

RESEARCH

Open Access



Rational in silico drug design of HIV-RT inhibitors through G-QSAR and molecular docking study of 4-arylthio and 4-aryloxy-3-iodopyridine-2(1-H)-one derivative

Debadash Panigrahi^{1*} , Amiyakanta Mishra¹ and Susanta Kumar Sahu²

Abstract

Background: Human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) is a spectrum of conditions caused by infection with the human immunodeficiency virus (HIV). Antiretroviral therapy (ART) against HIV infection offers the promise of controlling disease progression and prolonging the survival of HIV-infected patients. Reverse transcriptase (RT) inhibitors remain the cornerstone of the drug regimen to treat AIDS. In this direction, by using group-based QSAR study (G-QSAR), identification of the structural need for the development of lead structure with reverse transcriptase inhibition on 97 reported structures was carried out. Docking analysis was performed further and suggested the structural properties required for binding affinity with the receptor. The molecules in the data set were fragmented into six (R1, R2, R3, R4, R5, and R6) by applying the fragmentation pattern. Three G-QSAR models were selected based on the statistical significance of the model. The molecular docking study was performed to explain the structural properties required for the design of potent HIV-RT inhibitors.

Results: The statistically validated QSAR models reveal the presence of higher hydrophobic groups containing single-bonded –Br atom, 2 aromatic bonded –NH group with less electronegativity, and entropic interaction fields at R2 essential for better anti-HIV activity. The presence of a lipophilic group at R3, oxygen and sulfur connected with two aromatic bonds at R4, and –CH₃ group at R5 was fruitful for reverse transcriptase inhibition. Docking studies of the selected inhibitors with the active site of reverse transcriptase enzyme showed hydrogen bond, Van der Waal's, charge, aromatic, and π - π interactions with residues present at the active site.

Conclusion: The results of the generated models provide significant site-specific insight into the structural requirements for reverse transcriptase inhibition during the design and development of novel anti-HIV compounds. Molecular docking study revealed the binding interaction between the ligand and the receptor which gave insight towards the structure-based design for the discovery of more potent compounds with better activity against HIV infection.

Keywords: Quantitative structure activity relationship, G-QSAR, Antiretroviral therapy (ART), Anti-HIV, Reverse transcriptase inhibitor, Molecular docking

* Correspondence: debmpfarm@yahoo.co.in

¹Drug Research Laboratory, Nodal Research Centre, College of Pharmaceutical Sciences, Puri, Baliguali, Puri- Konark Marine Drive road, Puri, Odisha 752002, India

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

1 Background

Human immunodeficiency virus (HIV) attacks the body's immune system, specifically the CD4 cells (T cells), which help the immune system fight off infections. HIV reduces the number of CD4 cells in the body, making the person more likely to get other infections or infection-related cancers [1, 2]. Over time, HIV can destroy so many of these cells that the body cannot fight off infections and leads to acquired immunodeficiency syndrome (AIDS), the last stage of HIV infection [3]. It is one of the world's most significant public health challenges, particularly in low- and middle-income countries. In 2017, approximately 36.9 million people (35.1 million adults) were living with HIV and 1.8 million people became newly infected, globally. Nearly 1 million people died from AIDS-related illnesses in 2017 [4]. The medicine used to treat HIV is called antiretroviral therapy (ART) [5, 6]. There are various antiretroviral drugs available in the market such as entry or fusion inhibitors, nucleoside or non-nucleoside reverse transcriptase inhibitors (NRTI/NNRTI), integrase inhibitors (IN), protease inhibitors (PI), and maturation inhibitors [7]. The

resistance of the virus to the available antiretroviral drugs is the biggest challenge for ART, and the discovery of new anti-HIV agents to overcome this resistance is continually required [8]. Human immunodeficiency virus (HIV) is a retrovirus because of the presence of an enzyme reverse transcriptase (RT), which possesses both RNA-dependent DNA polymerase (RDDP) and ribonuclease-H (RNase H) activities that work in tandem to convert viral genomic single-stranded RNA to double-stranded DNA by the process reverse transcription and by retrotransposon, mobile genetic elements are then integrated into the DNA of the infected host cell to cause disease [9–11]. Hence, reverse transcriptase inhibitors both nucleoside reverse transcriptase inhibitor (NRTI's) and non-nucleoside reverse transcriptase inhibitor (NNRTI's) are active against HIV which inhibits the process reverse transcription that leads to inhibition of formation of DNA from viral RNA (Fig. 1) for insertion into the host DNA sequence cause treatment of the disease [12, 13].

Traditional drug development methods are based on random screening, which has a major disadvantage like a

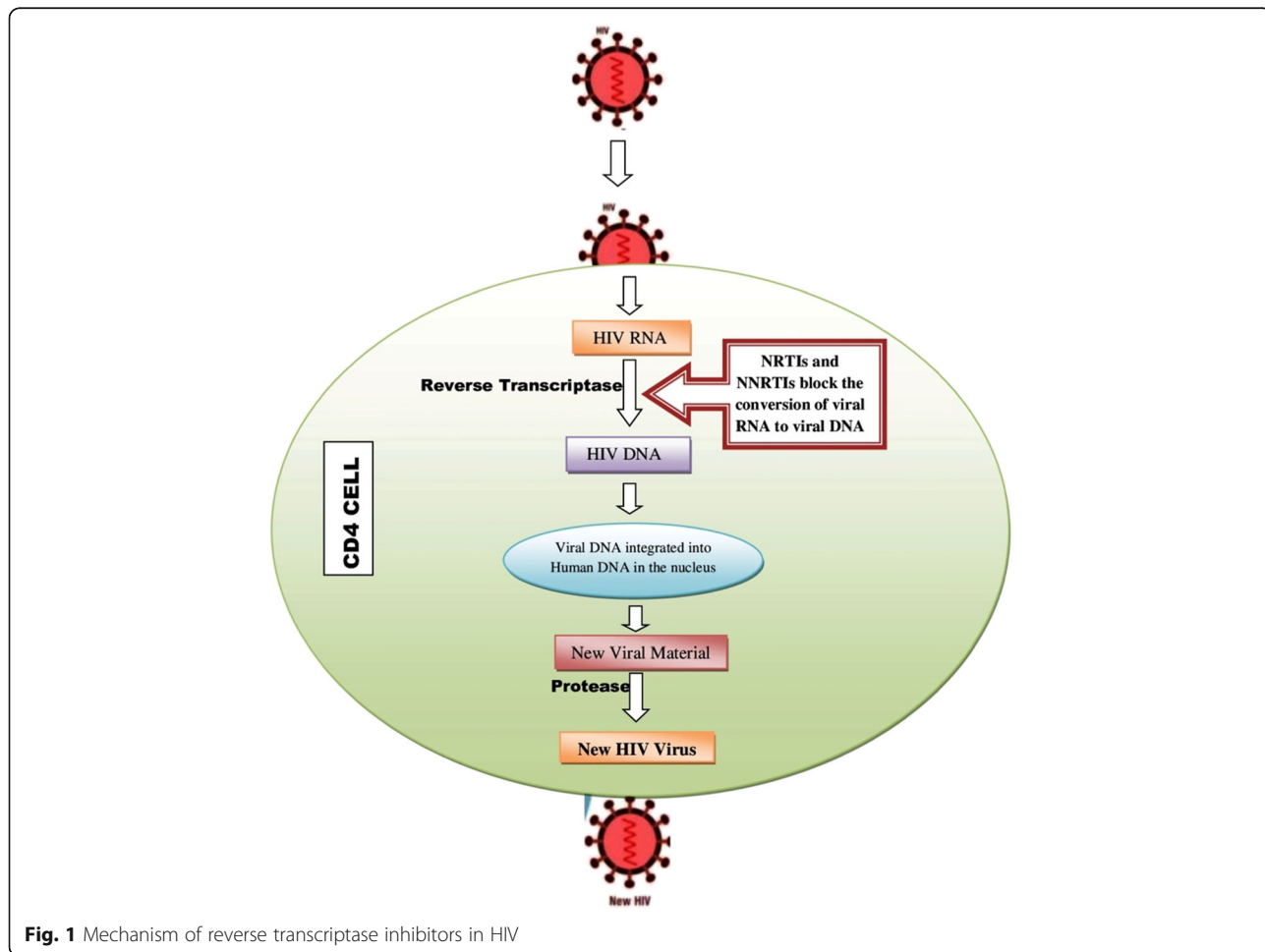


Fig. 1 Mechanism of reverse transcriptase inhibitors in HIV

lengthy, expensive, and intellectual method. To overcome these disadvantages in the recent decade, the emergence of computer-aided drug development (CADD) processes are taken place [14]. These methods are run at a lower cost and give a viable option for the screening of potential drug candidates. The number of methodologies is involved in the CADD, out of this quantitative structure-activity relationship (QSAR) is one such efficient chemoinformatic tool, which aims to identify and quantify the relationship between molecular structure and certain physicochemical properties for the development of predictive models [15]. QSAR models like 2D and 3D are having their own merits and demerits. Generated 2D QSAR models are only correlating the physicochemical properties of the molecules with their biological activity. These models do not specify the site at which modification is required for enhancement of activity. For this purpose, 3D-QSAR models are suitable for the prediction of the activity of the compounds based on their 3D grid points generated around the aligned set of molecules. One of the major limitations of the 3D-QSAR method is its dependency on molecular alignment and conformers chosen for the alignment. This becomes crucial when the information of confirmation is absent or when the molecular framework is not rigid. Hence, it is clearly understood that there is a requirement of a QSAR method which will allow flexibility to study molecular sites of interest and not depend upon conformational analysis and alignment of the molecules to provide information about sites and nature of interaction required for activity. Fragment-based G-QSAR has shown promise in current drug discovery and lead optimization. This method involves the calculation and evaluation of descriptors only for the substituent groups or molecular fragments instead of the whole molecule [16]. The biggest advantage of this method is it can be applied for both congeneric and non-congeneric series. G-QSAR method is helpful over 2D- and 3D-QSAR methods because this method provides useful information about the significant substitution sites, their chemical nature, and overall interaction that affects the therapeutic activity of the compounds [17, 18]. The focus of the present study was to develop fragment-based G-QSAR model correlating the biological activity of the molecular fragments with the 2D fragment-based descriptors and molecular docking study to interpret the structural requirement and the mechanism of the binding interaction between the ligand and the receptor on a congeneric series of 97 compounds taken from reported literature. The information derived from such predictive models could be utilized for accurate prediction of biological activities (dependent variables) of molecular fragments based on their chemical structure and properties (independent variables) that can be utilized as the

building blocks for designing of novel anti-HIV molecules before synthesis and experimental testing [19, 20]. The result of the docking study can improve the binding process of ligands with its receptor and provide insights into the structural features related to the activities of the new drug compounds.

2 Methods

2.1 Data set

To perform the present computational study, a data set of 97 compounds having reported IC_{50} values was taken from the literature [21]. The selected compounds for the study shared the same activity and assay procedure with significant variations in their structure and potency. Inhibitory potencies of the compounds in the data set have IC_{50} value ranges from 0.63 to 19.5 nm which were further converted to pIC_{50} by using the mathematical formula given as Eq. 1:

$$pIC_{50} = -\log_{10}(IC_{50}) \quad (1)$$

The chemical structure of the congeneric dataset was prepared by using MarvinSketch. The conversion of 2D structures to 3D structures and energy minimization of 3D compounds were performed by the help of force field batch minimization modules of the VLifeMDS software [22]. Energy minimization is performed to optimize the molecules up to their lowest stable state of energy. A template which is the representative of the entire molecules of the dataset under study was drawn with the presence of a dummy atom at the substitution sites. The template has 6 substitution sites depicted as R_1 – R_6 in Fig. 2. The study was performed by using VLife MDS version 4.6 from VLife Sciences Technologies Pvt Ltd, Pune, India.

2.2 Calculation of descriptors for G-QSAR modeling

This step is performed by using the G-QSAR module of VLife MDS. The common scaffold (Fig. 2) was used as a template for the generation of fragment-based G-QSAR models. The optimized molecules were imported into the QSAR worksheet, and their pIC_{50} values were incorporated manually, followed by the calculation of various physicochemical descriptors for the different substitution sites of the compounds [23]. A total of 321 physicochemical descriptors for classes like individual physicochemical descriptors like molecular weight, hydrogen bond donors and acceptors, retention index (χ), atomic valence connectivity index (χ_{iv}), chain path count, cluster, path cluster, kappa, estate numbers, estate contributors, and information theory index were calculated for the groups to present at the substitution site in each of 97 molecules presented in Table 1.

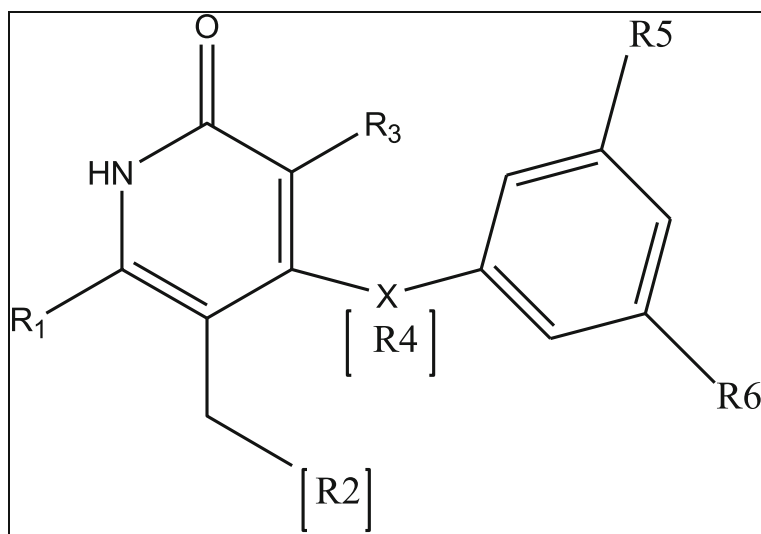


Fig. 2 Molecular template of 4-arylthio and 4-aryloxy-3-iodopyridine-2(1H)-one derivatives utilized for fragmentation pattern (R_1 – R_6 are substitution sites)

2.3 Data selection and building of G-QSAR model

The data set was divided into a training set of 75 compounds and a test set of 22 compounds based on Sphere Exclusion algorithms so that the activity of the selected test set is uniformly distributed throughout the activity column of the compounds [19]. Unicolumn statistics is performed for both training and test series to check the spread of data. The results of the unicolumn statistics study are presented in Table 2. From the result, the test set is interpolative, i.e., the activity of the test set is derived within the activity range of the training set. The mean and standard deviation of the training and test sets provide insight into the relative difference of mean and point density distribution of the two sets. The average value of the test set is higher than the training set shows the presence of relatively more active molecules as compared to the inactive ones.

For the building of G-QSAR models, a simulated annealing algorithm (SA) was utilized. After that, multiple G-QSAR models were generated using multiple linear regression (MLR), partial least squares regression (PLSR), and principal component regression (PCR).

2.4 Validation of the developed G-QSAR model

For validation of the developed G-QSAR models, the data set is divided into two sets as training and test sets. This division is based on the substitution groups and the inhibition of compounds. The training set is employed to produce the QSAR model, and the test set is used to validate the quality of the developed models [24]. The statistical parameters of the developed models and internal and external validations are adopted for testing the fitness, stability, and predictive ability of the QSAR

models [17]. The models are validated by considering many statistical parameters such as the number of compounds in regression (n), the number of variables (k), degree of freedom, squared correlation coefficient (r^2), cross-validated correlation coefficient (q^2), Fisher's value (F test), and r^2 for the external test set, (pred_r2) for external validation. For the internal predictive ability of the model, leave-one-out (LOO) method is used, showed as the value of q^2 (cross-validated explained variance). External validation of the developed QSAR models is performed by measuring the predictive power of the current models on the external test set by calculating the pred_r2 value as given in Eq. 2, which gives the statistical correlation between predicted and actual activities of the test set compounds

$$\text{pred}_r2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \quad (2)$$

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

Internal validation of the developed QSAR models is performed by calculating the q^2 value as given in Eq. 3, which gives the statistical correlation between predicted and actual activities of the training set compounds.

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \quad (3)$$

where y_i , \hat{y}_i , and y_{mean} are the actual, predicted activity of the i th molecule in the training set, and the average

Table 2 Unicolumn statistics of activity (pIC_{50}) for training and test set compounds for G-QSAR

Compounds	Average	Maximum	Minimum	Std. Dev	Sum
Training set	0.6399	1.2999	0.0969	0.3046	40.9567
Test set	0.8303	1.0000	0.0969	0.3019	15.3789

activity of all the molecules in the training set, respectively.

The generated models were considered having a significant predictivity when the squared correlation coefficient (r^2) between physiochemical descriptors (independent variable) and activity (dependent variable) was over 0.6. The developed model possesses significant internal and external predictivity when the cross-validated correlation coefficient of the leave-one-out method (q^2) > 0.6, the correlation coefficient of the training set ($pred_r2$) > 0.5, and higher value of F test.

2.5 Molecular docking study

A molecular docking study is a computational approach for searching a ligand that can fit both geometrically and energetically into the binding site of a target to show biological activity [25]. Docking study helps to predict drug/ligand or receptor/protein interactions by identifying the suitable active sites in the protein, getting the best geometry of ligand-receptor complex, and calculating the energy of interaction for different ligands to design more effective ligands with good binding affinity against RT enzyme [26].

2.5.1 Protein preparation

For protein preparation, three-dimensional crystallographic structures, and the coordinates of the target

protein (PDB-ID: 1S6G) having resolution 3.0 Å is retrieved from the RCSB PDB database (<https://www.rcsb.org>). The protocol for protein preparation was performed by deleting the bounded ligand, inserting missing atoms in incomplete residues, deleting alternate conformations, and modeling the missing loop regions with the help of Biopredicta (homology modeling) modules of the VLife-MDS software [27]. The final 3D structure of RT was evaluated using Biopredicta modules; Ramachandran plot showed that 87.95% of residues presented in most favored regions (Fig. 3). Before performing docking bond, orders of the ligands are assigned, hydrogen atoms are added, and the water molecules which do not involve in the interaction are deleted.

2.5.2 Protein-ligand docking

The molecular docking study was performed by using the Biopredicta tool of the V-Life MDS software version 4.6. The optimized and symmetrical crystalline protein structure got from Protein Preparation Wizard was used for docking study. The energy minimization of the crystal structure was carried out to relieve the steric clashes among the residues due to the addition of hydrogen atoms [28]. Crystallographic water molecules (water molecules without H bonds) were deleted. Hydrogen bonds corresponding to pH 6.8 were added considering the appropriate ionization states for both the acidic and basic amino acid residues. A grid box was generated at the centroid of the active site for docking, and the active site was defined with a 10 Å radius around the ligand present in the crystal structure [29]. To test the docking parameters, all compounds were docked into the catalytic pocket of the RT enzyme (PDB-ID: 1S6Q). Finally, the best-docked structure was selected using the dock score function [30].

3 Result

3.1 Development of G-QSAR model

In the present work, effective G-QSAR model for 97 molecules belonging to the congeneric series with viability in the biological activity is generated by calculating fragment-based molecular descriptors. By using the VLife MDS software, a total of 984 descriptors are calculated, and after removal of invariable columns, a total of 321 descriptors are used to generate G-QSAR models. The sphere exclusion method with a dissimilarity value of + 1 resulted in a training set of 75 and a test set of 22 compounds. The test set was selected as maximum pIC_{50} value of the test set compounds if it was less or equal to that of the training set and the lowest pIC_{50} value of the test set compound was more than or equal to that of the training set so that the test set has been derived from the maximum-minimum range of training set. Multiple G-QSAR models were built using a simulated annealing algorithm (SA) coupled with multiple

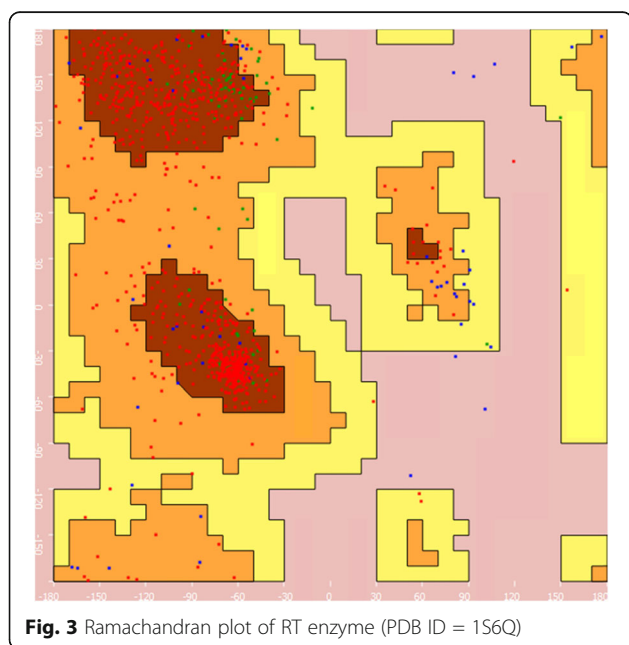
**Fig. 3** Ramachandran plot of RT enzyme (PDB ID = 1S6Q)

Table 3 Statistical parameters of the developed G-QSAR models

Parameter	Model-I (SA-MLR)	Model-II (SA-PLS)	Model-III (SA-PCR)
<i>n</i> (training/test)	75/22	75/22	75/22
Degree of freedom	83.59	89.35	80.67
r^2	0.7311	0.8524	0.7124
q^2	0.6874	0.6925	0.6421
<i>F</i> _test	44.0779	55.0374	43.4689
r^2 _se	0.5214	0.5778	0.5647
q^2 _se	0.4056	0.4265	0.4136
pred_ r^2	0.7325	0.7421	0.7025
pred_ r^2 _se	0.3445	0.3078	0.3487

r^2 squared correlation coefficient, q^2 cross-validated correlation coefficient of leave-one-out method (internal validation), *F* test Fisher test, *Pred_* r^2 correlation coefficient of the training set (external validation), *SE* standard error

linear regression (MLR), partial least squares regression (PLS), and principal component regression (PCR). Several models were generated, and the best three out of them are selected based on the statistical values like r^2 , q^2 , pred_ r^2 , *F* test, and standard error. The statistical parameters of each model are shown in Table 3.

The fitness plot between actual and predicted activity for training and test set compounds provides an idea about how well these models are trained and how well they predict the activity of the external test set. Further, the distribution curve of actual and predicted activity for training and test set compounds of the developed models is depicting closeness between the actual and predicted activity of the compounds for training and test sets. Descriptors like SK-hydrophobic area, SaaN-count, Chi-3-cluster, SsCHE-index, SaaNE-index, Delta epsilon-C, SsBr count, SaaNH count, smr, SssScount,

SaaOcount, SsCH3count, 1-path count, Id-average, Chi v3, and Chi v2 are showing contribution on the respective substitution sites.

3.1.1 Interpretation of model-I (SA-MLR)

The G-QSAR model-I was obtained by using a simulated annealing algorithm (SA) coupled with multiple linear regressions (MLR). The correlation equations between activity (pIC₅₀) and the selected parameters are expressed by Eq. 4;

$$\begin{aligned} \text{pIC}_{50} = & 0.4685 + 0.5031(\pm 0.0023)\text{R}_2\text{SKHydrophobicArea} - 0.2021(\pm 0.0757) \\ & \text{R}_2\text{SsSHE} - \text{index} - 0.0381(\pm 0.0090)\text{R}_2\text{SaaNE} - \text{index} \\ & + 0.4132(\pm 0.0003)\text{R}_4\text{H} - \text{acceptor Count} + 0.3217(\pm 0.0020)\text{R}_3\text{XlogP} \end{aligned} \quad (4)$$

The generated model-I was statistically significant with predictivity, $r^2 = 73.11\%$, internal (q^2), and external

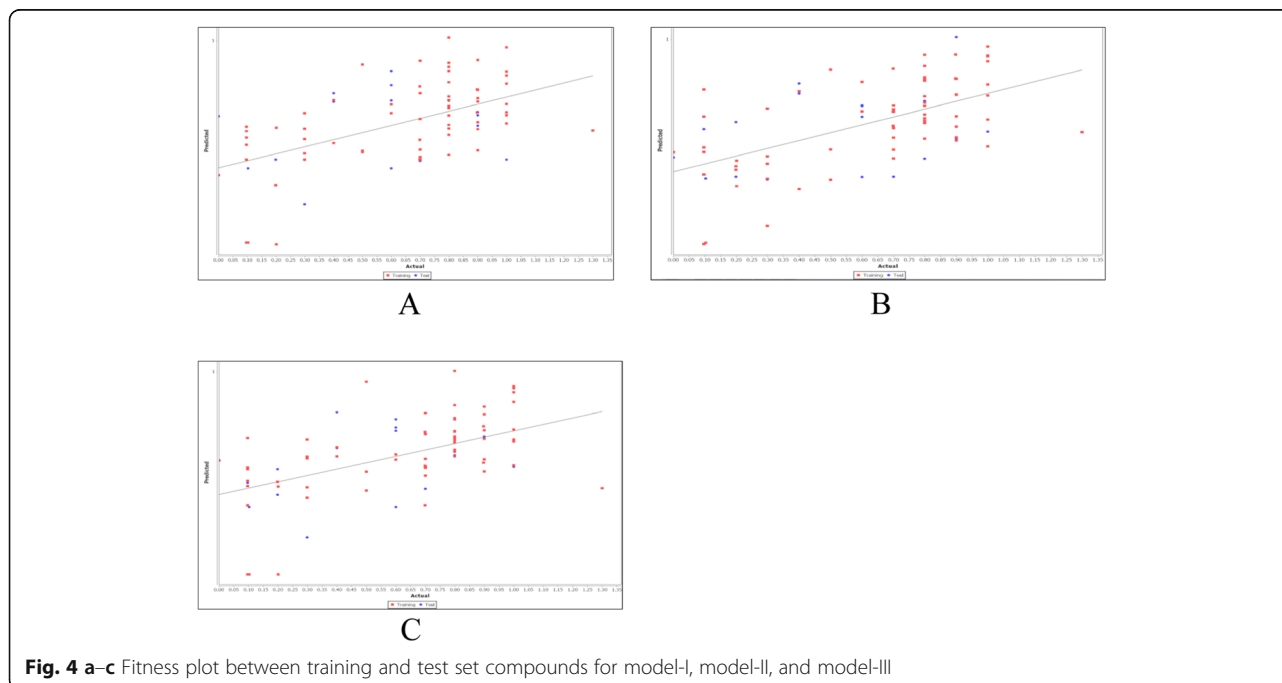


Fig. 4 a-c Fitness plot between training and test set compounds for model-I, model-II, and model-III

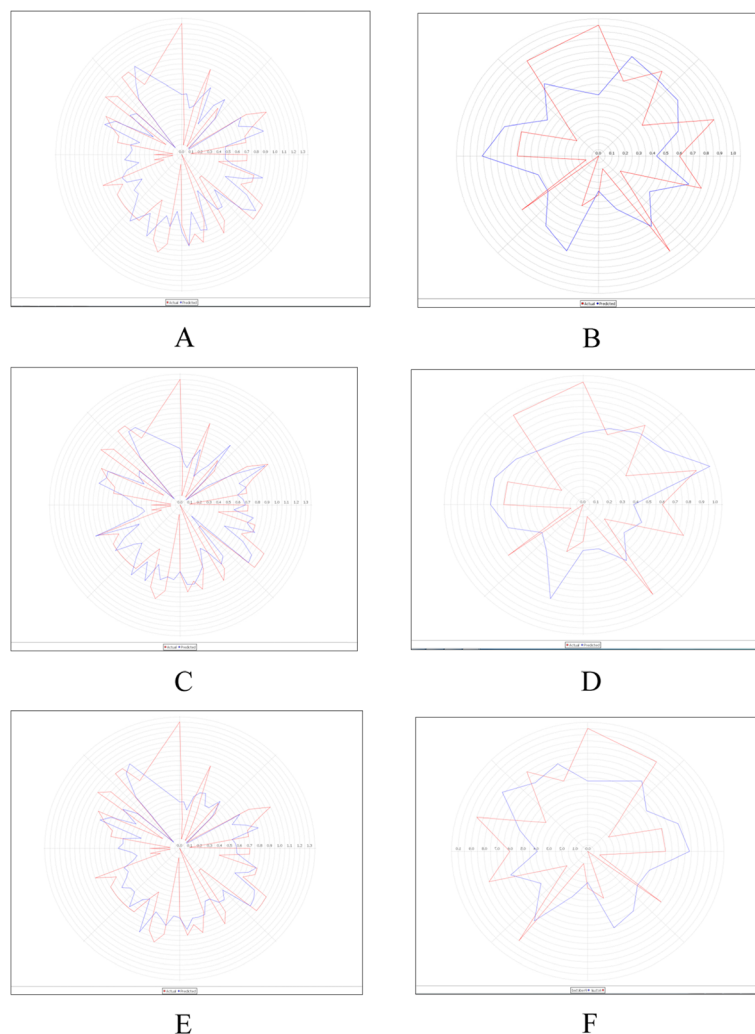


Fig. 5 a-f Actual vs predicted activity of training and test set compounds for model-I, model-II, and model-III

(pred_r²) validation revealed a predictive power of 68.74% and 73.25% respectively. The fitness plot between actual and predicted activity for training and test set compounds is given in Fig. 4a, which provides an idea about the predictivity of the model for training and test set compounds. Further, the distribution curve of actual and predicted activity for training and test set compounds for model-I are presented in Fig. 5a, b, depicting closeness between the actual and predicted activity of the compounds for training and test set. The contribution of different physiochemical descriptors towards activity is shown in Fig. 6a.

3.1.2 Interpretation of model-II (SA-PLS)

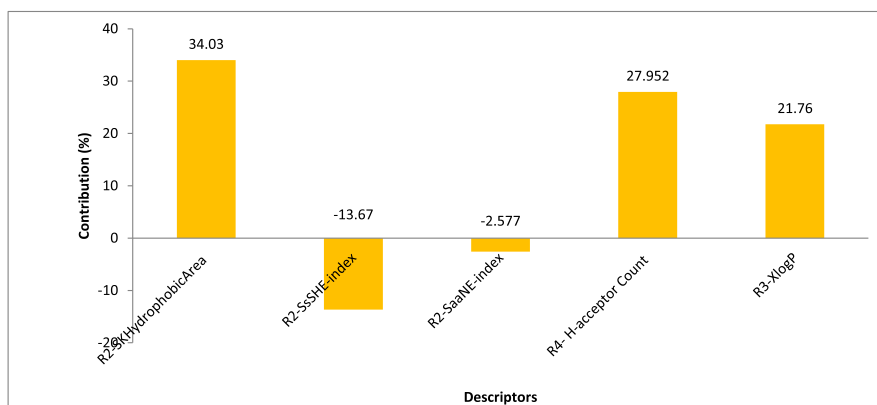
Model-II obtained by simulated annealing algorithm (SA) associated with partial least square regression (PLS) expressed as Eq. 5 explains an improved correlation coefficient of $r^2 = 85.24\%$, internal (q^2) and external (pred_r²)

validation predictive power of 69.25% and 74.21% respectively. The higher degree of freedom (89.35), F test (55.037), and low standard error (pred_r²se = 0.3078) value support the robustness of the model.

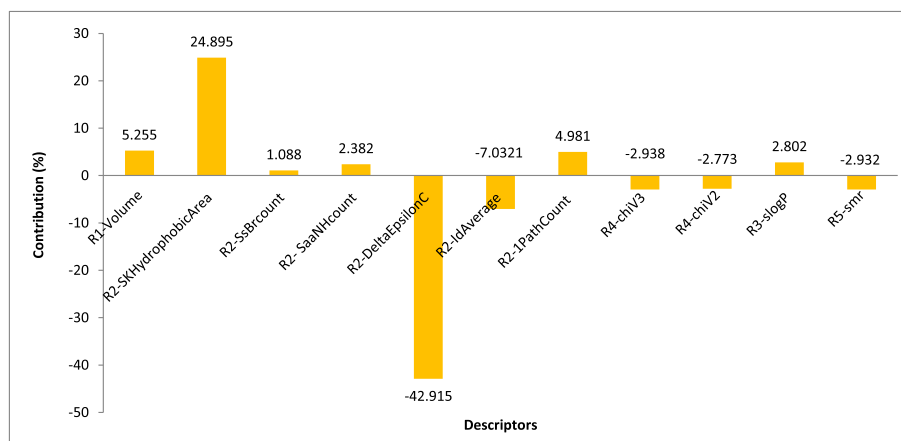
$$\begin{aligned}
 \text{pIC}_{50} = & 0.4227 - 0.2014(\pm 0.0017)\text{R}_1 \text{Volume} + 0.9541(\pm 0.0124) \\
 & \text{R}_2 \text{SKHydrophobicArea} + 0.0417(\pm 0.0033) \\
 & \text{R}_2 \text{SsBr count} + 0.0913(\pm 0.0002)\text{R}_2 \text{SaaNHcount} - 1.6448 \\
 & (\pm 0.0051)\text{R}_2 \text{DeltaEpsilonC} - 0.2695(\pm 0.0104)\text{R}_2 \text{IdAverage} + 0.1909 \\
 & (\pm 0.0057)\text{R}_2 \text{IPathCount} - 0.1126(\pm 0.0033) \\
 & \text{R}_4 \text{chiV3} - 0.1063(\pm 0.0033)\text{R}_4 \text{chiV2} + 0.1074(\pm 0.0001) \\
 & \text{R}_3 \text{slogP} - 0.1124(\pm 0.0048)\text{R}_5 \text{smr}
 \end{aligned}$$

(5)

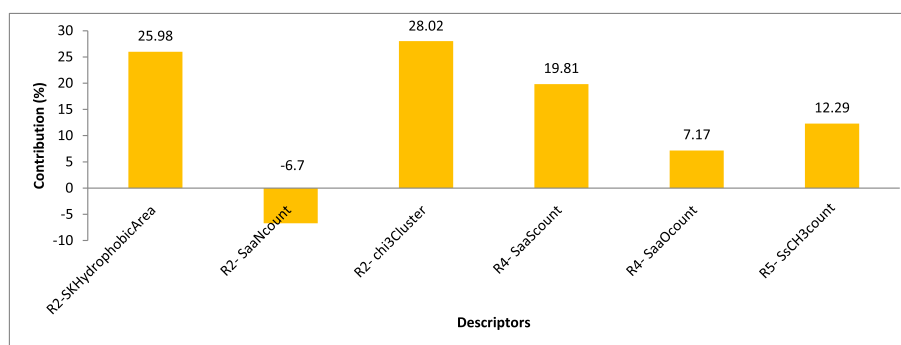
The fitness plot and the distribution curve of actual and predictive activity for training and test sets are given as Fig. 4b, and Fig. 5 c and d show that the minimum difference between the actual and



A



B



C

Fig. 6 a–c Contribution plot of descriptors towards pIC_{50} for model-I, model-II and model-III

predicted values of the compounds is a measure of the high quality of the model. The contribution curve of descriptors towards activity is presented as Fig. 6b.

3.1.3 Interpretation of model-III (SA-PCR)

The third model, model-III, was obtained by a simulated annealing algorithm (SA) associated with principal component regression (PCR) expressed as Eq. 6

explaining the biological activity as a function of some physicochemical descriptors like SKHydrophobicArea, SaaNcount, Chi3Cluster, SsCH3count, SaaScount, and SaaOcount at their respective fragmented substitution sites. The generated model has correlation coefficient $r^2 = 71.24\%$, internal (q^2) and external (pred_r2) predictive ability of 64.21% and 70.25% respectively.

Table 4 Docking score of compounds (Kcal/mol)

Compd. No	Dock score
1	- 11.225
2	- 9.508
3	- 10.783
4	- 9.531
5	- 13.039
6	- 10.816
7	- 10.747
8	- 8.66
9	- 9.206
10	- 11.714
11	- 12.655
12	- 10.574
13	- 10.958
14	- 10.385
15	- 11.348
16	- 10.397
17	- 9.356
18	- 11.127
19	- 10.421
20	- 10.286
21	- 10.22
22	- 9.728
23	- 10.799
24	- 13.278
25	- 10.811
26	- 9.397
27	- 9.994
28	- 11.258
29	- 10.716
30	- 8.001
31	- 8.857
32	- 7.659
33	- 10.126
34	- 8.39
35	- 10.129
36	- 7.847
37	- 12.319
38	- 10.922
39	- 8.296
40	- 8.561
41	- 10.199
42	- 9.321
43	- 9.074
44	- 8.775
45	- 10.248
46	- 9.961
47	- 10.951
48	- 9.515
49	- 10.232
50	- 10.768
51	- 12.478
52	- 11.422

Table 4 Docking score of compounds (Kcal/mol) (Continued)

Compd. No	Dock score
53	- 12.16
54	- 13.076
55	- 10.493
56	- 12.954
57	- 12.267
58	- 10.142
59	- 12.108
60	- 13.291
61	- 13.168
62	- 12.009
63	- 9.789
64	- 11.144
65	- 11.405
66	- 10.775
67	- 11.256
68	- 8.884
69	- 11.823
70	- 10.539
71	- 11.094
72	- 11.759
73	- 12.49
74	- 8.572
75	- 9.5
76	- 8.461
77	- 8.853
78	- 8.689
79	- 10.082
80	- 9.991
81	- 9.307
82	- 10.841
83	- 11.089
84	- 10.698
85	- 11.08
86	- 9.386
87	- 11.943
88	- 11.006
89	- 10.774
90	- 11.015
91	- 10.689
92	- 10.26
93	- 10.657
94	- 10.95
95	- 10.534
96	- 11.266
97	- 11.768

$$\begin{aligned}
 \text{pIC}_{50} = & 0.3998 + 0.4215(\pm 0.0014)R_2 - \text{SKHydrophobicArea} + 0.0088(\pm 0.0002) \\
 & R_2 - \text{SaaNcount} + 0.4545(\pm 0.0007)R_2 - \text{chi3Cluster} + 0.9244(\pm 0.0033)R_4 - \text{SaaScount} + 0.1163(\pm 0.0012) \\
 & R_4 - \text{SaaOcount} + 0.1995(\pm 0.0048)R_5 - \text{SsCH}_3\text{count}
 \end{aligned}
 \tag{6}$$

The fitness plot and the distribution curve of actual and predictive activity for training and test set (Fig. 4c and Fig. 5e, f) shows a correlation between actual and predicted activity. Contribution curve of descriptors towards activity is shown in Fig. 6c.

3.2 Molecular docking

The docking studies were carried out for 97 data set compounds into the catalytic pocket of the prepared target RT enzyme (PDB-ID: 1S6Q) by the V-Life MDS software. This study is useful to identify the binding potency and poses of active molecules that reveal the molecular mechanism of action. The compounds during docking showed several poses, orientation, and configurations. Each configuration is characterized by a combined score of van der Waal's forces, hydrogen bonding, pi interaction, charge interaction, halogen bond interaction, and salt bridge interaction. The docking scores of the listed compounds are presented in Table 4. The docking study revealed that interactions were dominated by the hydrophobicity or π -aromaticity due to the presence of aromatic and hetero atomic rings significant for stacking interactions. The docking interaction result of compound 51 given in Table 5 reveals that the interactions were dominated in the region of LYS-374, GLN-572, TRP-573, PRO-574, VAL-609, PHE-610, GLU-945, and GLU-948 amino acid residues due to the presence of the active site in the region (Fig. 7a). The 2D-docked pose of compound 51 in the active site of the target receptor is shown in Fig. 7b.

4 Discussion

All three models developed during the present study are more significant and found to have good predictivity. Model-I developed by SA-MLR is statistically significant and shows good predictivity for both training and test set of compounds. The contribution plot of descriptors towards pIC_{50} for model-I shows the positive contribution of SKHydrophobic area (34.03%), indicates the substitution of the higher hydrophobic group like substituted aromatic or heterocyclic rings at R2 increase activity, while the negative contribution of SssSHE-index (-13.67%) and SaaNE-index (-2.57%) at R2 position shows -NH₂ connected to one bond and -SH group connected with over one single bond are conducive for anti-HIV activity. The positive contribution of H-acceptor count (27.95%) at R4 and XlogP (21.76%) at R3 indicates the presence of H-bond forming elements like oxygen and sulfur at R4, and

higher hydrophobic group at R3 increases drug activity by increasing its binding with the receptor. G-QSAR equation (model-II) developed by using SA-PLC algorithm shows improved statistical study results for both training and test set of compounds. The contribution plot shows the positive contribution of molecular weight (5.25%) at R1 which indicates that the substitution of the bulkier group at R1 increases drug activity. The positive contribution of descriptors like SKHydrophobic area (24.89%), SsBr count (1.08%), SaaNHcount (2.38%), and 1Pathcount (4.98%) at R2 shows substitution of a higher hydrophobic group containing more -Br atom connected with one bond, the total number of -NH group connected with 2 aromatic bonds, and fragment R2 of first-order increase anti-HIV activity. Descriptors like DeltaEpsilonC (-42.91%) and IdAverage (-7.03%) showing significant negative contribution towards activity indicate a decrease in the contribution of electronegativity and entropic interaction fields of the fragment at R2 increase biological activity. The positive contribution of slogP (2.80%) at R3 suggesting the presence of the lipophilic group at R3 enhances inhibitory activity. Atomic valence connectivity index ChiV3 (-2.93%) and ChiV2 (-2.77%) at R4 contributes negatively towards activity indicates the atomic connectivity index of fragment R4 of order 3 and 2 decrease activity. The descriptor smr (-2.93%) at R5 deleterious towards activity shows the presence of a group having low molecular refractivity increase enzyme inhibition. Model-III developed by SA-PCR is also significant in predicting the role of descriptors towards activity for both training and test set. The contribution plot for descriptors and pIC_{50} shows a positive contribution of SKHydrophobicArea (25.98%) and Chi3Cluster (28.02%), and negative contribution of descriptor SaaNcount (-6.7%) at R2 means the presence of the group having higher hydrophobicity, 3rd order cluster chi index, and total nitrogen not connected with two aromatic bonds increase anti-HIV activity. The positive contribution of Subaccount (19.81%) and SaaOcount (7.17%) at R4 indicates the presence of oxygen and sulfur connected with two aromatic bonds at this site increases enzyme inhibition activity. Descriptor SsCH₃count (12.29%) at R5 contribute positively towards activity, indicates the presence of -CH₃ group increases drug activity by increasing its binding efficiency with the target site of the receptor. Molecular docking was performed for these compounds present in the data set with the active site of the target RT enzyme to derive the ligand-receptor interaction mechanism. The 2D-dock pose of compound 51 evident that the generated map of hydrophobic and hydrophilic fields, where benzene ring and heteroatomic (Pyridine and furan) rings are present in the chemical structure, is buried in the hydrophobic pocket. Further docking analysis shows the amide group of fragment R2 of the ligand forms a hydrogen bond between amino (-NH₂)

Table 5 Docking interaction of compound 51 with the binding pocket of 1S6Q

Amino acid residue	Ligand atoms	Distance (Å)	Interaction type
GLN572A	26C	3.708	Vander Waal's interaction
	27O	3.664	
	22C	3.461	
	24C	3.448	
	46H	3.354	
	10C	3.776	
	48H	3.193	
	49H	3.158	
	47H	2.835	
	50H	2.773	
	21C	3.269	
	22C	2.918	
	24C	2.916	
	25C	3.277	
	28 N	2.931	
	26C	3.234	
	28 N	3.135	
	28 N	2.889	
	53H	2.46	
	28 N	3.079	
TRP573A	20 N	3.559	Vander Waal's interaction
	21C	3.543	
	25C	3.552	
	26C	3.649	
	27O	3.496	
	51H	3.003	
	3C	3.568	
	42H	3.016	
	43H	3.021	
	45H	2.97	
	19C	3.622	
	20 N	3.473	
	51H	3.251	
	2C	3.195	
	10C	3.261	
	22C	3.378	
	24C	3.387	
	30H	2.697	
	44H	2.761	
	52H	2.816	
3C	3.814		
19C	3.542		
3C	3.8		
4C	3.783		

Table 5 Docking interaction of compound 51 with the binding pocket of 1S6Q (Continued)

Amino acid residue	Ligand atoms	Distance (Å)	Interaction type
	9O	3.364	
	42H	3.094	
	43H	3.09	
	9O	3.508	
	11C	3.837	
	12C	3.616	
	33H	3.361	
	9O	3.522	
	11C	3.586	
	12C	3.456	
	33H	3.007	
	9O	3.272	
	11C	3.647	
	12C	3.591	
	33H	2.942	
PRO608A	28 N	3.591	Vander Waal's interaction
	53H	2.703	
	28 N	3.624	
	53H	2.915	
	26C	3.163	
	28 N	2.833	
VAL609A	26C	3.7	Vander Waal's interaction
	54H	2.815	
	23C	3.684	
	26C	3.758	
	24C	3.233	
	46H	2.962	
	47H	2.964	
	50H	2.957	
PHE610A	28 N	3.406	Vander Waal's interaction
	49H	2.981	
	49H	2.957	
	54H	2.905	
	49H	2.767	
	22C	3.815	
	23C	3.543	
	24C	3.878	
	28 N	3.584	
	47H	3.188	
	50H	3.229	
	26C	3.303	
	46H	2.817	
LYS622A	54H	2.938	Vander Waal's interaction
	28 N	2.978	

Table 5 Docking interaction of compound 51 with the binding pocket of 1S6Q (Continued)

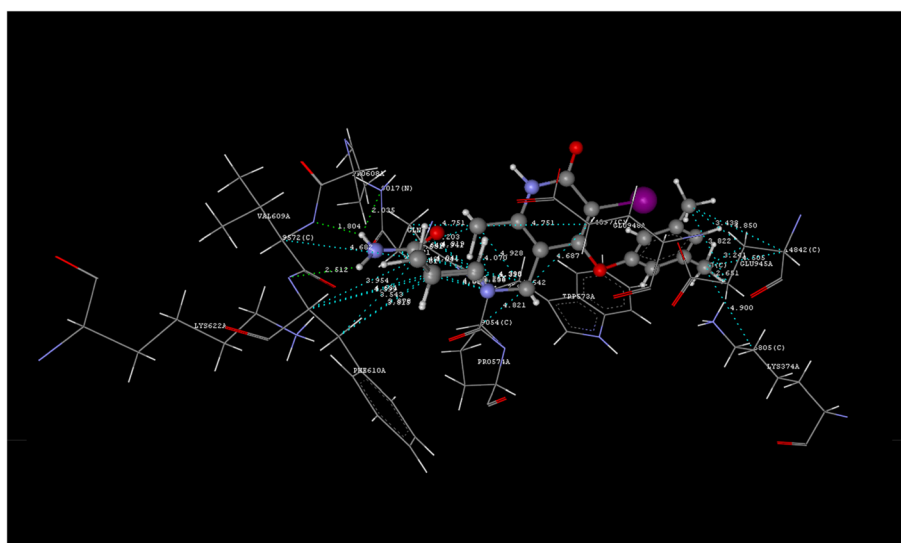
Amino acid residue	Ligand atoms	Distance (Å)	Interaction type
GLU945A	54H	2.881	Vander Waal's interaction
	17C	3.438	
	34H	3.143	
	37H	3.293	
	39H	3.08	
	11C	3.876	
	16C	3.783	
	17C	3.822	
	34H	3.355	
	39H	2.981	
	9O	3.45	
	34H	2.928	
	35H	2.987	
	5C	3.546	
	9O	3.144	
	40H	3.065	
	9O	3.085	
	39H	3.2	
	40H	2.82	
	13C	2.967	
15C	2.96		
37H	2.748		
38H	2.585		
39H	2.414		
14C	3.336		
15C	3.336		
17C	3.396		
13C	3.029		
14C	3.275		
40H	2.424		
41H	2.598		
GLU948A	1 N	3.266	Vander Waal's interaction
	29H	3.211	
	31H	3.176	
	32H	3.18	
	3C	3.382	
	4C	3.178	
	5C	3.102	
6C	3.216		
LYS374A	18C	4.9	Hydrophobic interaction
GLN572A	10C	4.751	Hydrophobic interaction
	21C	4.941	
	22C	4.681	
	24C	4.649	

Table 5 Docking interaction of compound 51 with the binding pocket of 1S6Q (Continued)

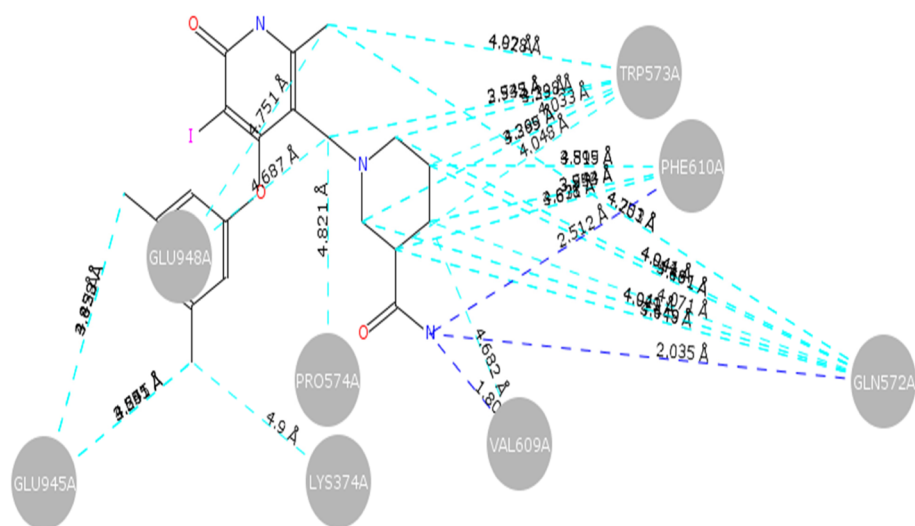
Amino acid residue	Ligand atoms	Distance (Å)	Interaction type
	25C	4.919	
	10C	4.203	
	21C	4.044	
	22C	3.461	
	23C	4.071	
	24C	3.448	
	25C	4.041	
TRP573A	10C	4.07	Hydrophobic interaction
	19C	2.935	
	21C	3.238	
	22C	4.033	
	24C	4.048	
	25C	3.265	
	10C	4.928	
	19C	3.542	
	21C	4.39	
	25C	4.399	
PRO574A	19C	4.821	Hydrophobic interaction
VAL609A	23C	4.682	Hydrophobic interaction
PHE610A	22C	4.599	Hydrophobic interaction
	23C	3.954	
	24C	4.631	
	22C	3.815	
	23C	3.543	
	24C	3.878	
GLU945A	17C	4.85	Hydrophobic interaction
	18C	4.505	
	17C	3.438	
	18C	3.241	
GLU945A	17C	3.822	Hydrophobic interaction
	18C	2.651	
GLU948A	10C	4.751	Hydrophobic interaction
	19C	4.687	
TRP573A	11C	4.553	Aromatic interaction
	11C	4.347	
TRP573A	20 N	3.559	Charge interaction
GLN572A	53H	2.035	H-bond interaction
VAL609A	53H	1.804	H-bond interaction
PHE610A	54H	2.512	H-bond interaction

and acidic group (–COOH) of GLN-572, VAL-609, and PHE-610 amino acid residue present at the active site of the receptor. Hydrophobic interactions were observed between groups present at R1, R2, R4, R5, and R6 with LYS-374, GLN-572, TRP-573, PRO-574, VAL-609, PHE-610,

GLU-945, and GLU-948 amino acids of receptor to form a stabilized complex suggesting a strong binding of inhibitors with the RT enzyme. The residue such as GLN-572, TRP-573, PRO-608, VAL-609, PHE-610, LYS-622, GLU-945, and GLU-948 was involved in the van der Waal's



a



b

Fig. 7 a Docking pose of compound 51 into active site of 1S6Q. b 2D-ligand interaction diagram of compound 51 in the binding pocket of 1S6Q

interaction, and amino acid TRP-573 involved in both aromatic and charged interaction with the ligand molecules increase the stabilization of inhibitor at the active site of RT enzyme. The result of the G-QSAR and molecular docking study provided a molecular level understanding to infer that identified compounds are promiscuous and might be a potential inhibitor of RT enzyme.

5 Conclusion

In the present study, an attempt was made to generate novel fragment-based QSAR (G-QSAR) models for a congeneric series of 97, 4-arylthio, and 4-aryloxy-3-iodopyridine-2(1H)-one derivative with known anti-HIV

activity. The whole data set of compounds were divided into training and test sets, and three G-QSAR models were developed by a simulated annealing algorithm (SA) coupled with multiple linear regression (MLR), partial least squares regression (PLS), and principal component regression (PCR). All generated models I, II, and III were statistically significant and provide site-specific clues for the design of new reverse transcriptase inhibitors. Model-II developed by SA-PLS was more significant among all; the result of the statistical parameters were $r^2 = 85.24\%$, $q^2 = 69.25\%$, $\text{pred}_r2 = 74.21\%$, the higher degree of freedom (89.35), F test (55.037), and low standard error ($\text{pred}_r2\text{se} = 0.3078$) values during validation

study for both training and test set compounds fulfilled the conditions for predictive and robustness. This model indicates the presence of higher hydrophobic substituent containing single-bonded –Br atom, 2 aromatic bonded –NH group with less electronegativity, and entropic interaction fields at fragment R2 essential for better anti-HIV activity. Similarly, the presence of a lipophilic group at R3, oxygen, and sulfur connected with two aromatic bonds at R4 and –CH₃ group at R5 increases inhibition activity by increasing binding efficiency with reverse transcriptase enzyme. G-QSAR method allows ease to interpretation, unlike conventional QSAR method which could only suggest important descriptors but does not reflect the site where it has to be optimized for further design of new compounds. Molecular docking results of the G-QSAR-generated compounds showed the interactions between the fragmented groups reported for these compounds and the residues located at the binding site. Dock pose of the selected compound reveals that the presence of a group at fragmented site R1, R4, R5, and R6 are responsible for hydrophobic interaction, and group at R2 is essential for both H-bond and hydrophobic interaction with the amino acid residues at the active site of the target enzyme. The findings got from G-QSAR and docking studies were utilized for designing newer RT-inhibitor anti-HIV agents. It is therefore concluded that the molecular manipulations at appropriate sites suggested by structure-activity relationship data will prove beneficial for identifying particular chemical variation at specific substitution sites and mathematical models for prediction of biological activity of newly designed molecules.

Abbreviations

G-QSAR: Group-based quantitative structure activity relationship; SA: Simulated annealing; MLR: Multiple linear regression; PLS: Partial least square; PCR: Principle component regression; RMSD: Root mean square deviation; RMSE: Root mean square error; SD: Standard deviation; r^2 : Coefficient of determination; q^2 : Cross-validated squared correlation coefficient; pred_r2 : External predictivity; F_{test} : Fisher test

Acknowledgements

The authors are thankful to V-Life Science Technologies Pvt. Ltd for providing the software for the study.

Authors' contributions

SKS and DP designed the research work. DP performed the whole research work, including computational studies. SKS and AM analyzed the data. All authors contributed equally to writing the paper. The authors read and approved the final manuscript

Funding

The present research work was not funded by any funding agencies.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Drug Research Laboratory, Nodal Research Centre, College of Pharmaceutical Sciences, Puri, Baliguali, Puri- Konark Marine Drive road, Puri, Odisha 752002, India. ²Department of Pharmacy, Utkal University, VaniVihar, Bhubaneswar, Odisha 751004, India.

Received: 4 June 2020 Accepted: 2 September 2020

Published online: 30 November 2020

References

1. Jegede O, Babu J, Di Santo R, McColl DJ, Weber J, Quinones-Mateu M (2008) HIV type 1 integrase inhibitors: from basic research to clinical Implications. *AIDS Rev* 10(3):172–189
2. Fajardo-Ortiz D, Lopez-Cervantes M, Duran L, Dumontier M, Lara M, Ochoa H, Castano VM (2017) The emergence and evolution of the research fronts in HIV/AIDS research. *PLoS One* 12(5). <https://doi.org/10.1371/journal.pone.0178293>
3. Wainberg MA, Jeang KT (2008) 25 years of HIV-1 research – progress and perspectives. *BMC Med* 6:31–37
4. Sweileh WM (2018) Global research output on HIV/AIDS-related medication adherence from 1980 to 2017. *BMC Health Serv Res* 18(1):765–777
5. Safadi YE, Vivet-Boudou V, Marquet R (2007) HIV-1 reverse transcriptase inhibitors. *Appl Microbiol Biotechnol* 75(4):723–737
6. Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E (2008) Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* 385(3):693–713
7. La J, Latham CF, Tinetti RN, Johnson A, Tyssen D, Huber KD, Sluis-Cremer N, Simpson JS, Headey SJ, Chalmers DK, Tachedjian G (2015) Identification of mechanistically distinct inhibitors of HIV-1 reverse transcriptase through fragment screening. *Proc Natl Acad Sci* 112(22):6979–6984
8. Cory TJ, Midde NM, Rao PSS, Kumar S (2015) Investigational reverse transcriptase inhibitors for the treatment of HIV. *Expert Opin Investig Drugs* 24(9):1219–1228
9. Asahchop EL, Wainberg MA, Sloan RD, Tremblay CL (2012) Antiviral drug resistance and the need for development of new HIV-1 reverse transcriptase inhibitors. *Antimicrob Agents Chemother* 56(10):5000–5008
10. Chinsambu KC (2019) Chemical diversity and activity profiles of HIV-1 reverse transcriptase inhibitors from plants. *Rev Bras Farmacogn* 29(4):504–528
11. Bethell RC, Lie YS, Parkin NT (2005) In vitro activity of SPD754, a new deoxycytidine nucleoside reverse transcriptase inhibitor (NRTI), against 215 HIV-1 isolates resistant to other NRTIs. *Antivir Chem Chemother* 16(5):295–302
12. Mehellou Y, Clercq ED (2010) Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go. *J Med Chem* 53(2):521–538
13. Tsibris AMN, Hirsch MS (2010) Antiretroviral therapy in the clinic. *J Virol* 84(11):5458–5464
14. Correa-Basurto J, Bello M, Rosales-Hernandez MC, Hernandez-Rodríguez M, Vazquez IN, Rojo-Domínguez A, Flores-Sandoval CA (2014) QSAR, docking, dynamic simulation and quantum mechanics studies to explore the recognition properties of cholinesterase binding sites. *Chem Biol Interact* 209:1–13
15. Wang Z, Cheng L, Kai Z, Wu F, Liu Z, Cai M (2014) Molecular modeling studies of atorvastatin analogues as HMGR inhibitors using 3D-QSAR, molecular docking and molecular dynamics simulations. *Bioorg Med Chem Lett* 24(16):3869–3876
16. Singh A, Goyal S, Jamal S, Subramani B, Das M, Admane N, Grover A (2016) Computational identification of novel piperidine derivatives as potential HDM2 inhibitors designed by fragment-based QSAR, molecular docking and molecular dynamics simulations. *Structural Chemistry* 27:993–1003
17. Ajmani S, Jadhav K, Kulkarni SA (2009) Group-based QSAR (G-QSAR): mitigating interpretation challenges in QSAR. *QSAR Comb Sci* 28(1):36–51
18. Joshi K, Goyal S, Grover S, Jamal S, Singh A, Dhar P, Grover A (2016) Novel group-based QSAR and combinatorial design of CK-1δ inhibitors as neuroprotective agents. *BMC Bioinformatics* 17(19):515–527

19. Abdullahi AD, Abdulkader AM, Abdulsamat N, Ingale K (2015) Application of group-based QSAR and molecular docking in the design of insulin-like growth factor antagonist. *Trop J Pharm Res* 14(6):941–951
20. Kale MA, Baheti KG (2015) G-QSAR studies of novel 1,2,4-triazole [3,4-b]-1,3,4-thiadiazole derivatives. *Der Pharmacia Sinica* 6(9):15–20
21. Guillemont J, Benjahad A, Oumouch S, Decrane L, Palandjian P, Vernier D, Queguiner L, Andries K, Bethune M, Hertogs K, Grierson DS, Nguyen CH (2009) Synthesis and biological evaluation of C-5 methyl substituted 4-arylthio and 4-aryloxy-3-iodopyridin-2(1H)-one type anti-HIV agents. *J Med Chem* 52(23):7473–7487
22. VLife MDS 4.6 Molecular design suite (2018). Vlife Sciences Technologies Pvt. Ltd. Pune.
23. Choudhari P, Kumbhar S, Phalle S, Choudhari S, Desai S, Khare S, Jadhav S (2016) Application of group-based QSAR on 2-thioxo-4-thiazolidinone for development of potent anti-diabetic compounds. *Molecular Structure* 1128:355–360
24. Kale MA, Sonwane GM (2020) Molecular docking, G-QSAR studies, synthesis and anticancer screening of some new 2-phenazinamines as Bcr-Abl tyrosine kinase inhibitors. *Curr Drug Discov Technol* 17(2). <https://doi.org/10.2174/1570163815666180913122542>
25. Behera DK, Behera PM, Acharya L, Dixit A (2017) Pharmacophore modelling, virtual screening and molecular docking studies on PLD1 inhibitors. *SAR QSAR Environ Res* 28:991–1009
26. Safarizadeh H, Garkani-Nejad Z (2019) Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: insight into mechanism of interactions and parameters for design of new inhibitors. *J Mol Graph Model* 87:129–143
27. Panigrahi D, Mishra A, Sahu SK (2020) Pharmacophore modeling, QSAR study, molecular docking and insilico ADME prediction of 1,2,3-triazole and pyrazolopyridones as DprE1 inhibitor antitubercular agents. *SN App Sci* 2(5). <https://doi.org/10.1007/s42452-020-2638-y>
28. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput* 6:1509–1519
29. Tang HJ, Yang L, LiJH CJ (2016) Molecular modelling studies of 3, 5-dipyridyl-1, 2, 4-triazole derivatives as xanthine oxidoreductase inhibitors using 3D-QSAR, Topomer CoMFA, molecular docking and molecular dynamic simulations. *J Taiwan Inst Chem Eng* 68:64–73
30. Athar M, Lone MY, Khedkar VM, Jha PC (2016) Pharmacophore model prediction, 3D-QSAR and molecular docking studies on vinyl sulfones targeting Nrf2-mediated gene transcription intended for anti-Parkinson drug design. *J Biomol Struct Dyn* 34:1282–1297

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
