy.

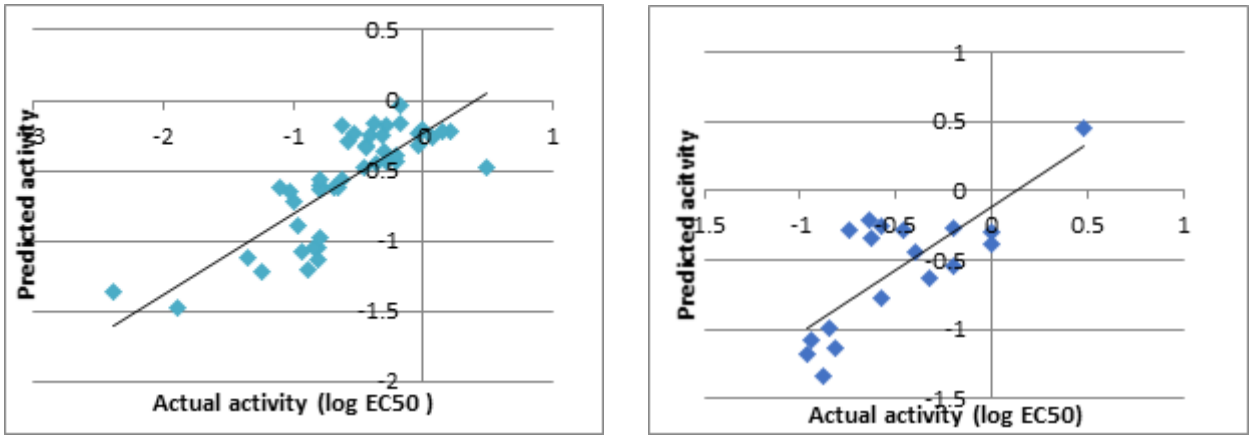
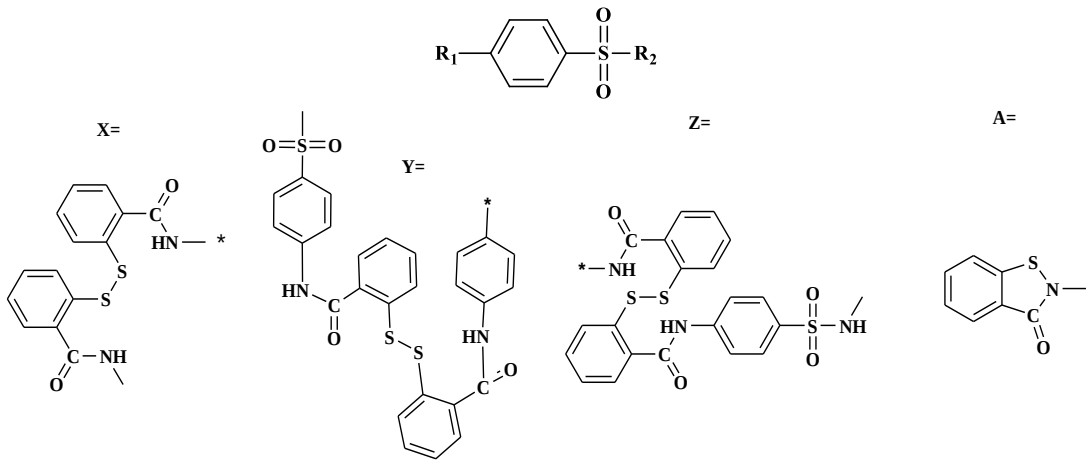
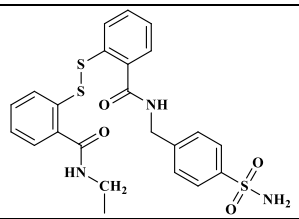
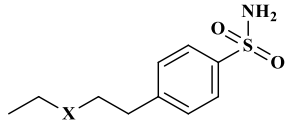
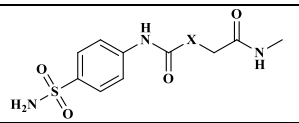
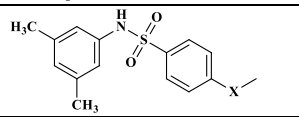
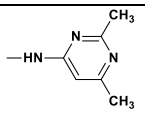
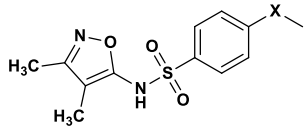
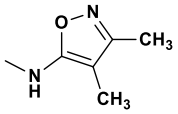
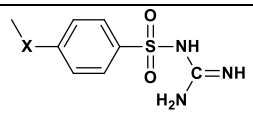
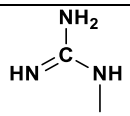
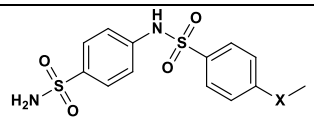
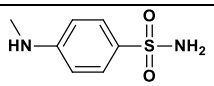
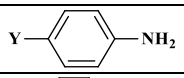
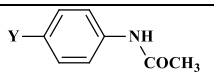
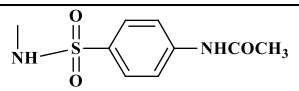
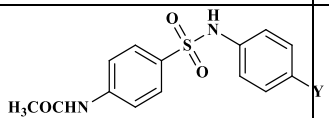
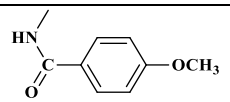
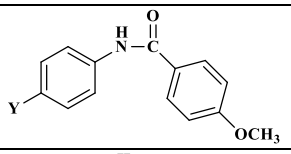
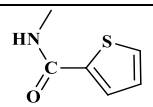
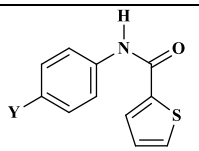
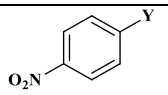


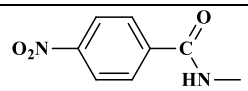
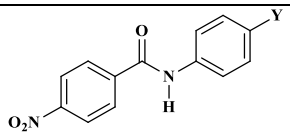
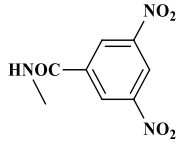
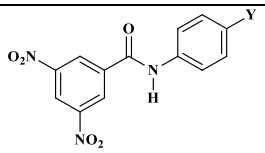
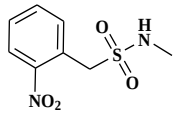
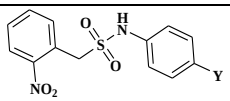
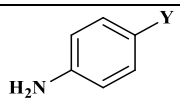
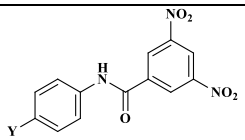
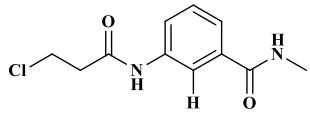
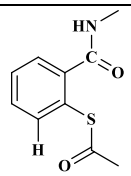
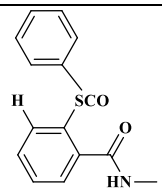
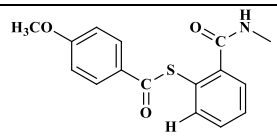
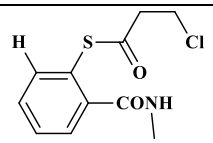
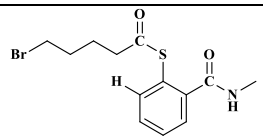
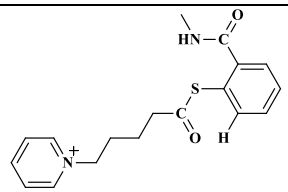
Fig. 1: Plot of Actual vs. Predicted EC_{50} values for training and test set (MLR analysis)

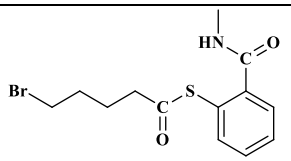
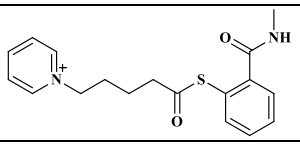
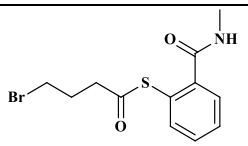
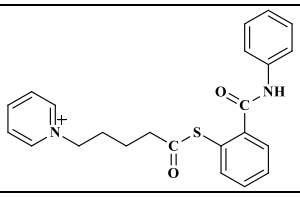
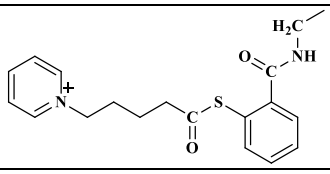
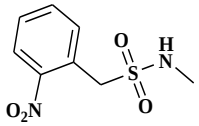
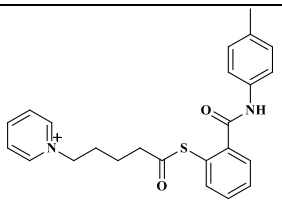
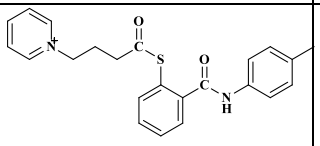
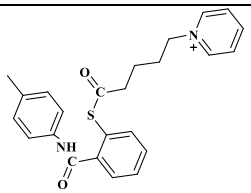
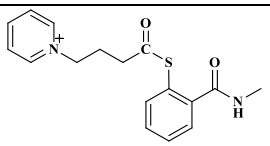
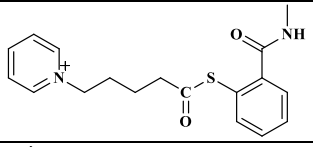
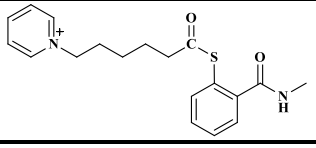


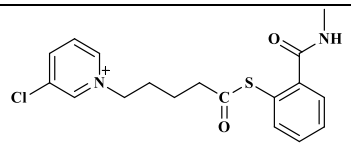
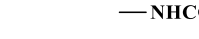
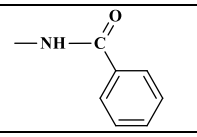
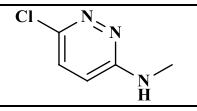
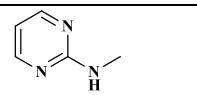
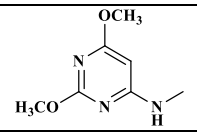
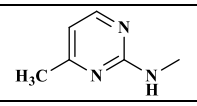
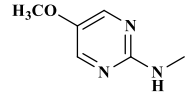
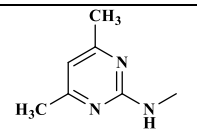
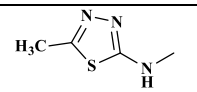
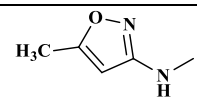
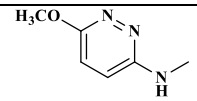
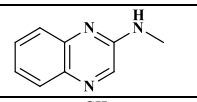
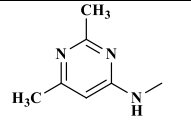
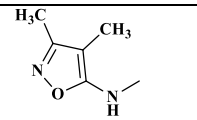
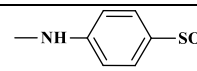
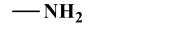
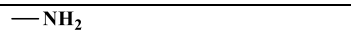
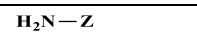
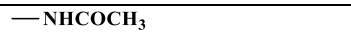
*Test Set Compounds

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

Comp	R_1	R_2	EC_{50}	Actual Activity (-log EC50)	Predicted Activity
1			0.33	0.481	-0.325
2			0.41		
*3			1.6	-0.204	-0.401
4			1.9	-0.278	-0.178
*5			2.5	-0.397	-0.439
6			111		
7			2.3	-0.361	-0.529
*8			4.3	-0.633	-0.392
9			1.5	-0.176	0.009
*10			1.0	0	-0.405
11			3.8		
12			12.2		
13			12.2		

14			78	-1.892	-1.644
15			240	-2.380	-1.544
16			12.3	0.206	-0.200
*17	—NH_2		1	0	-0.343
18	—$NHCOCH_3$		0.62	0.207	-0.200
19		—$NHCOCH_3$	38		
20		—$NHCOCH_3$	2.8	-0.414	-0.3006
21		—$NHCOCH_3$	2.6	-0.447	-0.401
22		—$NHCOCH_3$	2.8	-0.414	-0.3006
23		—$NHCOCH_3$	4.2	-0.447	-0.353
*24		—$NHCOCH_3$	3.8	-0.623	-0.443
*25		—$NHCOCH_3$	6.2	-0.579	-0.411

26		—NHCOCH$_3$	3.8	-0.792	-0.687
*27		—NHCOCH$_3$	6.2	-0.579	-0.690
28	—NO$_2$		1.6	-0.792	-0.665
*29	—NO$_2$		5.5	-0.204	-0.568
*30		—NH$_2$	1.1	-0.740	-0.339
31			2.9	-0.041	-0.288
*32	—NHCOCH$_3$		4.6	-0.462	-0.437
33	—NHCOCH$_3$		4.9	-0.692	-0.741
34		—NHCOCH$_3$	10.1	-1.004	-0.740
35		—NHCOCH$_3$	6.2	-0.792	-0.603
36		—NHCOCH$_3$	10.2	-1.021	-0.799

37			54	-0.968	-0.907
38	**A**		12.6		
39	**A**		9.3		
40	**A**		7.1	-0.812	-1.128
*41	**A**		6.5	-0.851	-0.731
42	**A**		7.6	-1.250	-1.282
*43	**A**		8.8	-0.880	-0.900
*44	**A**		17.8	-0.944	-0.787
45	**A**		6.9	-0.838	-1.149
46	**A**		7.8	-0.892	-1.264
47	**A**		8.7	-0.939	-1.138
48	**A**		22.6	-1.354	-1.210
49	**A**		9.2	-0.812	-1.247
*50	**A**		6.5	-0.963	-0.813
51	**A**		12.8	-0.812	-0.794
52	**A**		6.2	-0.792	-0.977
53			2	0.070	-0.277
54			0.85	-1.107	-0.640
55			1.5	-0.176	0.124

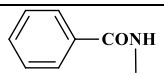
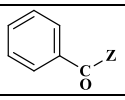
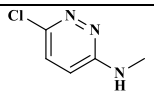
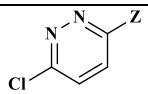
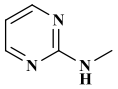
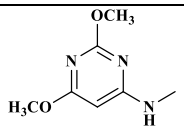
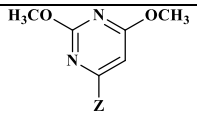
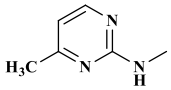
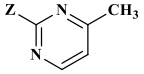
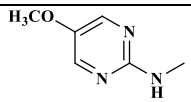
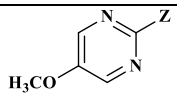
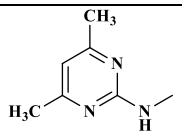
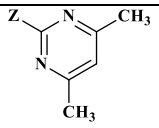
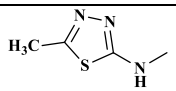
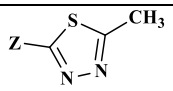
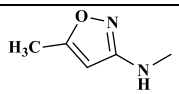
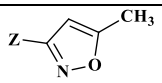
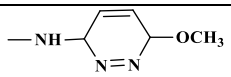
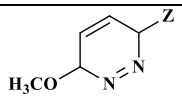
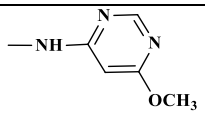
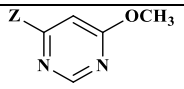
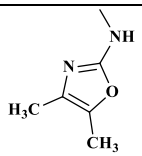
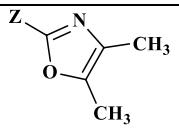
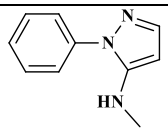
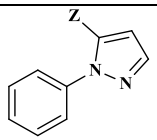
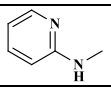
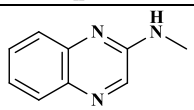
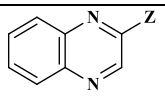
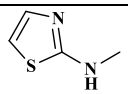
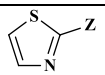
56			3.8	-0.579	-0.323
57			2.9	-0.462	-0.376
58			2	-0.301	-0.281
59			2.1	-0.322	-0.223
60			4.2	-0.623	-0.538
61			2.1	-0.204	-0.374
*62			1.6	-0.322	-0.574
63			3.4	-0.531	-0.091
64			1.9	-0.278	-0.319
65			4.2	-0.204	-0.446
66			1.6	-0.530	-0.091
67			2.4	-0.380	-0.075
*68			0.33	-0.623	-0.364
*69			1.1	0.481	0.496
70			0.72	-0.041	-0.287
71			1	0.142	-0.161

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

Statistical Tests	Values
s value	0.298
f value	95.183
F probability	2.332e-102
Regression coefficient, r	0.836
r^2(training)	0.655
r^2(test)	0.605
Cross Validation, $r^2_{(CV)}$	0.658
Residual Sum of Squares	3.651
Predictive Sum of Squares	4.139

Table 3: Correlattion matrix showing descriptor entered model

	Refractivity Similarity vs. Molecule 55	Shape Similarity vs. Molecule 68	EC_{50}
EC_{50}	1	-0.58528	0.83616
Refractivity Similarity vs. Molecule 55	-0.58528	1	-0.50619
Shape Similarity vs. Molecule 68	0.83616	-0.50619	1

Acknowledgement

The author thanks to Dr. Aditya Sharstri for providing facility.

References

1. Camarasa MJ, Velazquez S, San-Felix A., Perez-Perez M, Gago F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: a single mode of inhibition for the three HIV enzymes?. Antiviral Res. 2006;71:260–267.
2. Fader LD, Bethell R, Bonneau P, Bös M, Bousquet Y, Cordingley MG, Coulombe R., Deroy P, Faucher AM, Gagnon A., Goudreau N, Grand-Maître 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. Bioorganic & Medicinal Chemistry Letters. 2011,21(1):398–404.
3. *AIDS Epidemic Update*: UNAIDS/WHO. http://www.unaids.Org (November, 2009).
4. Sogaar OS, Lohse N, Gerstoft J, Kronborg G, Ostergaard L, Pedersen C, Pedersen G, Sørensen, HT, Obel N. Hospitalization for pneumonia among individuals with and without HIV infection. Clinical. Infectious. Disease, 2008**,**47(10):1345-53.
5. Hallenberge S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature.,1992:360,358–61.
6. Paliwal S, Narayan A, Paliwal S. Quantitative Structure Activity Relationship Analysis of Dicationic Diphenylisoxazole as Potent Anti-Trypanosomal Agents. QSAR Combinatorial. Science. 2009,28:1367–1375.
7. Balakrishnan, N, Childs A. Outlier, *Encyclopaedia of Mathematics*, 2001, Kluwer Academic Publishers, ISBN 978-1556080104.
8. Hodgkin EE, Richards WG. Molecular similarity based on electrostatic potential and electric field. International Journal. Quantum. Chemistry. 1987,14:105–110.
9. Richards WG, Hodgkin EE. Molecular Similarity. Chem. in Brit., 1988,*2*4,1141-1144.
10. Fayyad Usman M, Data Mining and Knowledge Discovery In Databses: Applications in Astronomy and Planetary Science, 1996, AAA/MIT Press.
11. Duhman MH, Data mining: Inroductory and advanced topics. Princtine Hall, Upper Saddle River, New Jersey, USA 2002.
12. Turpin Jim A, Song Y, Inman JK, Huang M, Wallqvist A, Maynard A., Covell DG, Rice WG, Apella E. Synthesis and Biological Properities of Novel Pyridinoalkanoyl Thiolesters (PATE) as Anti-HIV Agents That Target The Viral Nucleoacapsid Protein Zinc Fingers. Journal of Medicinal Chemistry,1999,42**:**67-86.
13. Rameshwar N, Krishna K., Ashok Kumar B, Parthasarathy T. QSAR studies of N1-(5-chloro-2- pyridyl)-2-{[4-(alkyl methyl)] amino}-5-chlorobenzamide analoges. Bioorganic & Medicinal Chemistry, 2006**,**14,319- 325.