International Journal of Drug Design and Discovery

***Volume* 4 • *Issue* 3 • *July–September* 2013. 1117-1123**

Designing Hypothesis of Substituted N-Arylsulfonyl-3-Acetylindoles as Anti-HIV Agents: 3D QSAR Study

Ravichandran V[1*], Harish R[2], Shalini S[1], Dhanaraj SA[1] and Suresh Kumar K[3]

[1]*Faculty of Pharmacy, AIMST University, Semeling – 08100, Kedah, Malaysia.*

[2]*Faculty of Pharmacy, Guru Khasidas University, Madhyapradesh, India.*

[3]*KMCH College of Pharmacy, Coimbatore, Tamilnadu, India.*

ABSTRACT: The 3D QSAR models were developed by k-nearest neighbour-molecular field analysis using a database consisting of 35 recently discovered N-arylsulfonyl-3-acetylindole derivatives. Presence of electrostatic and steric field descriptors significantly affects the ability of N-arylsulfonyl-3-acetylindole derivatives to inhibit HIV-1 activity. Negative range of steric potential are favorable for increase in activity and hence small steric groups (Cl, F, CH_3, C=O, SO_2 or CN) is preferred in the specific region. Positive range of electro potential is favorable for increase in activity and hence a less electronegative is preferred in the particular region (branched alkyl groups, cyclic groups). The selected descriptors could serve as a primer for the design of novel and potent anti-HIV agents.

KEYWORDS: Human immunodeficiency virus; Anti-HIV; Quantitative Structure Activity Relationship; N-arylsulfonyl-3-acetylindole derivatives.

Introduction

Since the first case of acquired immune deficiency syndrome (AIDS) was reported in 1981, the human immunodeficiency virus (HIV)/AIDS has always been a global health threat and the leading cause of deaths[1]. Therefore, the rapid worldwide spread of AIDS has prompted an intense research effort to discover compounds that could effectively inhibit HIV. In the past two decades, 25 drugs, including nucleoside/nucleotide viral reverse transcriptase (RT) inhibitors (NRTIs), non-nucleoside RT inhibitors (NNRTIs), protease inhibitors (PIs), integrase inhibitors (INIs) and fusion (or entry) inhibitors (FIs) were approved for clinical use in the world[2]. However, these drugs have only limited or transient clinical benefit due to their severe side effects and the emergence of viral variants resistant to HIV-1 inhibitors[3]. Consequently, it is imperative that the design and development of new, selective and safe drugs for the treatment of HIV-1[4,5].

In search of effective N-arylsulfonyl-3-acetylindole derivatives with minimal viral resistance problem, Jun-Qiang *et al.* synthesized and evaluated a novel set of N-arylsulfonyl-3-acetylindole for their anti-HIV activity[6]. The 2D QSAR studies have been carried out for anti-HIV activity of different group of compounds[7-15]. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) have also been carried out for anti-HIV activity of a given set of molecules in rational drug design and its related applications[16-25].

As a part of ongoing efforts to design novel molecules with potent anti-HIV activity, the 3D QSAR analysis were performed on 35 N-arylsulfonyl-3-acetylindole derivatives (7) (Table 1) to correlate their anti-HIV activity with 3D descriptors.

Experimental

General Procedure

Molecular modeling studies were performed using software QSARPro (Product of VLife Sciences Technologies Private Limited, India with website: www.vlifesciences.com).

Biological data

Biological and chemical data of thirty five N-arylsulfonyl-3-acetylindole derivatives from the work of Jun-Qiang *et al.*[6] (Table 1) were used. A high structural diversity and a sufficient range of biological activity were observed in the selected series of N-arylsulfonyl-3-acetylindole derivatives. Anti-HIV activity used in the present study were expressed as $pEC_{50} = -\log_{10} EC_{50}$, where EC_{50} is the micro molar concentration of the compounds producing 50 % reduction in the HIV-1 activity stated as the mean of at least two experiments.

* *For correspondence:* Ravichandran

Tel: 006-04-4298000 Ext. 1029; Fax: 006-04-4298007

Email: phravi75@rediffmail.com

Sketching of molecules

The 2D structures of the compounds were drawn by using QSARPro software. The structures were then checked for errors, cleaned up and transferred to 3D form.

Energy minimization

The geometry of 3D structure was optimized to local minima by Merck Molecular Force Field (MMFF) by considering distance-dependent dielectric constant of 1.0, convergence criterion or root-mean-square (RMS) gradient at 0.001 kcal/mol Å and the iteration limit to 10,000. The most stable structure for each compound was generated and saved as .mol 2 files for computing various physico-chemical and alignment independent descriptors.

3D QSAR analyses

The molecular modeling studies were performed using QSARPro, supplied by VLife. The energy minimized N-arylsulfonyl-3-acetylindole derivatives were used for 3D QSAR studies. The most significant requisite for any 3D-QSAR study is to align the data set on a suitable conformational template, either by taking a reported crystal structure of a bioactive compound or by considering the most active compound.

The most active compound was considered as the template for the alignment due to the absence of any reported crystal structure. The indole moiety of the bioactive compound **30** was used as a substructure and the rest of the molecules were aligned on it using database alignment method. Electrostatic, steric and hydrophobic field descriptors were selected for 3D QSAR studies. For calculation of field descriptor values, the cutoff values were set at 10.0 and 30.0 kcal/mole for electrostatic and steric field respectively. The charge type was selected as Gasteiger-Marsili. Dielectric constant was set to 1.0 considering the distance dependent dielectric function.

Calculation of Descriptors (Independent variables)

The QSARPro (VLife) software allows the user to choose probe, grid size, and grid interval for the generation of descriptors. After suitable alignment of a given set of molecules, a common rectangular grid (lattice) was generated around the molecules. The steric, electrostatic and hydrophobic interaction energies were computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values were considered for relationship generation and could be utilized as descriptors. The descriptors were considered as independent variables and biological activity as dependent variables.

Training and test set selection

The sphere exclusion (SE) and random selection methods were adopted for division of training and test data set. The dissimilarity value used in sphere exclusion method was 5, where the dissimilarity value gives the sphere exclusion radius. The compounds 7, 13, 15, 27 and 35 were selected as test sets and remaining 30 compounds as training set. The predictive ability of the QSAR model was found to be less when the dissimilarity level is high. The selection of training and test set was further justified by uni-column statistics calculated for each case of study.

Feature selection and model development

The kNN-MFA method was used to generate 3D QSAR models using the variable selection methods viz. stepwise (SW) forward- backward, genetic algorithm (GA) and simulated annealing (SA). In all the methods F-test 'in' was set to '4.0' and F-test 'out' was set to '3.99'. As some additional parameters, variance cut off was set as 2 kcal/mol Å, scaling was set as auto-scaling, additionally the k-nearest neighbour parameter was set as 5 and the prediction method was selected as distance base weighted average.

Model quality and validation

The developed QSAR models were evaluated using the following statistical measures: n, (the number of compounds in regression); k, (number of variables); DF, (degree of freedom); F test, (Fischer's value) for statistical significance; q^2, (cross-validated correlation coefficient); q^2_se, (standard error of cross-validated square correlation co-efficient); pred_r^2, (r^2 for external test set); pred_r^2se, (standard error of predicted squared regression). The F-test reflects the ratio of variance explained by the model and variance due to the error in regression. High values of F-test indicate that the model is statistically significant. The low standard error of q^2 (q^2_se) and pred_r^2 (Pred_r^2se) value shows absolute quality of fitness of the model.

Internal validation was carried out using 'leave-one-out' (q^2, LOO) method.[26, 27] The cross-validated coefficient, q^2, was calculated using the following formula [1]:

$$q^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad \text{.....[1]}$$

where y_i, and $\hat{y}_i$ are the actual and predicted activity of the i^{th} molecule in the training set, respectively and y_{mean} is the average activity of all molecules in the training set. However, a high q^2 value does not necessarily give a suitable representation of the real predictive power of the model for anti-HIV activity. So, an external validation was also carried out in the present study. The external predictive power of the model was assessed by predicting -logEC_{50} value of the test set molecules by using the formula [2], which were not included in the QSAR model development. The predictive

ability of the selected model was also confirmed by $r^2–r^2_0/r^2$, $r^2–r'^2_0/r^2$, k and k'.

$$\text{pred\_}r^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad \text{.....[2]}$$

where y_i, and $\hat{y}_i$ are the actual and predicted activity of the i^{th} molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Another term to check the external predictability of the selected model is r^2_m, which was proposed by Roy and Roy (2007)[28] and it was calculated by the following formula [3]:

$$r^2_m = r^2\left(1 - \sqrt{\left|r^2 - r^2_0\right|}\right) \quad \text{.....[3]}$$

Where r^2 is squared correlation coefficient between observed and predicted values and r_0^2 is squared correlation coefficient between observed and predicted values with intercept value set to zero. A value of r^2_m is greater than 0.5 may be taken as an indicator for good external predictability.

In 3D QSAR, electrostatic, steric and hydrophobic fields were generated around the aligned molecules in the grid. Negative values in electrostatic field descriptors (blue points in the contour maps) indicates that negative electronic potential is essential for its activity and more electronegative substituent group is preferred at that position, and positive electronic potential range indicates the vice-versa. Negative range in steric field (green points in the contour maps) signifies that negative steric potential is required for activity and less bulky substituent group is preferred in that region. The positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region. Positive value in hydrophobic field descriptors (yellow points in the contour maps) indicates that positive hydrophobic potential is favorable for activity and more hydrophobic substituent group is preferred at that position, and negative hydrophobic potential range indicates the vice-versa

Results and Discussion

The 3D QSAR study of template aligned 35 N-arylsulfonyl-3-acetylindole derivatives (Figure 1) for anti-HIV activity (Table 1) was performed through kNN-MFA methodology.

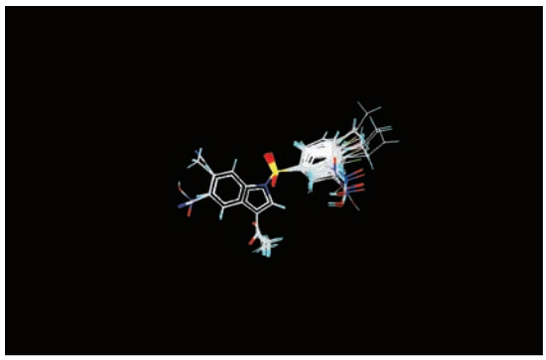
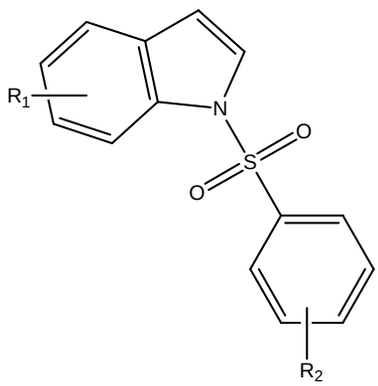
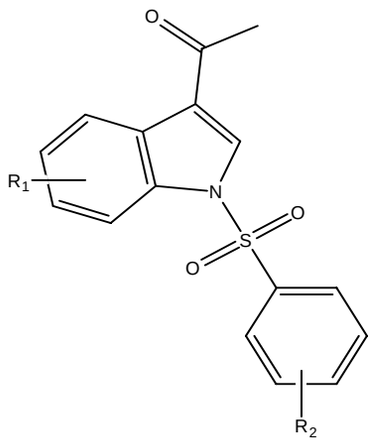


Fig. 1 Template based alignment of N-arylsulfonyl-3-acetylindole derivatives.

Table 1 Structures and Selected 3D Parameters of the N-arylsulfonyl-3-acetylindole Derivatives.



Compound 1-18



Compound 19-35

Compound No	R^1	R^2	S_740	S_775	E_781
1	H	H	–0.02649	–0.14758	0.07203
2	H	3-NO_2	–0.04413	–0.16829	0.092652
3	H	3-NO_2, 4-Cl	–0.04811	–0.18953	0.01084
4	H	4-Br	–0.02695	–0.16053	–0.10215

Compound No	R^1	R^2	S_740	S_775	E_781
5	H	4-Cl	–0.02704	–0.16167	–0.14398
6	5-NO_2	4-Br	–0.0271	–0.16408	–0.03641
7a	5-NO_2	H	–0.02658	–0.14924	0.150323
8	5-NO_2	4-CH_3	–0.02707	–0.14608	0.157209
9	6-CH_3	H	–0.02825	–0.17816	0.014919
10	6-CH_3	3-NO_2	–0.04899	–0.19768	0.062093
11	6-CH_3	4-CH_3	–0.0292	–0.17598	–0.00131
12	6-CH_3	4-CH_2CH_3	–0.03004	–0.17848	–0.02482
13a	6-CH_3	4-Cl	–0.0287	–0.19561	–0.15718
14	5-CN	4-CH_3	–0.02697	–0.14199	0.147817
15a	5-CN	4-Br	–0.0259	–0.15577	–0.10197
16	5-CN	H	–0.02646	–0.14502	0.138795
17	5-CN	4-Cl	–0.02707	–0.16017	–0.08209
18	5-CN	4-CH_2CH_3	–0.02739	–0.14285	0.127359
19	H	4-CH_3	–0.06309	–0.15127	0.314457
20	H	3-NO_2	–0.07448	–0.17183	0.170528
21	H	3-NO_2, 4-Cl	–0.07426	–0.17163	–0.08199
22	H	4-Br	–0.0665	–0.16553	–0.02408
23	H	4-Cl	–0.06475	–0.16987	–0.03155
24	5-NO_2	4-Br	–0.06475	–0.16987	–0.03155
25	5-NO_2	H	–0.06228	–0.15938	0.264747
26	6-CH_3	H	–0.0701	–0.1741	0.167898
27a	6-CH_3	4-Br	–0.07276	–0.1858	–0.05752
28	6-CH_3	3-NO_2	–0.08205	–0.19047	0.146596
29	6-CH_3	4-CH_3	–0.06842	–0.1686	0.196787
30	6-CH_3	4-CH_2CH_3	–0.07006	–0.17965	0.136913
31	6-CH_3	4-Cl	–0.07288	–0.18655	–0.11544
32	5-CN	H	–0.05836	–0.14921	0.289801
33	5-CN	4-Cl	–0.06035	–0.15988	–0.04449
34	5-CN	3-NO_2	–0.06801	–0.16423	0.21899
35a	5-CN	4-CH_2CH_3	–0.05892	–0.15127	0.326962

a – test set compound

Based on various feature selection methods viz. SW, GA and SA using QSARPro (VLife) software many 3D QSAR models were generated, considering the term selection criterion as q^2. The training and test sets were selected by sphere exclusion (dissimilarity value 5) and the models were validated by both internal and external validation. A uni-column statistics for training set and test set was generated to check the correctness of selection criteria for trainings and test set molecules. Some of the selected 3D QSAR models are given below (equations 4, 5 and 6):

Model 1: kNN-MFA -SW-SE - (dissimilarity value 5)

Training Set Size = 30, Test Set Size = 5

Statistics:

k Nearest Neighbour = 2, n = 30, Degree of freedom = 26, q^2 = 0.783, q^2_se = 0.400,

Pred r^2 = 0.770, pred_r^2se = 0.375

Descriptor Range:

S_740: –0.0820 to -0.0701, E_781: 0.1466 to 0.1679, S_775: –0.1905 to -0.1741 …..[4]

Model 2: kNN-MFA-GA-SE-(dissimilarity value 5)

Training Set Size = 30, Test Set Size = 5

Statistics:

k Nearest Neighbour = 3, n = 30, Degree of freedom = 26, q^2 = 0.485, q^2_se = 0.524,

Pred r^2 = 0.489, pred_r^2se = 0.541

Descriptor Range:

E_4: -0.1854 to -0.1668, E_128: -0.2172 to -0.1688, H_680: 0.5323 to 0.6703 …..[5]

Model 3: kNN-MFA-SA-SE-(dissimilarity value 5)

Training Set Size = 30, Test Set Size = 5

Statistics:

k Nearest Neighbor= 5, n = 30, Degree of freedom = 25, q^2 = 0.519, q^2_se = 0.384,

Pred r^2 = 0.663, pred_r^2se = 0.443

Descriptor Range:

S_754: –0.2558 to 0.6858, E_140: –0.1814 to 0.0653,
S_267: –0.0075 to –0.0073, E_236: –3.1221 to 5.2697

…..[6]

Developed kNN-MFA models showed the relative position and ranges of the corresponding electrostatic, steric and hydrophobic fields and provided the guidelines for designing new molecules. Negative range of steric potential are favorable for increase in activity and hence small steric groups (Cl, F, CH_3, C=O, SO_2 or CN) is preferred in that region. Positive range of electro potential is favorable for increase in activity and hence a less electronegative is preferred in this region (branched alkyl groups, cyclic groups).

The points which contribute to the selected kNN-MFA model 1 in data sets are illustrated in Figure 2. The residual values of observed and predicted anti-HIV activity are shown graphically in Figure 3. The range of property values for the chosen points may also assist in design of new potent molecules. The range is based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbor set.

The positive range of hydrophobic field descriptors in GA feature selection method indicates that the substitution of hydrophobic groups at these positions is conductive for anti-HIV activity. It also indicates that these hydrophobic groups may interact with amino acid residues (Leu234, Tyr318, His235, Phe227, Trp229, Tyr118) of HIV-1 reverse transcriptase and inhibit the HIV multiplication. Thus kNN-MFA models provide direction for design of new molecules in a rather convenient way.

Uni-column statistics for models 1, 2 and 3 are given in Table 2. Cross validated correlation coefficient of the best model (model 1) was found to be 0.783 which explains the efficiency of internal prediction of the model. The external predictive power of the model was 0.770. A closer view of the selected descriptors suggests that descriptors included in the model S_740, E_781and S_775 (Table 1) play a significant role in the structure activity relationship (Table 3). The other best models are reported as model 2 and 3.

The proposed QSAR model 7 was predictive since it satisfies the following conditions:

r^2_{Pred} = 0.770 > 0.6, r^2_m = 0.511 > 0.5; k = 0.900; k' = 1.089 (0.85 < k or k' < 1.15); r^2-r^2_0/r^2 = -0.085; r^2-r'^2_0/ r^2 = -0.059 (r^2-r^2_0/r^2 or r^2-r'^2_0/ r^2 < 0.1).

The applicability domain was established for model 1, determining the leverage values for each compound. Figure 4 shows the Williams plot (plot of standardized residuals (*y*-axis) *versus* leverages (*x*-axis)) for each compound.

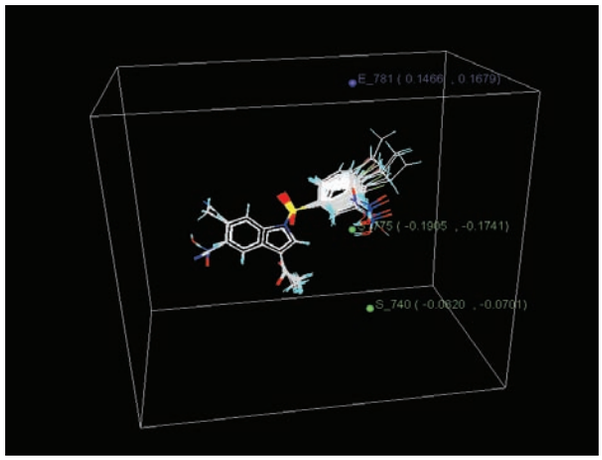


Fig. 2 Contour maps for Model 1 E- Electrostatic, S- Steric.

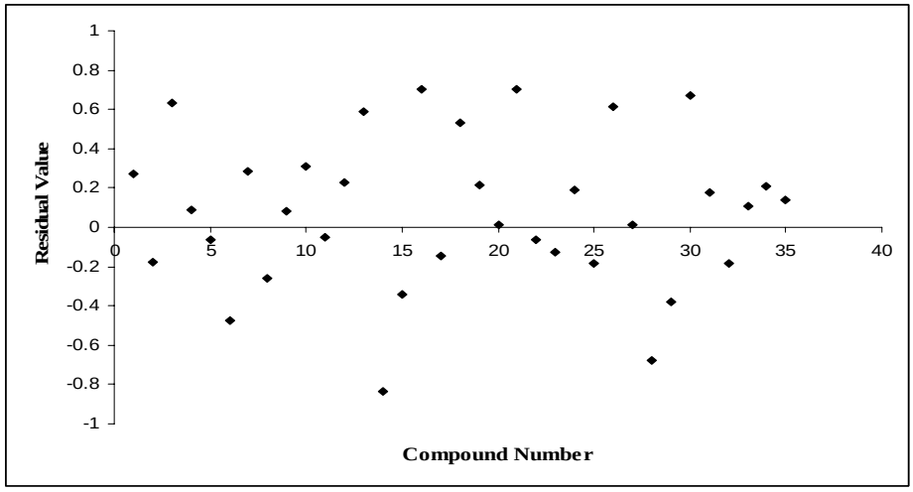


Fig. 3 Residual plot between experimental and predicted anti-HIV activity of N-arylsulfonyl-3-acetylindole derivatives by Model 1.

Table 2 Uni-column Statistics for Model 1, 2 and 3.

Column Name	Average	Max	Min	Std. Dev	Sum
Training Set-logEC_{50}	2.1577	3.8861	1.0541	0.8594	64.7318
Test Set-logEC_{50}	2.1004	3.1427	1.1584	0.7801	10.5022

Table 3 Experimental and Predicted Anti-HIV Activity of the N-arylsulfonyl-3-acetylindole Derivatives.

Compound No	Anti-HIV activity (-logEC_{50})	
	Experimental	Predicted
1	3.032	2.760
2	3.131	3.307
3	3.420	2.787
4	1.690	1.604
5	1.909	1.971
6	1.203	1.680
7a	2.318	2.031
8	1.580	1.841
9	3.051	2.970
10	3.585	3.275
11	2.523	2.572
12	2.089	1.861
13a	3.143	2.554
14	1.198	2.035
15a	1.158	1.498
16	2.491	1.789
17	1.305	1.449
18	2.381	1.847
19	1.507	1.289
20	2.674	2.660
21	2.465	1.762
22	1.093	1.159
23	1.054	1.178
24	1.264	1.073
25	1.339	1.523
26	3.444	2.831
27a	2.368	2.354
28	2.991	3.666
29	1.859	2.241
30	3.886	3.216
31	2.256	2.081
32	1.240	1.423
33	1.264	1.159
34	1.811	1.602
35a	1.516	1.374

a – test set compound

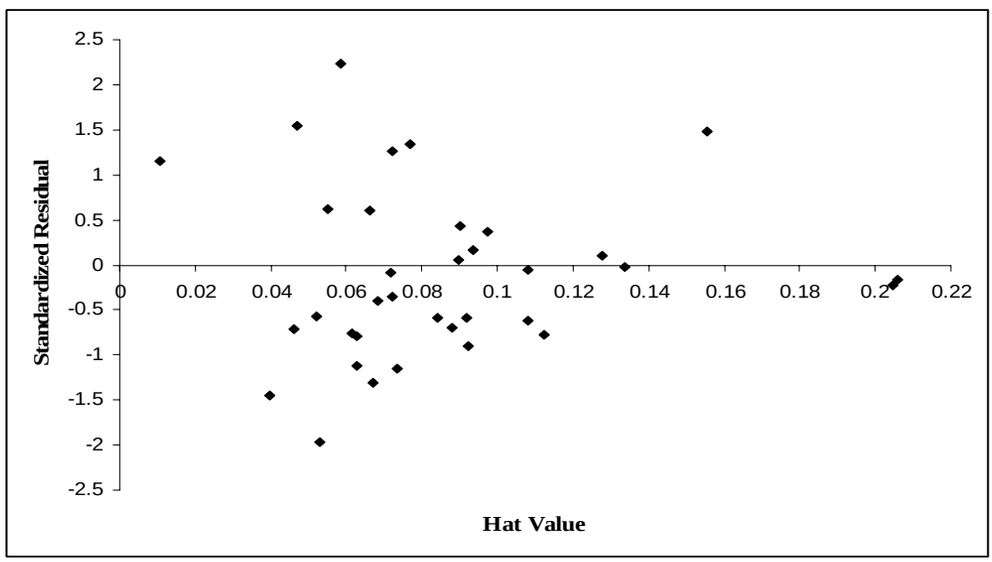


Fig. 4 Williams plot for Model 1: plot of standardized residuals (*y*-axis) *versus* leverages (hat values; *x*-axis) for each compound.

The applicability domain was established inside a squared area with standard deviations of 2.5 for this model. A leverage threshold value of $h^* = 0.342$ ($h^* = 3p'/n$, being p' the number of model parameters + 1, and n the number of compounds) was observed for this model. As seen in figure 4 all compounds of training and test set were found to be inside of the square area. The predicted anti-HIV activity data must be considered reliable only for those chemicals that fall within the applicability domain on which the model was constructed for future predictions.[29]

It was also stated that the reported QSAR models may not be useful to predict the activity of other type of molecules against HIV-1. The applicability domain of the derived QSAR models is specifically related to N-arylsulfonyl-3-acetylindole derivatives. However, it is very important to point out an eventual limitation of QSAR models i.e. activity cliffs shown by similar molecules exhibits different biological activity[30].

Conclusion

The established 3D QSAR models for N-arylsulfonyl-3-acetylindole derivatives were found to be statistically significant and highly predictive to anti-HIVactivity. Small steric groups (Cl, F, CH_3, C=O, SO_2 or CN) and less electronegative groups (branched alkyl groups, cyclic groups) in the specific positions are essential to design potent anti-HIV agents. It was concluded that the reported 3D QSAR models might be used specifically for predicting anti-HIV activity of N-arylsulfonyl-3-acetylindole derivatives.

Acknowledgment

Author is thankful to Dr. Yogesh Wagh, VLife Sciences, Pune, India for providing software to carry out this work.

References

[1] Barreca M. L.; Carotti A.; Chimirri A.; Monforte A. M. Bioorg. Med. Chem. 1999, 7, 2283.

[2] Buolamwini J.K.; Assefa H. J. Med. Chem. 2002, 45, 841.

[3] Cramer R. D.; Patterson D. E.; Bunce J. D. J. Am. Chem. Soc. 1988, 110, 5959.

[4] De Clercq E. Biochim. Biophys. Acta, 2002, 1587, 258.

[5] Debnath A.K. J. Med. Chem. 1999, 42, 249.

[6] Deng J.; Dayam R.; Al-Mawsawi L. Q.; Neamati N. Curr. Pharm. Des. 2007, 13, 129.

[7] Golbraikh A.; Tropsha A. J. Mol. Graph. Model. 2002, 20, 269.

[8] Hu R.; Barbault F.; Maurel F.; Delarmar M.; Zhang R. Chem. Biol. Drug Design, 2010, 76, 518.

[9] Jayatilleke P. R. N.; Nair A. C.; Zauhar R.; Welsh W. J. J. Med. Chem. 2000, 43, 4446.

[10] Johnston M. I.; Hoth D. F. 1993, 260, 1286.

[11] Jun-Qiang R.; Ning H.; Xu H.; Yang L. M.; Min L.; Zheng, Y. T. Bioorg. Med. Chem. Lett. 2010, 20, 3534.

[12] Klebe G.; Abraham U.; Mietzner T. J. Med. Chem. 1994, 37, 4130.

[13] Leonard J. T.; Roy K. Bioorg. Med. Chem. 2006, 14, 1036.

[14] Maggiora G. J. Chem. Inf. Model. 2006, 46, 1535.

[15] Martins S.; Ramos M. J.; Fernandes P. A. Curr. Med. Chem. 2008, 15, 1083.

[16] Pungpo P.; Hannongbua S. J. Mol. Graph. Modell. 2000, 18, 581.

[17] Qin B.; Jiang, X.; Lu H.; Tian X.; Barbault F.; Huang L.; Qian K.; Chen C. H.; Huang R. J. Med. Chem. 2010, 53, 4906.

[18] Ravichandran V. Mol. Simulation, 2010, 36, 878.

[19] Ravichandran V.; Agrawal R. K. Bioorg. Med. Chem. Lett. 2007, 17, 2197.

[20] Ravichandran V.; Jain A.; Mourya V. K.; Agrawal R. K. Chem. Pap. 2008, 62, 596.

[21] Ravichandran V.; Jain P. K.; Mourya V. K.; Agrawal R. K. Med. Chem. Res. 2007, 16, 342.

[22] Ravichandran V.; Mourya V. K.; Agrawal R. K. Digest J. Nanomat. Biostruct. 2008, 3, 19.

[23] Ravichandran V.; Mourya V. K.; Agrawal R. K. J. Enzym. Inhib. Med. Chem. 2011, 26, 288.

[24] Ravichandran V.; Prashanthakumar B. R.; Sankar S.; Agrawal R.K. Eur. J. Med. Chem. 2009, 44,1180.

[25] Ravichandran V.; Prashanthakumar B. R.; Sankar S.; Agrawal R. K. Med. Chem. Res. 2008, 17, 1.

[26] Ravichandran V.; Shalini S.; Sundram K.; Dhanaraj S. A. Eur. J. Med. Chem. 2010, 45, 2791.

[27] Roy P. P.; Roy K. Expert. Opin. Drug Discov. 2007, 2, 1567.

[28] Sahu K. K.; Ravichandran V.; Jain P. K.; Sharm S.; Mourya V. K.; Agrawal R. K. Med. Chem. Res. 2007, 5, 418.

[29] Vighi M.; Gramatica P.; Consolaro F.; Todeschini R. Ecotoxicol. Environ. Safety, 2001, 49, 206.

[30] Xu H.; Lv M. Curr. Pharm. Des. 2009, 15, 2120.