Structural requirements for potential HIV-integrase inhibitors identified using pharmacophore-based virtual screening and molecular dynamics studies†

Md Ataul Islam[a], Tahir S. Pillay[*a,b]

Acquired immunodeficiency syndrome (AIDS) is a life-threatening disease which is a collection of symptoms and infections caused by a retrovirus, human immunodeficiency virus (HIV). There is currently no curative treatment and therapy is reliant on the use of existing anti-retroviral drugs. Pharmacoinformatics approaches have already proven their pivotal role in the pharmaceutical industry for lead identification and optimization. In the current study, we analysed the binding preferences and inhibitory activity of HIV-integrase inhibitors using pharmacoinformatics. A set of 30 compounds were selected as the training set of a total 540 molecules for pharmacophore model generation. The final model was validated by statistical parameters and further used for virtual screening. The best mapped model (R = 0.940, *rmsd* = 2.847, Q^2 = 0.912, *se* = 0.498, R^2_{pred} = 0.847 and $r^2_{m\ (test)}$ = 0.636) explained that two hydrogen bond acceptor and one aromatic ring features were crucial for the inhibition of HIV-integrase. From virtual screening, initial hits were sorted using a number of parameters and finally two compounds were proposed as promising HIV-integrase inhibitors. Drug-likeness properties of the final screened compounds were compared to FDA approved HIV-integrase inhibitors. HIV-integrase structure in complex with the most active and final screened compounds were subjected to 50ns molecular dynamics (MD) simulation studies to check comparative stability of the complexes. The study suggested that the screened compounds might be promising HIV-integrase inhibitors. The new chemical entities obtained from the NCI database will be subjected to experimental studies to confirm potential inhibition of HIV integrase.

Introduction

Human immunodeficiency virus (HIV) is the aetiological agent of acquired immunodeficiency syndrome (AIDS) which destroys the immune system of the body leaving the victim vulnerable to infections, malignancies and neurological disorder. Owing to its rapid spread it has become a serious global threat and there is no curative treatment for this fatal disease. According to statistics by World Health Organization (WHO), a total of 37.20 million people are living with AIDS and 1.70 million people died in 2013 alone. Currently, there are 3 categories of therapeutic anti-HIV drugs based on their inhibitory mechanisms[1] and these include nucleoside reverse transcriptase inhibitors (NRTIs)[2], non-nucleoside reverse transcriptase inhibitors (NNRTIs)[3], and protease inhibitors (PIs)[4, 5]. To date the highly active antiretroviral therapy (HAART)[6] which is combined therapy using the above classes of inhibitors is widely used for patients with advance infection but has failed to eradicate the virus. HAART is intended to slow down viral replication and lower the patient's total burden of HIV infection, but this treatment is not entirely cost effective and is often out of reach of people worldwide. The genome of the HIV encodes for three enzymes viz. the protease, reverse transcriptase and integrase. The integrase is a 32 kDa enzyme made of three functional domains included an N-terminal domain, catalytic core domain and a less conserved C-terminal domain[7, 8]. The HIV

[a]*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa.*
[b]*Division of Chemical Pathology, University of Cape Town, , South Africa.*
Correspondence should be addressed to T.S. Pillay, Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria, 0007
Email: tspillay@gmail.com
Phone: +27-12-319-2114
Fax: +27-12-3283600

integrase has no equivalent counterpart or sequence homologue in the human host cell and consequently it can be considered as an attractive drug target[9]. After the approval of Raltegravir[10] as anti-HIV drug several integrase inhibitors emerged as promising class of therapeutics for the treatment of AIDS. Raltegravir and several other inhibitors have been identified to possess anti-HIV integrase activity but these have adverse effects on prolonged use and the development of drug resistance drives the need to explore new novel and potential chemical scaffolds for the treatment of AIDS.

The computational methods in drug discovery collectively termed pharmacoinformatics, includes structure activity relationship (SAR), pharmacophore, virtual screening and molecular docking and these have proven their pivotal role in the pharmaceutical industry for lead identification and optimization. Several research groups worldwide identified integrase inhibitors[11-17] using pharmacoinformatics approaches for potential application for HIV therapy. Consistent with the objective of developing new potent and less toxic integrase inhibitors the current research explores the binding preferences of the inhibitory molecules of HIV integrase in terms of space modelling study and virtual screening along with molecular docking and molecular dynamics.

A pharmacophore model is an collection of steric and electronic features and provides an intuitive way of depicting and understanding the binding properties of small molecules along with an explanation of optimum supra-molecular interactions with a precise biological target , to activate (or block) its biological response[18, 19]. It can also be defined that the pharmacophore idea is based on the kinds of interaction observed in molecular appreciation, *i.e.*, hydrophobic, hydrogen bonding, and charge interaction. For the HIV integrase inhibitors the hydrogen bond acceptor (HBA) and donor (HBD), hydrophobic (H) and aromatic ring (RA) pharmacophore features were found to be the important functional features associated with the selectivity and potency. The pharmacophore models are widely used in the field of drug discovery by providing valuable information to study SAR and reveals the mechanism of ligand-target relationship by deducing the nature of functional groups and non-covalent bonding patterns[20]. It can also be used in virtual screening to identify potential molecules, predict the activity of the newly synthesized compound before animal experiment; or understand the possible mechanism of action[21, 22]. In the current study, an attempt was made to identify the pharmacophore hypothesis using the *HypoGen* module[23] based on key chemical features of HIV-integrase inhibitors with inhibition constant covering a satisfactory wide range of magnitude. The model was validated using several statistical approaches including Fischer's randomization and test set prediction. The validated model was utilized for the virtual screening to select the virtual hits from structural database. The molecular docking study was also performed to elucidate the binding interactions and preferred orientation of proposed potential molecules. The potential of the work is displayed by the identification of two potential lead molecules as integrase inhibitors. Finally, the molecular dynamics study was performed to analyse stability and precise binding interactions of the screened molecules inside the receptor cavity of HIV integrase.

Materials and methods

At present, several popular commercial and freeware packages are used for ligand-based method to derive 3D pharmacophore models and also help in estimation of biological activities. Here we used Discovery Studio 4.0 (DS)[24] for the 3D QSAR pharmacophore, virtual screening and molecular docking studies. This is commercially available software containing several module packages and widely used in pharmacoinformatics drug discovery[25-28]. The *3D QSAR Pharmacophore Generation* module takes input of structure and activity data for a set of potential HIV-integrase ligand to create hypotheses. Two modules, *HypoGen* and *HipHop* are used for ligand-based pharmacophore modelling. The *HypoGen* allows identification of hypotheses that are common to the 'active' molecules of training set but absent in the 'inactive' molecules, whilst *HipHop* identifies hypotheses present both in 'active' and 'inactive' compounds. In the present work the *HypoGen* module was used to generate the hypotheses.

Dataset

1437 compounds belong to a collection of HIV-1 integrase inhibitors were downloaded from BindingDB (http://www.bindingdb.org/) with data on inhibitory activity (IC_{50}). Duplicate and compounds without definite activity values were deleted and finally 540 compounds considered as whole dataset for the study. Training and test set compounds were separated from whole dataset for pharmacophore model generation and validation of generated model respectively. The molecules of the dataset have a wide range

of IC_{50}, from 2.000 to 1000000.000 nM. For simplicity the whole dataset was divided into three sets on the basis of inhibitory activities values; highly active (IC_{50} < 100.000 nM, +++), moderately active (100.000 ≤ IC_{50} < 1000.000 nM, ++) and least active (IC_{50} ≥ 1000.000 nM, +). For selection of the training set for pharmacophore model generation in DS basic guidelines laid down by Li et al.[29] were followed. The guidelines reported as a) molecules should be selected to provide clear and brief information with structure features and range of activity, b) at least 16 diverse molecules for training set should be considered to ensure the statistical significance and avoid chance correlation, c) the training set must include the most and the least active molecules and d) the biological activity data of the molecules should have spanned at least 4 orders of magnitude. Following the above guidelines four training sets were (Set 1, Set 2, Set 3 and Set 4) generated containing 30 compounds each. It was also kept in mind that no compounds were common in any two training sets except for the most active and least active molecules. The remaining 410 molecules were considered as test set molecules for each set and used for assessing the performance of pharmacophore model. The 2D/3D visualizer[24] of DS was used to generate three-dimensional coordinates of the compounds. For each compound, the coordinates were corrected, atoms were typed and energy was minimized using the modified CHARMm force field[30, 31]. The several packages of DS were used for pharmacophore, virtual screening and molecular docking studies.

Pharmacophore model generation

In order to generate the pharmacophore space model the *3D QSAR Pharmacophore Model Generation* module of DS was used. Conformations of the training set molecules were generated by *Cat-Conf* program of the DS software package. Out of BEST/FAST, the BEST method was considered to obtain multiple acceptable conformations which provides complete and enhanced coverage of conformational space with help of rigorous energy minimization and optimizing the conformations by the poling algorithm[32]. In the BEST algorithm, the chemical features are arranged in space instead of simply the arrangement of atoms[33]. For prediction of the favourable features for the highly active compounds of the dataset the *Feature mapping* was considered. Mapped features were given as input features for pharmacophore model generation. Using the conformer along with chemical features the modules operates in two modes such as *HipHop* and *HypoGen*. The *HipHop* approach generates the pharmacophore models by using active compounds only, while the *HypoGen* approach considered both active and inactive compounds in order to find out a hypothesis which is common in the active molecules and absent in the inactive compounds[33]. Top ten hypotheses are generated by the *HypoGen* with consideration of the training set, conformational models and chemical features through three steps: constructive, subtractive and optimization[34]. In the first step, hypotheses are generated that are common in the most active compounds; in subtractive phase, inactive compounds are removed from those that fit the hypotheses. In final step, the remaining hypotheses improve the score with help of small perturbations[33, 35]. The best hypothesis was selected based on the best correlation coefficient (*R*), low root mean square deviation (*rmsd*), cost function analysis and good predictive ability.

All four sets were used to develop the pharmacophore models and statistical parameters were calculated based on training & test sets molecules. Statistical results are depicted in Table S2 in Supplementary file. It was observed that Set 1 gives the better statistical results compared to Sets 2 – 4. Hence the Set 1 (n_{tr} = 30, Fig. 1) was considered as the training set in the current study. In the remaining section "training set" is explained as Set 1 compounds and "test set" as test compounds of Set 1. Training set molecules in SMILES format with observed and estimated activity along with fit values of Set 2, Set3 and Set 4 are depicted in Tables S2, S3 and S4 respectively in Supplementary data. The information concerning the structure and the biological activity of test set compounds of Set 1 is provided in Table S5 in the supplementary information, while all the data regarding the training set (Set 1) molecules are reported in Fig. 1. The activity value distribution of the training set molecules of Set 1 is given in Figure S1 (Supplementary file).

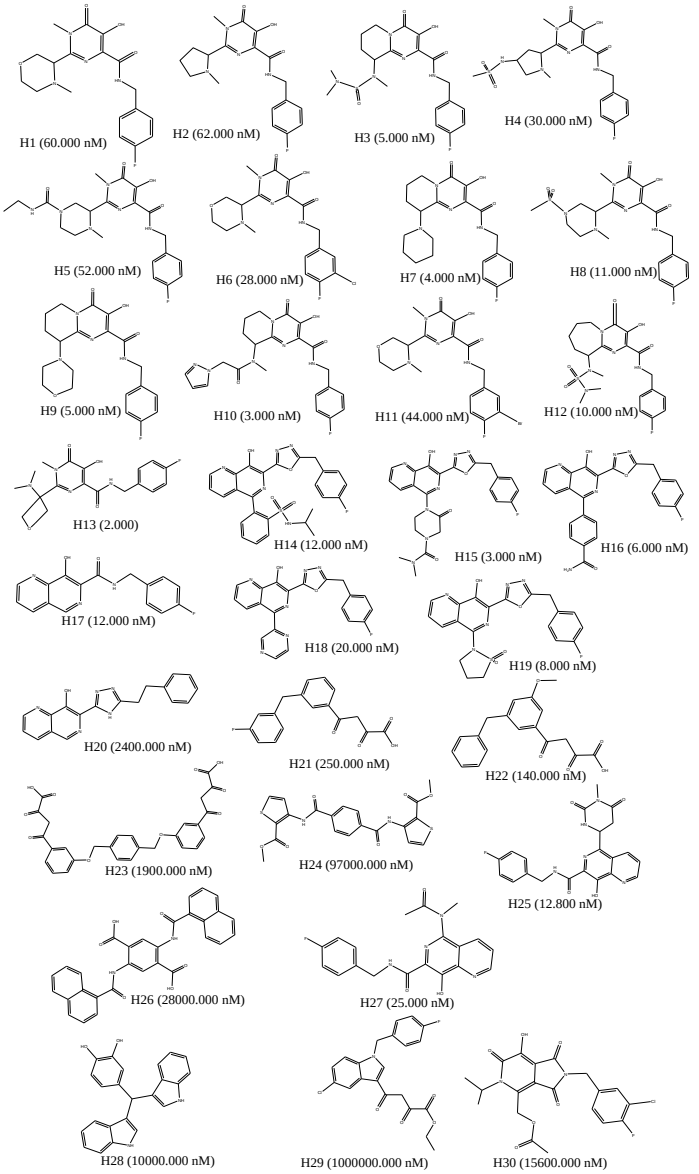

Fig. 1 2D chemical structures of the training set compounds and the activity values (IC_{50}) are given in the parentheses.

Validation of pharmacophore model

In order to check the predictivity and applicability as well as robustness of the pharmacoinformatics model, the pharmacophore hypotheses developed were validated by five different methods, (1) internal validation, (2) cost function analysis, (3) Fischer's randomization test, (4) test set prediction and (5) decoy set.

Internal validation

Leave-one out (LOO) cross-validation is one of the important internally validation protocol of the selected model, in which one compound was randomly deleted from training set in each cycle and model redeveloped using the rest of the compounds with the same parameters used in original model. The activity of the deleted compounds was calculated based on the newly developed model. The above procedure applied for all molecules of the training set and predicted activity recorded. Two important statistical parameters, the LOO cross-validated correlation coefficient (Q^2) and error of estimation (*se*) were calculated based upon predicted activity of training compounds. It is reported that high Q^2 (>0.5) and low *se* explained better predictive ability [36].

Another parameter, the modified r^2 ($r^2_{m(LOO)}$) developed by Roy *et al.*[37, 38] was calculated to confirm the good predictive ability of the training set molecules. This parameter measure the degree of deviation of the predicted activity from the observed ones. It was stated that model may be considered with $r^2_{m(LOO)}$>0.5.

Cost function analysis

To select the final pharmacophore model several statistical parameters were employed at the time of hypothesis generation, these included spacing, uncertainty, and weight variation. The spacing represents the minimum inter-features distance which may be permissible in the final hypothesis, while the weight variation reflects the level of magnitude explored by the hypothesis in which every feature implies some degree of magnitude of the biological activity of the compound. The default values of spacing and weight variation are 300 and 0.3 respectively but some cases it varies from 400 to 100 and 1 to 2 respectively. The uncertainty returns the error of prediction which signifies the standard deviation of the error cost. The default value of uncertainty parameter is 3 but it may vary from 1.5 to 4.0 depending upon the nature of work. The cost-function analysis is an important aspect for selection of final model which minimized three terms, *viz.*, weight cost, error cost, and configuration cost. The weight cost is directly proportional to the deviation of weight variation from its input value. The error cost is the deviation between the predicted activity and experimentally determined activity of the molecule in the training set. A fixed cost is determined by the complexity of the hypothesis space. The configuration cost implies entropy of hypothesis space and it is reported that value should have <17 for a good pharmacophore model. *HypoGen* module also calculates the null hypothesis which is the assumption that there is no relationship in the data, and the experimental activities are distributed about their mean. It is

illustrated that the higher (>60) cost difference (*Δcost* = null cost - total cost) indicted that the hypothesis does not reflect a chance correlation.

Fischer's randomization test

In order to check the strong relationship between the chemical compound and the biological activity of the training set molecules the Fischer's randomization test was performed. In this method, the biological activity was scrambled and assigned new values. After that the pharmacophore hypotheses were generated with new values of activity using the original pharmacophoric features and constraints used to generate the original pharmacophore hypotheses. If the randomization run generates improved correlation coefficient and/or better statistical parameters than the original hypothesis may be considered to be developed by chance. Different number of spreadsheets are generated based on the statistical significance randomization run. The statistical significance is given by following equation.

$$Significance = [1-(1+a)/b] \quad (1)$$

Where, *a* denotes the number of hypotheses with a total cost less than the best hypothesis, whereas *b* implies a collection of *HypoGen* runs and random runs. For example, at 95% confidence level total number of random spreadsheet are generated as 19 (*b* = 20) and each generated spreadsheet is submitted to *HypoGen* using the same parameters as the initial run. In the present study, the developed pharmacophore model was tested at 95% confidence level which produced 19 spreadsheets.

Test set prediction

The ability to judge the external predictivity of the model beyond the training set molecules is an important step which verifies ability of prediction of test compounds. Accordingly, in the present work 410 test compounds were predicted using the developed pharmacophore model by *Ligand Pharmacophore Mapping* protocol in DS, given in Table S1 (Supplementary file). Quality of prediction of the pharmacophore model was adjudged best on statistical parameters, R^2_{pred} (correlation coefficient) and s_p (error of prediction)[39, 40].

The R^2_{pred} value depends on the average experimental activity of the training set molecules. Since these parameters depend on average value, it might be achieved for compounds with a wider range of activity value, but this may not be guaranteed that the estimated activity values are very close to those experimental activity. As a result, instead of a good overall correlation being maintained, there is chance of a significant numerical difference between the two values. In order to better indicate the predictivity of the pharmacoinformatics model, modified r^2 [$r^2_{m(test)}$][41, 42] values were calculated (threshold value=0.5).

Decoy set

In order to check the efficiency of the screening capacity of the selected pharmacophore model, it was validated using the decoy set approach. Decoy set method checks how the model can select active molecules over inactive molecules on screening with amalgamated active and inactive molecules. In this purpose a set decoy was generated by DecoyFinder1.1[43]. Decoys physically resemble active inhibitors but differ chemically to avoid bias in the enrichment factor calculation. Based on five parameters decoys were selected and these included molecular weight, number of rotational bonds, hydrogen-bond donor count, hydrogen-bond acceptor count, and the octanol–water partition coefficient of the active inhibitors. In order to discriminate decoys and active inhibitors chemically, the MACCS fingerprints were calculated according to the maximum Tanimoto coefficient values. The screening dataset consisting of 80 active HIV-integrase inhibitors and 900 decoys obtained from DecoyFinder was used for queries of the selected hypothesis. Different statistical parameters including the accuracy and enrichment factor (EF) were calculated to validate the pharmacophore model. Screened molecules based on pharmacophore model were ranked on basis of fit value. For the assessment of effectiveness of screening the enrichment calculation of the dataset was performed. The EF implies total known active inhibitors retrieved from the part of screened database. In the current study, EF (1%) was calculated from the top 1% hits. Another parameter Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) which gives the significance of the dataset screening was also calculated. The BEDROC is a comprehensive form of receiver operating characteristic (ROC), which recognises problems in the screening method. Calculation of the enrichment factor and BEDROC are described by Bhayye et. al[44].

Virtual screening

Virtual screening is a crucial technique to discover novel potent compounds that can bind to a particular receptor site to block or trigger the activity. In the current research, the NCI (National Cancer Institute) database was used to search to obtain novel chemical entities for HIV integrase inhibitors. The NCI database contains 265,242 compounds. Both best pharmacophore and hyporefined models were submitted separately to the NCI database with set 'Limit Hits' as 'Best N' and 'Maximum Hits' as 600. The initial hit molecules were filtered with a number of criteria to get final potential molecules for the HIV integrase. Furthermore, molecular docking study was carried out to analyse binding interactions between the potential HIV-integrase compounds and catalytic residues of the active site.

Molecular docking

Molecular docking is one of the finest filtering techniques and an important method in the drug design process. The *LigandFit* protocol of DS was used in order to understand how the screened drug-like virtual hits bind to the receptor through molecular docking. The protocol of this module first detects the cavity to identify and select the region of the protein as the active site followed by docking the ligands to the selected site. In order to detect the site the 3D regular grids of points were employed. The protein receptor of the HIV integrase was retrieved from RCSB Protein Data Bank (RCSB-PDB) for the molecular docking study. It is reported that a protein structure may be suitable for molecular docking study with the low resolution (<2.5Å) and R-factor (<0.28)[45]. In the present study PDB ID: 1QS4[46] was selected among several HIV integrase keep in mind the above criteria along with receptor size and date of deposit. The *Prepare Protein* and *Prepare Ligand* tools of DS were used to prepare receptor and ligands respectively. Both protein and ligand were minimized using CHARMm force field[47]. For protein preparation using *Prepare Protein* module of DS the 'Build Loop' and 'Protonate' parameters were fixed to 'True' while, dielectric constant, pH, ionic strengths and energy cut-off were considered as default value. In *Prepare Ligand* module, preparation 'Change ionization', 'Generate Tautomers' and 'Generate isomers' were fixed to 'False', and 'Generate Coordinates' was set to '3D'. Followed by protein preparation, the binding site was identified on the basis of volume occupied by the ligand at the active site the binding site was identified. In order to avoid false positive results of molecular docking the validation of docking protocol is an essential step and for which the co-crystal small molecule in the PDB complex file was initially redrawn and the same docked into the active site of HIV integrase (PDB ID: 1QS4). The best docked pose of co-crystalized ligand was checked for the binding interactions at the active site followed by superimposing the docked pose and the co-crystal.. The RMSD value was calculated in order to verify docking parameters that were capable to regenerate a comparable conformation to that of the co-crystal at the active site of HIV integrase. Further molecular docking studies of potential molecules were carried out with the same settings as in the co-crystalized docking. For the analysis of binding interactions and dock score values, top ten conformations for each ligand were considered.

Molecular Dynamics

In order to perform MD simulation and part of the analysis of the trajectories the AMBER 12[48] was used for the selected docked poses. The generalized amber force field was used for preparation of both ligand and receptor. A rectangular box of TIP3P water was created with boundary of 10Å around the each system. Each system was minimized for 500 steps of each conjugate gradient and steepest descent method. The system was heated at constant volume and temperature of 10 to 150K with the Langevin thermostat over the course of 100ps with 25kcal mol^{-1} $Å^{-2}$.

Drug-likeness analysis of screened compounds

In order to compare the drug-likeness of the screened compounds with existing Food and Drug Administration (FDA) approved HIV integrase inhibitors different parameters including dockscore, estimated activity, fit value, molecular weight, logP, violation of Lipinski's rule of five, molecular volume, molecular refractivity, number of H-bonds and number of bump interactions were analysed. For this purpose DS[24] and online program Molinspiration (www.molinspiration.com) were used. Dockscore is defined as the collection of internal energy of ligand-receptor complex and ligand only. The estimated activity and fit value were calculated after fitting with best pharmacophore model. logP measures the hydrophobicity of the molecules. Lipinski's rule of five states that for a drug-like molecule logP value should be less than 5, hydrogen bond acceptor and donor should be less than 10 and 5 respectively,

and molecular weight should be less than 500. Molecular volume explains transport characteristics of molecules, such as intestinal absorption or blood-brain barrier penetration.

Results and Discussion

The *HypoGen* module was used to develop the pharmacophore model based on training set (n_{tr} = 30) compounds selected from the whole dataset. The training set molecules are depicted in the Fig. 1 (**H1 – H30**) and the inhibitory activity values (IC_{50}) are given within the parentheses. The *Feature mapping* protocol of DS was used to select the pharmacophoric features, 'HBA', 'HBD', 'H' and 'R' as required for chemical features and were given as input to the 3D QSAR pharmacophore generation along with keeping minimum and maximum feature value '0' and '5' respectively. On successful completion of pharmacophore generation the top ten hypotheses were output with excellent statistical parameters. The statistical parameters of the selected model derived based upon the biological activity of the training molecules are given in the Table 1. For selection of best pharmacophore model Debnath's[21, 49] analysis states that model should have the less *rmsd*, high correlation coefficient, lowest cost value and highest cost difference. In the current study, the predictivity of first hypothesis (*Hypo 1*) was confirmed by Debnath's method[21, 49]. A valid hypothesis should have the overall cost of the hypothesis different from the null cost and near to the fixed cost, and it is reported that the cost difference (Δ*cost* = null cost – total cost) in the range of 40–60 bits explains the probability of the predictive correlation of 75–90%, while the Δ*cost* more than 60 bits means the hypothesis has a correlation probability of greater than 90%[50]. In the selected pharmacophore in the present study, the cost difference for *Hypo1* (Table 1) was found to be 914.671, i.e. much more than 60 bits, explaining that selected model has a >90% chance of being able to select HIV integrase inhibitors.

Table 1: Statistical results and predictive power (presented as cost, measured in bits) of the top ten hypotheses of training set molecules of HIV integrase inhibitors

Hypo No.	Spacing	[1]Unc.	[2]Wt. Var.	[3]*R*	[4]*Rmsd*	Costs					Output features
						Total	Null	Fixed	[5]Δ	[6]Config.	
1	250	1.5	0.5	0.940	2.847	223.699	1138.370	89.046	914.671	16.664	2xa, r
2	250	1.5	0.5	0.929	3.087	251.522	1138.370	89.046	886.848	16.664	2xa, r
3	250	1.5	0.5	0.895	3.730	299.662	1138.370	89.046	838.708	16.664	3xa, p
4	250	1.5	0.5	0.897	3.688	304.535	1138.370	89.046	833.835	16.664	2xa, r
5	250	1.5	0.5	0.882	3.943	323.382	1138.370	89.046	814.988	16.664	3xa, p
6	250	1.5	0.5	0.868	4.150	357.335	1138.370	89.046	781.035	16.664	3xa, p
7	250	1.5	0.5	0.862	4.241	362.352	1138.370	89.046	776.018	16.664	2xa, r
8	250	1.5	0.5	0.857	4.310	368.198	1138.370	89.046	770.172	16.664	2xa, d, p
9	250	1.5	0.5	0.841	4.520	417.176	1138.370	89.046	721.194	16.664	2xa, r
10	250	1.5	0.5	0.829	4.678	425.714	1138.370	89.046	712.656	16.664	2xa, r

[1]Uncertainty; [2]Weight variation; [3]Correlation coefficient; [4]Root mean square deviation; [5](Null cost- Total cost); [6]Configiration cost, a = HBA; d = HBD; p = H; r = RA

For all ten hypotheses the correlation coefficient (*R*) values were recorded (Table 1) and it was observed that all selected hypotheses had correlation value of >0.820, but the best hypothesis revealed with the highest *R* value (0.940), which explained good predictive ability of the selected hypothesis. In case of *Hypo1* the fixed and total cost values were obtained as 89.046 and 223.699, respectively, while the difference between total and null cost was found to be 914.671. From Table 1 it is delineated that the high correlation coefficient, less *rmsd*, highest cost difference and minimum error values were observed for *Hypo1* in comparison to other hypotheses. Hence, *Hypo1* was considered as the best model for further analyses.

The best model (*Hypo1*, Fig. 2a) obtained with critical features of two 'HBA' (HBA1 and HBA2) and one 'RA' and portrayed in Fig. 2a after mapping with most active compound of the dataset. In order to nullify over-prediction of the bioactivity for inactive molecules

using *Hypo1* the HypoRefine was performed. In this purpose the *Steric refinement with Excluded Volumes* module of DS was used. The algorithm identifies areas of space that are occupied by inactive ligands and strategically places excluded volumes (*ev*) in these regions to reduce the number of inactives that can map the pharmacophore. It was observed that ten *ev* factors were critical for the inhibitory activity of HIV integrase. The model with *ev* is delineated in Fig. 2b. Quality of the selected hypothesis was justified by its ability to predict the activity of individual compounds within the training set. It was observed that all the training set molecules were accurately predicted with low error values between the actual and estimated IC_{50}. The predicted activity of the training set molecules explained that one active compound was overestimated as moderately active and two moderately active molecules were underestimated as active compounds. The remaining compounds were classified correctly within their region. Based on the discussion above it can be concluded that the *Hypo1* predicted biological activity of the dataset correctly which is reflected by the high correlation between observed and estimated biological activities of training and test sets. The error values explain the consistency between the experimental and estimated biological activity and defined as ratio between them. It was observed that except three (**H5**, **H23** and **H29**) all compounds of the training set have error values within reasonable range. The observed and estimated activities on the based on *Hypo1* for the training set compounds are given in Table 2 and Fig. 3. It was observed that best hypothesis gives Q^2 of 0.912 and *se* of 0.498. The high Q^2 and low *se* of the selected model suggested that model is robust in nature.

Fig. 2 a) Mapped pharmacophore features (*Hypo 1*) with most active compound; b) *Hypo 1* with excluded volume; c) Inter-feature distances of *Hypo 1*.

The most active compound of the dataset was mapped in the best model (*Hypo 1* in Table 2) and depicted in Fig. 2a. Mapped features suggested that amine and hydroxyl group present in the six membered non-aromatic ring as HBA1 and HBA2. The aromatic ring presence in the molecular system of most active compounds was found to be important as ring aromatics factor. Gupta et al.[16, 51] developed QSAR, pharmacophore and comparative molecular field analysis (CoMFA) studies on curumine derivatives as HIV-1 integrase inhibitors. Both studies found that hydrophobicity and ring aromatic along with hydrogen bond acceptor and donor were crucial for inhibitory activity. These findings are comparable with the outcomes of current research. Therefore it can be postulated that to design or synthesize new chemical entities of HIV integrase inhibitors HBA and R factors with critical inter-feature distances (Fig. 2c) will be crucial factors.

Test set prediction

Any robust pharmacoinformatics model should have the capability to predict the activities of the compounds other than training set. Out of whole dataset, 510 compounds were considered as test set and the activity (IC_{50}) converted into logarithm value (pIC_{50}) and given in Table S5 in supplementary file and Fig. 3. The entire test set of molecules were divided into three groups based on their inhibitory activity (pIC_{50}) values: highly active (pIC_{50} < 2.000 nM*,* +++), moderately active (2.000 nM ≤ pIC_{50} < 3.000 nM*, ++)* and least active/inactive (pIC_{50} > 3.000 nM*,* +). The *Ligand Pharmacophore Mapping* of DS activity was used to estimate the inhibitory activity of test set compounds. It was observed that six compounds estimated far from the observed activity hence were considered as outliers and not used in further analysis. Analysis of observed and estimated activity of test compounds revealed that out of sixteen active compounds from the highly active category, seven and nine compounds were overestimated as moderately active and least active respectively. In case of moderately active compounds, 17 and 6 compounds were underestimated as highly active and overestimated as least active respectively. Seven and six compounds belong to the least active category and were underestimated as moderately and highly active respectively. The remaining 450 compounds were classified in their observed and predicted activity correctly (Table S5 in supplementary file) which suggests that the selected model was able to provide accurate estimation for the biological activities of external compounds. The correlation (*R*) between observed and estimated activity of test

compounds was found to be 0.915 and the R^2_{pred} value of 0.847 with error of prediction (s_p) of 0.697.

Furthermore, to check the power of predictive capability of the selected model, the values of $r^2_{m(test)}$ were also calculated. The $r^2_{m(test)}$ explains how the estimated activities are close to the corresponding experimental values as a good correlation coefficient value (R^2_{pred}) may not always suggest a low residual between the observed and estimated activity data. The $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ were found to be 0.636 and 0.130 respectively, explaining that selected model has adequate predictive potential. Hence, the above results suggested that the selected model can reasonably predict the biological activities of new chemical compounds.

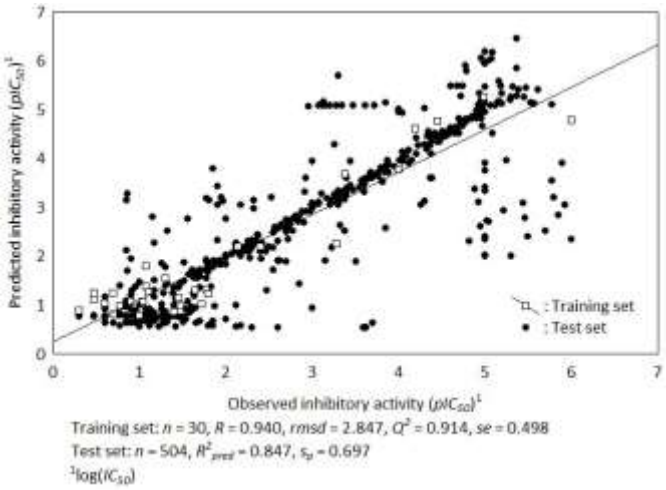

Fig. 3 Observed and predicted inhibitory activity of HIV integrase inhibitors based on *Hypo 1*.

Table 2: Observed, predicted activities and fit values of the training set molecules, obtained using the pharmacophore model *Hypo1*

Comp.	Activity (IC_{50} nM)		Error	Activity scale		Fit value
	[1]Obs.	[2]Pred.		[1]Obs.	[2]Pred.	
H1	60.000	20.198	-2.971	+++	+++	12.877
H2	62.000	16.214	-3.824	+++	+++	12.973
H3	5.000	17.129	+0.292	+++	+++	12.949
H4	30.000	9.523	-3.150	+++	+++	13.204
H5	52.000	10.359	-5.020	+++	+++	13.167
H6	28.000	13.877	-2.018	+++	+++	13.040
H7	4.000	10.800	+0.370	+++	+++	13.149
H8	11.000	11.548	+0.953	+++	+++	13.120
H9	5.000	6.418	+0.779	+++	+++	13.375
H10	3.000	17.153	-0.175	+++	+++	12.948
H11	44.000	20.032	-2.197	+++	+++	12.881
H12	10.000	6.031	-1.658	+++	+++	13.402
H13	2.000	7.453	+0.268	+++	+++	13.310
H14	12.000	23.149	+0.518	+++	+++	12.818
H15	3.000	12.650	+0.237	+++	+++	13.080
H16	6.000	9.553	+0.628	+++	+++	13.202
H17	12.000	60.462	+0.198	+++	+++	12.401
H18	20.000	34.581	+0.578	+++	+++	12.644
H19	8.000	11.500	+0.696	+++	+++	13.122
H20	2400.000	4786.360	+0.501	+	+	10.503
H21	250.000	156.173	-1.601	++	++	11.989
H22	140.000	157.416	+0.889	++	++	11.986
H23	1900.000	174.085	-10.914	++	++	11.942
H24	97000.000	176422.000	+0.550	+	+	8.936
H25	12.800	17.849	+0.717	+++	+++	12.931
H26	28000.000	57400.900	+0.488	+	+	9.424
H27	25.000	7.106	-3.518	+++	+++	13.331
H28	10000.000	5908.690	-1.692	+	+	10.411
H29	1000000.000	60599.000	-16.502	+	+	9.400
H30	15600.000	40312.200	+0.387	+	+	9.577

[1]Observed; [2]Predicted

Fischer randomization test

The *hypo 1* was select for the Fischer randomization test to adjudge the statistical significance of the model of interest by allocating a particular confidence level. Selected hypotheses were considered for the Fischer randomization test at confidence level of 95%. In this confidence level, observed activities of training set molecules are reshuffled and generated into 19 random spreadsheets, and each spreadsheet creates a hypothesis. The implication of the hypothesis was calculated as per equation (1). The correlation values of all 19 spreadsheets obtained and depicted in the Table 3 indicated that not a single value obtained after randomization-produced hypotheses that exhibited predictive powers similar to or better than that of *hypo 1* (Table 1).

Table 3: Correlation and total cost of the 19 random spreadsheets

Validation	Correlation	Total cost
Hypo1	0.940	223.699
random1	0.592	770.212
random2	0.509	862.437
random3	0.743	572.127
random4	0.622	732.664
random5	0.570	799.428
random6	0.686	646.582
random7	0.553	818.048
random8	0.593	767.586
random9	0.461	912.705
random10	0.716	605.744
random11	0.778	511.697
random12	0.756	551.685
random13	0.749	572.692
random14	0.527	849.479
random15	0.573	794.599
random16	0.752	568.053
random17	0.725	608.415
random18	0.756	553.663
random19	0.541	832.001

From Table 3, the average of correlation coefficient for all 19 trials was found to be 0.642. It was also observed that the total costs of randomized runs were much higher than the total cost of *Hypo 1*. The total cost of *Hypo 1* and all other 19 random hypotheses is depicted in the Fig. 4 and Table 3. The above discussion undoubtedly demonstrated that the selected pharmacophore model was not produced by chance.

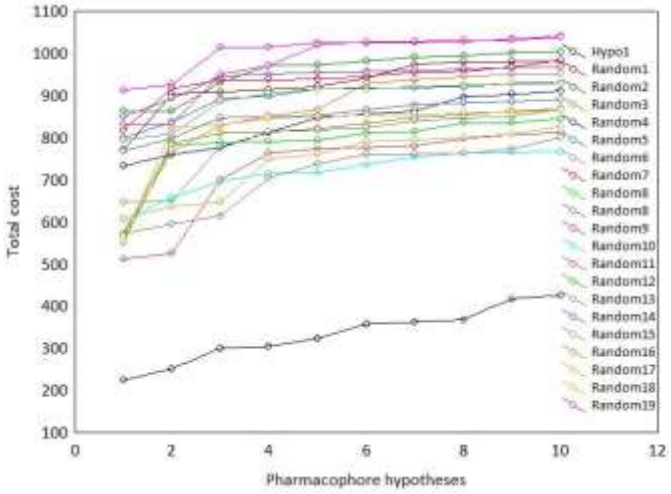

Fig. 4 Comparison of the total costs of *Hypo 1* and 19 random hypotheses generated in the Fischer's randomization test.

Decoy set

Decoy set validation of pharmacophore model is one of the important approaches to evaluate the screening capability of the model. In this regard, the *hypo1* was screened by a set of 900 HIV-integrase decoys obtained by DecoyFinder1.1 amalgamated with 80 active HIV-integrase inhibitors. The accuracy and enrichment analyses of decoy set were calculated based on the screened results. The model was successful used to identify and differentiate actives and decoys with good accuracy of 0.93. The true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values were found to 57, 42, 858 and 23 respectively. The ROC plot was derived for the model and given in Fig. 5 which indicated that actives and decoys are well-classified. The area under curve (AUC) was also calculated and the value was found to be 0.69 that clearly

indicated more true positives have been verified. The enrichment factor (EF 1%) and Boltzmann-enhanced discrimination of ROC were calculated. Average EF 1 % value for pharmacophore model was found to be 9.80 which indicated that model has identified active compounds very well and the top 1% hit is enriched with active compounds. The average BEDROC was obtained as 0.85 which signifies that the top hits is not only enriched with active compounds but also ranked higher than decoy. The abovementioned results of decoy validation strongly suggest that the developed pharmacophore features in the selected model are perfectly acceptable for the mapping of HIV-integrase inhibitors

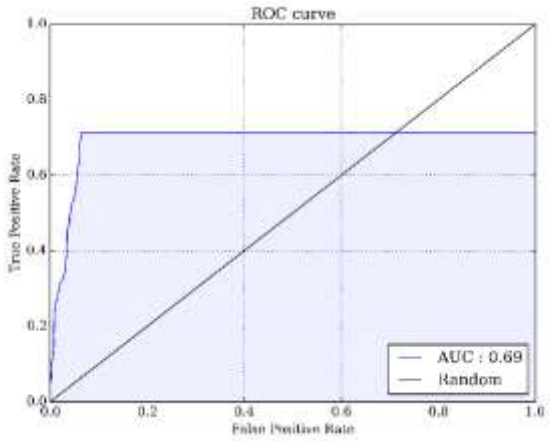

Fig. 5 ROC curve for pharmacophore model, derive from true positive rate of actives vs. false positive rate of inactive compounds.

Virtual screening

In order to find potential molecules that are HIV integrase inhibitors virtual screening is a powerful technique and also effective as an alternative to high-throughput screening methodologies. For this purpose the validated pharmacoinformatics models including QSAR, pharmacophore or molecular docking can be used to search the molecular database for lead identification. The *Hypo 1* and hyporefined of *Hypo 1*(Fig. 2a and 2b) were used to explore the NCI database (comprising of 265,242 compounds) separately to find potential HIV integrase inhibitors. The '*Search Database*' under '*Pharmacophore*' module of DS was used for screening of the database, where the protocol 'Search Method' and 'Limit Hits' were set to 'Best' and 'Best N' respectively. 'Maximum Hits' was set to 600 for each screening method. As per the given criteria each hypothesis retrieved 600 best compounds. Compounds from both models were merged and found there were 79 redundant molecules in the merged file. After deletion of redundant molecules, the remaining 1121 compounds were fitted to the pharmacophore model by the *Ligand Pharmacophore Mapping* protocol of DS with maximum omitted feature set to '0'. It was observed that except for one compound all were fitted successfully to pharmacophore model. Estimated biological activity and fit score value of all 1120 molecules was recorded. The estimated inhibitory activity and fitscore values of all 1120 compared with estimated inhibitory activity (7.453nM) and fitscore value (13.210) of most active (**H13** in Fig. 1) compound of the dataset were obtained. Compounds with less estimated activity (IC_{50}) than **H13** and greater fit value than **H13** were considered for further screening. Out of 1120 only 13 compounds satisfied the above criteria and were considered for further screening. Furthermore the Lipinski's rule of five[52] and Veber's[53] rule were checked for 13 compounds. It was observed that 8 compounds failed to pass both rules. The remaining 5 compounds further were taken into consideration for molecular docking study in the active site of HIV integrase (PDB ID: 1QS4). The dock score of these 5 compounds were compared with dock score of **H13**. Three compounds (**NSC91705**, **NSC651812** and **NSC666331**) were found to have higher dock scores compared to **H13**. Finally, the above 3 compounds were considered for ADMET and synthetic accessibility check. It was observed that blood brain barrier (BBB) level and synthetic accessibility of **NSC666331** were 4 and 6.51 respectively which indicated that this molecule is difficult to synthesize and penetration may be problematic, hence **NSC666331** was deleted and finally **NSC91705** and **NSC651812** (Fig. 6) were considered as promising compounds and further subjected to assess the critical interactions with the catalytic amino acid residues present in the active site cavity of the HIV integrase.

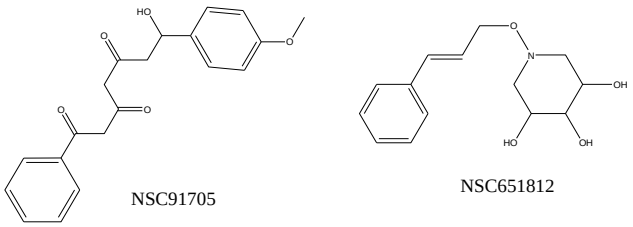

Fig. 6 Screened promising HIV-integrase inhibitors from NCI database

Molecular docking

The molecular docking study gives the accurate and preferred orientations of the molecule at the receptor site of the macromolecule. Best docked poses of the promising HIV-integrase ligands obtained from NCI database (Fig. 6) and most active compound of the dataset (**H13** in Fig. 1) were considered to evaluate the optimal orientation and binding abilities. The crystal structure of HIV integrase (PDB ID: 1QS4) was collected from RCSB-Protein Data Bank. In order to validate the docking protocol, self-docking[54] is one of the important techniques in which bound ligand is docked at the catalytic site of protein molecule and the conformer of the original bound ligand is superimposed to the docked poses to calculate root mean square deviation (RMSD) values. It is illustrated that low RMSD (<2 Å) value of original bound ligand validates the docking procedure[54]. The RMSD values was found 1.406Å, which indicated that the protocol was selected in the docking method was validated.

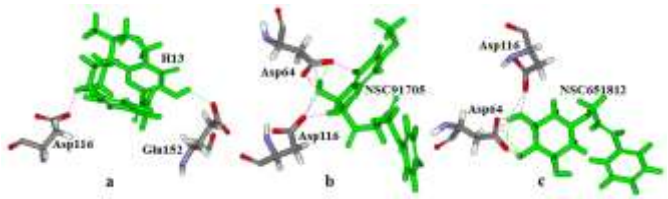

Fig. 7 Binding modes of a) most active compound of the training set and NCI database screened potential HIV-integrase molecules b) **NSC91705** and c) **NSC651812**.

The molecular docking study (Fig. 7) of between most active compound (**H13**) of the training set and HIV protease revealed that Asp116 and Glu152 were important catalytic amino residues. One of each hydrogen bond and bump interactions were observed between the ligand with Asp116 and Glu152 respectively. Both screened compounds **NSC91705** and **NSC651812** were found to interact with Asp64 and Asp116. The **NSC91705** formed two and 1 hydrogen bonding with Asp64 and Asp116 respectively, while 3 and 2 bump interactions with Asp64 and Asp116 respectively. Four hydrogen bond and three bump interactions were observed between **NSC651812** and Asp64 inside the HIV integrase receptor site. Asp116 was also found to be important to form one of each hydrogen bond and bump interactions with **NSC651812.** The dock score of **H13**, **NSC91705** and **NSC651812** were found to be 185.253, 194.48 and 217.798 respectively. The above observation indicated that both screened compounds formed more number of binding interactions than **H13** which suggested that **NSC91705** and **NSC651812** show more promising than **H13**.

Molecular dynamics

In order to analyse stability of molecular docked complexes of HIV-integrase with **H13**, **NSC91705** and **NSC651812** molecular dynamics studies were performed. In this purpose AMBER12 was used for a time span of 50,000ps. Backbone RMSD and RMSF were recorded to observe the complex constancy during simulation time and depicted in Fig. 8.

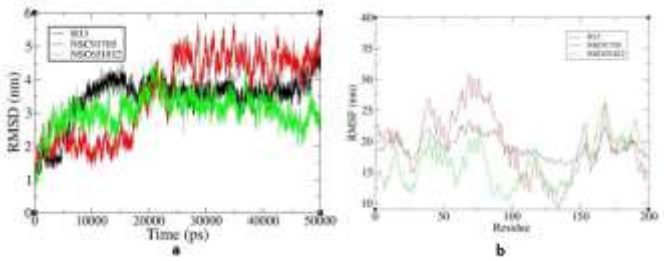

Fig. 8 Plot of a) RMSD vs simulation time and b) RMSF vs residue number.

In visual inspection of the simulation trajectories (Fig. 8a) it can be observed that complex with most active compound (**H13**) showed high RMSD values compared to the complexes with screened compounds (**NSC91705** and **NSC651812**) until 15ns of time span. Further it was noticed that RMSD of **NSC91705** increased and finally both **NSC91705** and **H13** achieved stability with new conformations in between RMSD of 4 and 5nm. RMSD curve for **NSC651812** indicated that initially it achieved high RMSD conformers but after 20ns of time span it formed stable conformation at about 2.5nm. Average and maximum RMSD values were observed as 3.457nm and 4.838nm, 3.577nm and 5.726nm, and 3.060nm and 4.640nm respectively for complexes with **H13**, **NSC91705** and **NSC651812** correspondingly. Minimum RMSD value of all complexes were found to be 0nm. In order to check fluctuation of residues of docked complexes the RMSF parameter were collected and analysed (Fig. 8b). Average RMSF of complexes with **H13**, **NSC91705** and **NSC651812** were found to be 20.372nm, 22.491nm and 23.043nm respectively, while differences between maximum and minimum RMSF were perceived as 12.333nm, 15.864nm and 18.253nm correspondingly. The results above indicated that the complex with **NSC651812** achieved stable conformation at low RMSD while complex with **NSC91705** reached stability at higher

RMSD value. The MD simulation study contradicts the docking results in the case of **NSC91705** but correlates with complexes with **H13** and **NSC651812**. The potential and total energies of the system were recorded during molecular dynamics simulation and plotted in Fig. 9.

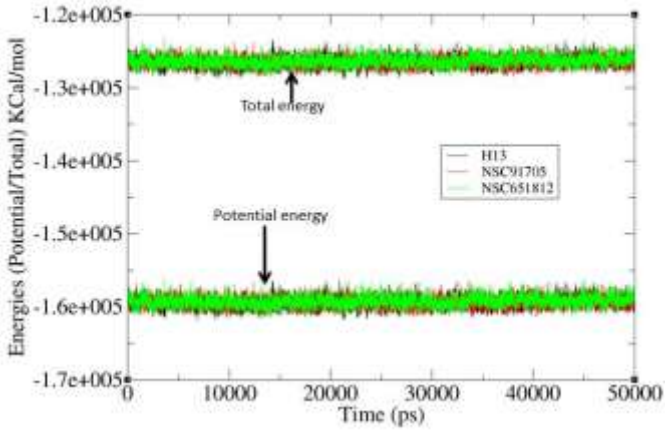

Fig. 9 Time vs energies (potential/total) in Kcal/mol

Lowest energy complexes with screened compounds and most active compound of the dataset were collected after MD simulation study. The binding mode of all three complexes is given in the Fig. 10. Both molecular docking and lowest energy complexes of MD simulation of screened compounds explain the importance Asp64 and Asp116 amino residues at the active site cavity. In case of binding mode of **H13** shows importance of Asp116 where molecular docking of this compounds revealed with importance of Glu52 and Asp116. Moreover polar amino residues Pro90, Tyr99, Lys103, Pro142, Tyr143, Asn144, pro145, Gln146, Ser147, Gln148, Glu152, Lys156 and Lys173, and non-polar Val88, Ile141, Gly149 and Val150 were found in the active site cavity and might be crucial for binding interactions. Importance of two HB acceptor and one ring aromatic sites in the pharmacophore can be correlated with binding mode of the most active and screened compounds. Presence of catalytic amino residues at the receptor site explained the possibility of formation of binding interactions with ligands that substantiated the finding of pharmacophore model.

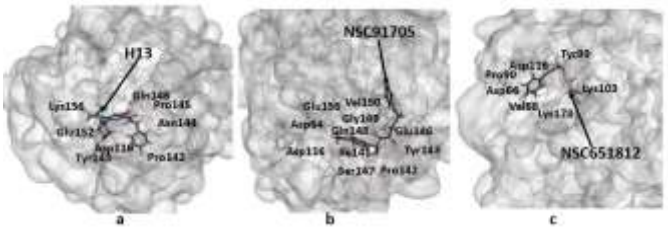

Fig. 10 Binding mode of a) **H13**, b) **NSC91705** and c) **NSC651812** in lowest energy conformation in MD simulation.

Comparison of drug-likeness with FDA approved HIV integrase inhibitors

Dolutegravir[55], **Elvitegravir**[56], **Raltegravir**[57] are FDA approved HIV-integrase inhibitors available in the market for the treatment of AIDS. In order to compare the drug-likeness of screened compounds with FDA approved HIV-integrase inhibitors different parameters of **H13**, **Dolutegravir**, **Elvitegravir**, **Raltegravir**, **NSC91705** and **NSC651812** were calculated and reported in Table 4. Parameters are included as dockscore, estimated activity, fit value, hydrophobicity, molecular weight, violation of Lipinski's rule of five, molecular volume, molecular refractivity, number of H-bonds and number of bump interactions.

Table 4: Comparative analysis of standard HIV-integrase inhibitors and screened compounds.

Molecules	[1]DS	[2]EA	[3]FV	logP	[4]MW	[5]vROF	[6]MV	[7]MR	[8]HBond	[9]Bump
H13	185.25	7.45	13.31	-1.11	404.44	0	287.70	103.70	1	1
Dolutegravir	143.41	12.79	13.08	-1.07	419.39	0	280.60	100.75	0	1
Elvitegravir	190.89	33.32	12.89	4.57	447.90	0	316.00	119.88	2	0
Raltegravir	222.09	19.73	12.89	-0.86	444.43	0	300.80	101.80	1	2
NSC91705	194.48	3.46	13.64	1.67	340.38	0	260.30	92.59	3	5
NSC651812	217.80	4.12	13.57	0.69	265.31	0	208.40	71.02	5	4

[1]Dockscore; [2]Estimated activity; [3]Fit value; [4]Molecular weight; [5]Violation of Lipinski's rule of five; [6]Molecular volume; [7]Molecular refractivity; [8]Number of H-bonds; [9]Number of bump interactions

It was observed that dockscore of both screened compounds is greater than **H13**, **Dolutegravir** and **Elvitegravir** but less than that of **Raltegravir**. The estimated activity and fit value were also predicted after mapping with pharmacophore model and both parameters showed that **NSC91705** and **NSC651812** have more promising than **Dolutegravir**, **Elvitegravir** and **Raltegravir**. The number of hydrogen

bond and bump interactions were also found to be much higher for **NSC91705** and **NSC651812** compared to **Dolutegravir**, **Elvitegravir** and **Raltegravir** when docked inside the receptor cavity of HIV-integrase. Moreover molecular weight, molecular volume and molecular refractivity were also found to be lower in the case of **NSC91705** and **NSC651812** compare to **H13**, **Dolutegravir, Elvitegravir** and **Raltegravir**. These findings show that the final screened molecules may be potential HIV-integrase inhibitors for the treatment for the wider community of HIV/AIDS.

Conclusion

Pharmacophore-based virtual screening studies were carried out to identify potential molecules for therapeutic application in HIV/AIDS. Several pharmacophore models were generated using 30 molecules out of 540 compounds in the whole dataset. After analysis of statistical parameters finally 10 hypotheses were considered for further validation. Hypotheses were validated using R^2_{pred}, s_p, $R^2_{m(test)}$, $\Delta r^2_{m(test)}$, Fischer's randomization and decoy set, and finally *Hypo1* was selected as the best model. The model indicated that HB acceptor and ring aromatic features were critical factors for the inhibitory activity. In order to find out excluded volumes, *Hypo1* was used for "hyporefine" and it found several excluded volumes that were important for HIV-integrase activity. Furthermore *Hypo1* and *Hypo1* with excluded volumes were used for virtual screening of NCI database. From the initial hits, finally two compounds (**NSC91705** and **NSC651812**) were found to be promising HIV-integrase inhibitors after screening with a number of criteria including mapping with omitted feature = 0, comparison of fitscore and estimated activity with most active compound of data set (**H13**), passing through Lipinski's rule of five and Veber's rule[53], comparison of dockscore with **H13**, ADMET and synthetic accessibility analysis. In the molecular docking study, a number of binding interactions were observed between final screened compounds and catalytic amino acid residues of HIV-integrase. The final potential compounds were subjected to 50,000ps molecular dynamics simulation to analyse binding stability in the receptor cavity of HIV-integrase. RMSD, RMSF, potential energy and total energy were recorded of the most active compound and final screened molecules. MD simulation study suggested that complex of **NSC651812** achieved stable conformation at lower RMSD of **H13** while **NSC91705** obtained at higher RMSD value of **H13**. Comparison of different parameters of screened compounds with FDA approved HIV integrase inhibitors confirmed that **NSC91705** and **NSC651812** were more promising than **Dolutegravir, Elvitegravir, Raltegravir**. Finally, from the discussion above, it can be postulated that the final screened compounds may be might be promising candidates for the treatment of HIV/AIDS but need further experimental validation for better confirmation.

Acknowledgement

MA Islam and TS Pillay were funded by the University of Pretoria Vice Chancellor's post-doctoral fellowship and National Research Foundation (NRF), South Africa Innovation post-doctoral fellowship schemes.

References

1. K. S. Rao, A. Ghorpade and V. Labhasetwar, *Expert opinion on drug delivery*, 2009, **6**, 771-784.
2. M. O. Akanbi, K. K. Scarsi, B. Taiwo and R. L. Murphy, *Expert opinion on pharmacotherapy*, 2012, **13**, 65-79.
3. Z. K. Sweeney and K. Klumpp, *Current opinion in drug discovery & development*, 2008, **11**, 458-470.
4. A. G. Tomasselli and R. L. Heinrikson, *Biochimica et biophysica acta*, 2000, **1477**, 189-214.
5. G. S. Balint, *Orvosi hetilap*, 1998, **139**, 1471-1474.
6. J. E. Cook, S. Dasgupta, L. D. Middaugh, E. C. Terry, P. R. Gorry, S. L. Wesselingh and W. R. Tyor, *Annals of neurology*, 2005, **57**, 795-803.
7. J. Y. Wang, H. Ling, W. Yang and R. Craigie, *The EMBO journal*, 2001, **20**, 7333-7343.
8. J. C. Chen, J. Krucinski, L. J. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt and R. M. Stroud, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 8233-8238.
9. M. Thomas and L. Brady, *Trends in biotechnology*, 1997, **15**, 167-172.
10. M. A. Sanchez-Olivas, M. P. Valencia-Zavala, G. B. Vega-Robledo, J. A. Sanchez-Olivas, G. Velazquez-Samano and G. Sepulveda-Velazquez, *Revista alergia Mexico*, 2015, **62**, 142-148.
11. S. P. Kumar, Y. T. Jasrai, V. P. Mehta and H. A. Pandya, *Journal of biomolecular structure & dynamics*, 2015, **33**, 706-722.
12. B. Sangeetha, R. Muthukumaran and R. Amutha, *SAR and QSAR in environmental research*, 2013, **24**, 753-771.
13. H. Bhatt, P. Patel and C. Pannecouque, *Chemical biology & drug design*, 2014, **83**, 154-166.

14. K. K. Reddy, S. K. Singh, N. Dessalew, S. K. Tripathi and C. Selvaraj, *Journal of enzyme inhibition and medicinal chemistry*, 2012, **27**, 339-347.
15. V. N. Telvekar and K. N. Patel, *Chemical biology & drug design*, 2011, **78**, 150-160.
16. P. Gupta, A. Sharma, P. Garg and N. Roy, *Current computer-aided drug design*, 2013, **9**, 141-150.
17. S. Kaushik, S. P. Gupta, P. K. Sharma and Z. Anwar, *Medicinal chemistry*, 2011, **7**, 553-560.
18. C. Laggner, C. Schieferer, B. Fiechtner, G. Poles, R. D. Hoffmann, H. Glossmann, T. Langer and F. F. Moebius, *Journal of medicinal chemistry*, 2005, **48**, 4754-4764.
19. T. Steindl, C. Laggner and T. Langer, *Journal of chemical information and modeling*, 2005, **45**, 716-724.
20. J. H. van Drie, *Current pharmaceutical design*, 2003, **9**, 1649-1664.
21. A. K. Debnath, *Journal of medicinal chemistry*, 2003, **46**, 4501-4515.
22. J. Wei, S. Wang, S. Gao, X. Dai and Q. Gao, *Journal of chemical information and modeling*, 2007, **47**, 613-625.
23. H. Li, J. Sutter and R. Hoffmann, *Pharmacophore Perception, Development, and Use in Drug Design*, International University Line, La Jolla, CA, 2000.
24. Discovery Studio, Accelrys Inc., San Diego, USA, 2014.
25. S. K. Middha, A. K. Goyal, S. A. Faizan, N. Sanghamitra, B. C. Basistha and T. Usha, *Journal of biosciences*, 2013, **38**, 805-814.
26. Q. A. Al-Balas, H. A. Amawi, M. A. Hassan, A. M. Qandil, A. M. Almaaytah and N. M. Mhaidat, *Pharmaceuticals*, 2013, **6**, 700-715.
27. D. Huang, X. Zhu, C. Tang, Y. Mei, W. Chen, B. Yang, J. Han, H. Qian and W. Huang, *Medicinal chemistry*, 2012, **8**, 1117-1125.
28. M. T. Chhabria, P. S. Brahmkshatriya, B. M. Mahajan, U. B. Darji and G. B. Shah, *Chemical biology & drug design*, 2012, **80**, 106-113.
29. H. Li, J. Sutter and R. Hoffman, in *Pharmacophore Perception, Development, and Use in Drug Design*, ed. O. F. Guner, International University Line, La Jolla, CA, 1999, pp. 173-189.
30. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187–217.
31. F. A. Momany and R. Rone, *Journal of Computational Chemistry*, 1992, **13**, 888–900.
32. A. Smellie, S. L. Teig and P. Towbin, *Journal of Computational Chemistry*, 1995, **16**, 171-187.
33. R. Kristam, V. J. Gillet, R. A. Lewis and D. Thorner, *Journal of chemical information and modeling*, 2005, **45**, 461-476.
34. B. R. Sadler, S. J. Cho, K. S. Ishaq, K. Chae and K. S. Korach, *Journal of medicinal chemistry*, 1998, **41**, 2261-2267.
35. H. Li, J. Sutter and R. Hoffman, *Pharmacophore Perception, Development, and Use in Drug Design*, International University Line, California, 2000.
36. H. Kubinyi, F. A. Hamprecht and T. Mietzner, *Journal of medicinal chemistry*, 1998, **41**, 2553-2564.
37. K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das and H. Kabir, *Journal of chemical information and modeling*, 2012, **52**, 396-408.
38. P. K. Ojha, I. Mitra, R. N. Das and K. Roy, *Chemometrics and Intelligent Laboratory Systems*, 2011, **107**, 194–205.
39. A. Golbraikh and A. Tropsha, *Journal of molecular graphics & modelling*, 2002, **20**, 269-276.
40. I. Mitra, A. Saha and K. Roy, *Journal of molecular modeling*, 2010, **16**, 1585-1596.
41. P. P. Roy, S. Paul, I. Mitra and K. Roy, *Molecules*, 2009, **14**, 1660-1701.
42. P. P. Roy and K. Roy, *QSAR & Combinatorial Science*, 2008, **27**, 302–313.
43. A. Cereto-Massague, L. Guasch, C. Valls, M. Mulero, G. Pujadas and S. Garcia-Vallve, *Bioinformatics*, 2012, **28**, 1661-1662.
44. S. S. Bhayye, K. Roy and A. Saha, *Letters in Drug Design & Discovery*, 2016, **13**.
45. A. C. Anderson, *Chemistry & Biology*, 2003, **10**, 787–797.
46. Y. Goldgur, R. Craigie, G. H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai and D. R. Davies, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 13040-13043.
47. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, Jr., *Journal of computational chemistry*, 2010, **31**, 671-690.
48. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Computational Molecular Science*, 2013, **3**, 198-220.
49. A. K. Debnath, *Journal of medicinal chemistry*, 2002, **45**, 41-53.
50. S. Sakkiah, S. Thangapandian, S. John, Y. J. Kwon and K. W. Lee, *European journal of medicinal chemistry*, 2010, **45**, 2132-2140.
51. P. Gupta, P. Garg and N. Roy, *Molecular diversity*, 2011, **15**, 733-750.
52. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Advanced drug delivery reviews*, 2001, **46**, 3-26.
53. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *Journal of medicinal chemistry*, 2002, **45**, 2615-2623.
54. M. O. Taha, M. Habash, Z. Al-Hadidi, A. Al-Bakri, K. Younis and S. Sisan, *Journal of chemical information and modeling*, 2011, **51**, 647-669.
55. *The Medical letter on drugs and therapeutics*, 2013, **55**, 77-79.
56. K. K. Pandey, *Hiv/Aids*, 2014, **6**, 81-90.
57. *AIDS patient care and STDs*, 2007, **21**, 288.