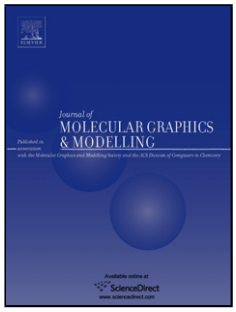
Accepted Manuscript




Title: Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification

Author: Md Ataul Islam Tahir S. Pillay

PII: S1093-3263(14)00203-4
DOI: http://dx.doi.org/doi:10.1016/j.jmgm.2014.11.015
Reference: JMG 6494

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 10-9-2014
Revised date: 24-11-2014
Accepted date: 30-11-2014

Please cite this article as: M.A. Islam, T.S. Pillay, Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification, *Journal of Molecular Graphics and Modelling* (2014), http://dx.doi.org/10.1016/j.jmgm.2014.11.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

# Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification

Md Ataul Islam, Tahir S. Pillay*

*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, 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: +27-123192155
Fax: +27-123283600

**Highlights:**

- A 3D-QSAR pharmacophore model was developed for HIV protease inhibitors.
- The 3D-QSAR pharmacophore model was validated using different statistical parameters.
- In order to identify hit molecules virtual screening of the NCI database was performed.
- A number of criteria were imposed on hit molecules to obtain the lead molecules.
- Seven compounds were identified as potential HIV-protease molecules, amongst which one has reportedly been confirmed as active agent for anti-HIV.
- Binding interactions between lead molecules and catalytic residues of HIV protease were analysed.

**Abstract**

Pharmacoinformatics approaches are widely used in the field of drug discovery as it saves time, investment and animal sacrifice. In the present study, pharmacophore based virtual screening was adopted to identify potential HIV-protease ligand as anti-HIV agents. Pharmacophore is the 3D orientation and spatial arrangement of functional groups that are critical for binding at the active site cavity. Virtual screening retrieves potential hit molecules from databases based on imposed criteria. A set of 30 compounds were selected with inhibition constant as training set from 129 compounds of dataset set and subsequently the pharmacophore model was developed. The selected



Accepted Manuscript

ACCEPTED MANUSCRIPT

best model consists of hydrogen bond acceptor and donor, hydrophobic and aromatic ring, features critical for HIV-protease inhibitors. The model exhibits high correlation ($R$ = 0.933), less *rmsd* (1.014), high cross validated correlation coefficient ($Q^2$ = 0.872) among the ten models examined and validated by Fischer's randomization test at 95% confidence level. The acceptable parameters of test set prediction, such as $R^2_{pred}$ = 0.768 and $r^2_{m(test)}$ = 0.711 suggested that external predictivity of the model was significant. The pharmacophore model was used to perform a virtual screening employing the NCI database. Initial hits were sorted using a number of parameters and finally seven compounds were proposed as potential HIV-protease molecules. One potential HIV-protease ligand is reportedly confirmed as an active agent for anti-HIV screening validating the current approach. It can be postulated that the pharmacophore model facilitates the selection of novel scaffold of HIV-protease inhibitors and also can be allow the design of new chemical entities.

**Keywords:** HIV protease inhibitors, Pharmacophore, Molecular Docking, Virtual screening

## 1. Introduction

Acquired immunodeficiency syndrome (AIDS) is an epidemic disease with an estimated two million deaths each year[1] and remains one of the world's most significant public health challenges, predominantly in low- and middle-income countries. The causative agent of the AIDS is human immunodeficiency virus type -1 (HIV-1)[2-5] which is characterized by extensive and dynamic genetic diversity[6] that has implications for the understanding of viral transmission, pathogenesis and diagnosis, and strongly influences strategies for vaccine development. The HIV polyprotein precursor is encoded by relatively simple genomes consisting of *gag*, *pol* and *env* open reading frames. The *gag* gene encodes the structural capsid, nucleocapsid, and matrix protein; *env* undergoes multiple alternative splicing events to regulatory protein; while, *pol* encodes essential viral enzymes necessary for viral replication. The HIV-1 protease receptor (HIV-1 PR) is an aspartyl protease that is required for proteolytic processing of the *gag* and *gag-pol* polyprotein precursors to yield the viral enzyme and structural proteins and is absolutely indispensible for proper viron assembly and maturation[7]. For this reason this protein is one of the major targets for the design of anti-HIV inhibitors[8] for the treatment of AIDS due to its critical role in virus maturation and replication. HIV-1 PR contains a homodimeric C-2 symmetric structure and each monomer contributes one catalytic aspartic residue along with threonine and glycine residues which are flexible and a flap that favors the binding of substrate and inhibitors. The highly active antiretroviral therapy (HAART) and protease inhibitors (PIs) along with reverse-transcriptase inhibitors have resulted in the unprecedented success of HIV/AIDS chemotherapy[9-12]. However owing to the rapid emergence of drug-resistant HIV-1 variants and transmission of these resistant

ACCEPTED MANUSCRIPT

viral strains along with the adverse side effects of currently used HIV-1 PIs, are remain critical factors that limiting the clinical effectiveness of HAART[13-15]. Numerous groups worldwide have developed HIV-1 protease inhibitors, showing excellent antiviral profiles[16-22]. Up to now, some clinically approved HIV-1 protease inhibitors including atazanavir, indivanir, nelfinavir and sequinavir are available in the market for HIV treatment but they are very peptide-like and have poor bio-availability. Therefore to overcome these problems, there is a need for the development of new PIs with improved activity against drug resistant variants and excellent pharmacokinetic and safety profiles. The pharmacoinformatics approaches including structure activity relationship (SAR), pharmacophore, virtual screening and molecular docking have become pivotal techniques in the pharmaceutical industry for lead discovery. Many groups have applied the pharmacoinformatics approaches to identify inhibitors[23-29] against HIV protease. Hence the current study explores the binding preferences of the inhibitory molecules of HIV protease in terms of space modelling study and virtual screening along with molecular docking.

A pharmacophore defined as ensemble of steric and electronic features that is required to ensure optimal supra-molecular interactions with a specific biological target and to trigger (or block) its biological response[30]. It also can be stated that the pharmacophore concept is based on the kinds of interaction observed in molecular recognition, *i.e.*, hydrogen bonding, charge, and hydrophobic interaction. The pharmacophore features: hydrogen bond acceptor (HBA) and donor (HBD), hydrophobic (H) and aromatic ring (R) were found to be the key features associated with the selectivity and potency of HIV protease inhibitors. The pharmacophore model can 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[31, 32]. In this study, an attempt was made to identify the pharmacophore hypothesis using the *HypoGen* algorithm[33] based on key chemical features of HIV-protease 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 significance of the work is clearly reflected by the identification of seven potent lead molecules as protease inhibitors. Among these seven potential HIV-protease ligand one compound is reportedly confirmed as an active anti-HIV agent, thus validating the approach.

## 2. Materials and methods

ACCEPTED MANUSCRIPT

The pharmacophore space modelling study is one of the most widely used and versatile techniques to discover novel scaffolds for various targets. Mainly, two types of pharmacophore modelling approaches can be used and adopted for searching novel active scaffolds, ligand-based and structure-based. In the present research ligand-based pharmacophore modelling approach was considered for a set of HIV protease inhibitors with inhibitory constant ($K_i$).

The Discovery Studio 3.5 (DS)[34] was used for the 3D QSAR pharmacophore, virtual screening and molecular docking studies. The DS is commercially available software containing several module packages and widely used in the pharmacoinformatics drug discovery[35-38]. The *3D QSAR Pharmacophore Generation* module enables the use of structure and activity data for a set of potential HIV-protease ligand to create hypotheses. Two algorithms, *HypoGen* and *HipHop* are used for ligand-based pharmacophore modelling. The *HypoGen* allows identification of hypotheses that are common to the 'active' compounds of training set but not present in the 'inactive' compounds, whilst *HipHop* identifies hypotheses present both in 'active' and 'inactive' compounds. In the present work the *HypoGen* algorithm was used to generate the hypotheses.

Accepted Manuscript

ACCEPTED MANUSCRIPT

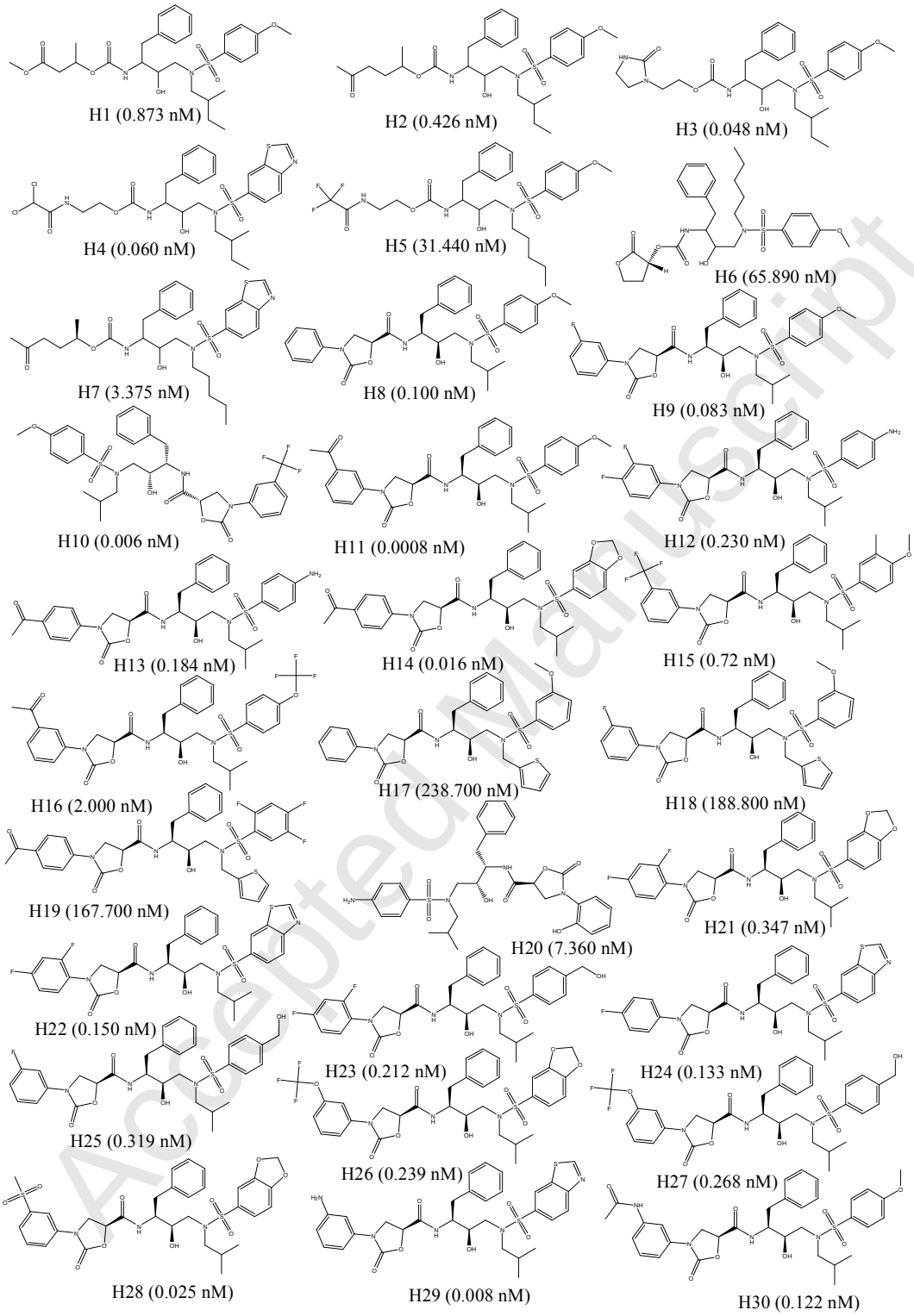



**Figure 1:** 2D chemical structures of the training set compounds and the activity values ($K_i$) are given in the parentheses.

*2.1 Dataset*

A dataset of 129 HIV protease inhibitors[39-41] were collected from literature. The experimental $K_i$ values were determined by the same group of authors using fluorescence resonance energy transfer

ACCEPTED MANUSCRIPT

(FRET) method. The whole dataset was divided into training and test set compounds for pharmacophore model generation and validation of generated model respectively. The molecules of the dataset have a wide range of $K_i$, from 0.0008 to 237.8 nM. The whole dataset was divided into three categories based on their activities values; highly active ($K_i$ < 1.000 nM, +++), moderately active (1.000 ≤ $K_i$ < 66.000 nM, ++) and least active ($K_i$ ≥ 66.000 nM, +). To select the training set for pharmacophore model in DS basic guidelines laid down by Li et al.[42] were followed. The guidelines are a) molecules should be selected to provide clear and concise information including structure features and activity range. b) a minimum of 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. d) the biological activity data of the molecules should have spanned at least 4 orders of magnitude. Based on the above criteria 30 compounds were selected as training set ($n_{tr}$ = 30, Figure 1) and the remaining 99 compounds (Table S1 in Supplementary file) were considered as test set ($n_{ts}$) compounds used for assessing the performance of pharmacophore model. The information concerning the structure and the biological activity of test set compounds is provided in the supplementary information, while all the data regarding the training set molecules are reported in Figure 1. The three-dimensional coordinates of the compounds were generated using the 2D/3D visualizer[34] of DS. For each compound, the geometries were corrected, atoms were typed and energy minimization was performed based on the modified CHARMm force field[43, 44]. The various protocols in the molecular modelling package, DS were utilized for 3D-QSAR modelling, virtual screening and molecular docking studies.

*2.2 Pharmacophore model generation*

The pharmacophore model was developed using *3D QSAR Pharmacophore Model Generation* module of DS. The training set molecules were considered to generate conformation by *Cat-Conf* program of the DS software package. The BEST method was applied during generation of multiple acceptable conformations which provides complete and improved coverage of conformational space by performing a rigorous energy minimization and optimizing the conformations in both torsional and Cartesian space using the poling algorithm[45]. The algorithm best conformer generation considers the arrangement in space of chemical features rather than simply the arrangement of atoms[46]. The *Feature mapping* was used to predict the favourable features for the highly active compounds of the dataset. Mapped features were considered as input features for model generation. Followed by the conformer generation, the algorithm also considers chemical features and conformers, and operates in two modes: *HipHop* and *HypoGen*. *HipHop* generates pharmacophore models using active compounds only, whereas *HypoGen* takes activity data into account and uses both active and inactive compounds in an attempt to identify a hypothesis that is common among

ACCEPTED MANUSCRIPT

the active compounds but not in the inactive compounds[46]. It builds top ten scoring hypothesis models with consideration of the training set, conformational models and chemical features by three steps: a constructive step, a subtractive step and an optimization step[47]. The constructive step generates hypotheses that are common among the most active compounds, while in subtractive step, the hypotheses that fit to the inactive compounds are removed. Finally, the optimization step attempts to improve the score of the remaining hypotheses by applying small perturbation[46, 48]. The best hypothesis was selected based on the high correlation coefficient (*R*), low root mean square deviation (*rmsd*), cost function analysis and good predictive ability.

*2.3 Validation of pharmacophore model*

Validation is an essential step of any *in-silico* model to judge the predictivity and applicability as well as robustness. In the current study, the developed pharmacophore hypotheses were validated by four different methods, (1) internal validation, (2) cost function analysis, (3) Fischer's randomization test and (4) test set prediction.

*2.3.1 Internal validation*

The best model was validated internally using the leave-one out (LOO) cross-validation method. In this procedure, one compound was randomly deleted from training set in each cycle and model regenerated using the rest of the compounds with the same parameters used in original model. The new generated model was used to predict the activity of deleted compound. The procedure was continued until all molecules of the training set were deleted and activity predicted. The LOO cross-validated correlation coefficient ($Q^2$) and error of estimation (*se*) were calculated based on predicted activity of training compounds as explained above. High $Q^2$ (>0.5) and low *se* (<0.5) explained better predictive ability [49].

Further to confirm the good predictive ability of the training set compounds, the modified $r^2$ ($r^2_{m(LOO)}$) developed by Roy *et al.*[50, 51] was calculated. The $r^2_{m(LOO)}$ is a measure of the degree of deviation of the predicted activity from the observed ones. It was reported that model may be considered with $r^2_{m(LOO)}$>0.5.

*2.3.2 Cost function analysis*

The statistical parameters employed for hypothesis generation were spacing, uncertainty, and weight variation. Spacing is a parameter representing the minimum inter-features distance that may be allowed in the resulting hypothesis. The weight variation is the level of magnitude explored by the hypothesis in where each feature signifies some degree of magnitude of the compound's activity. This varied in some cases from 1 to 2. In other cases, the default value of 0.3 is generally considered. The uncertainty parameter reflects the error of prediction and denotes the standard



Accepted Manuscript

ACCEPTED MANUSCRIPT

deviation of a prediction error factor called the error cost. In the present work, values of 1.5 to 3.0 were considered as the uncertainty parameter. The total cost function is minimized encompassing three terms, *viz.*, weight cost, error cost, and configuration cost. Weight cost is the value that increases as the weight variation in the model deviates from the input weight variation value. The deviation between the estimated activity of the molecule in the training set and their experimentally determined value is the error cost. A fixed cost depends on the complexity of the hypothesis space being optimized and is also denoted as the configuration cost. The configuration cost is equal to the entropy of hypothesis space and should have a value <17 for a good pharmacophore model. The hypogen algorithm also calculates the cost of a null hypothesis that assumes no relationship in the data, and the experimental activities are normally distributed about their mean. Accordingly, the greater the difference ($\Delta cost$) between the total and the null costs, it is more likely that the hypothesis does not reflect a chance correlation.

*2.3.3 Fischer's randomization test*

The Fischer's randomization test was used to ensure strong correlation between the chemical structures and the biological activity of the training set molecule. In this method, the biological activity was scrambled and assigned new values. Thereafter, the pharmacophore hypotheses were generated using the same features and parameters as those used to develop the original pharmacophore hypotheses. If the randomization run generates better correlation coefficient and/or better statistical parameters than the original hypothesis may be considered to be developed by chance. Depending upon the statistical significance randomization run produces a different number of spreadsheets. The statistical significance is given by following equation.

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

Where, $a$ defined as total number of hypotheses having a total cost lower than best significant hypothesis, whereas $b$ denoted by total number of *HypoGen* runs and random runs. In case of 95% confidence level 19 random spreadsheet are generated ($b$ = 20) and every generated spreadsheet is submitted to *HypoGen* using the same experimental conditions as the initial run. In the present study, the developed pharmacophore model was checked at 95% confidence level and produced 19 spreadsheets.

*2.3.4 Test set prediction*

The prediction of the test set compound is a crucial step in order to verify whether the pharmacophore model was able to accurately predict the activities of compound beyond the training

ACCEPTED MANUSCRIPT

set molecule. Consequently, in the present research 99 test compounds were predicted using the developed pharmacophore model by *Ligand Pharmacophore Mapping* protocol in DS, depicted in Table S1 (Supplementary file) and Figure 4 . External validation provides the ultimate proof of the true predictivity of the model, and the predictive capacity of the model was judged best on statistical parameters, $R^2_{pred}$ (correlation coefficient) and $s_p$ (error of prediction). The threshold value of $R^2_{pred}$ is ≥0.5, whereas for $s_p$ it is ≤0.5[52, 53].

The $R^2_{pred}$ value depends on the mean observed activity of the training set compounds. Consequently high values of this parameter may be obtained for compounds with a wide range of activity data, but this may not indicate that the predicted activity values are very close to those observed. Though a good overall correlation is maintained, there may be a significant numerical difference between the two values. In order to better indicate the predictive ability of the model, modified $r^2$ [$r^2_{m(test)}$][54, 55] values were calculated (threshold value=0.5).

*2.4 Virtual screening*

Virtual screening is a vital technique which is used to identify novel potent compounds that can repress or trigger the activity of a particular target. In this work, the best developed pharmacophore model was considered for 3D query in the NCI (National Cancer Institute) database screening to retrieve the novel scaffold for HIV protease inhibitors. The NCI database contains 265,242 compounds. The filtered compounds were screened with a number of criteria. The flow diagram of the screening protocol is given in the Figure 2. Biological activity ($K_i$) of the retrieved compounds was estimated using the pharmacophore model and only those compounds whose activity falls within range of the training set compounds were selected for next step. Further the compounds were fitted in the pharmacophore model using maximum omitted feature set to '0'. Lipinski's rule of five[56] was applied and then molecular docking was performed. The dock score and binding energy of the compounds were compared with most active compound of the dataset. Finally the binding interactions were observed between the potential HIV-protease compounds and catalytic residues of the active site.

ACCEPTED MANUSCRIPT

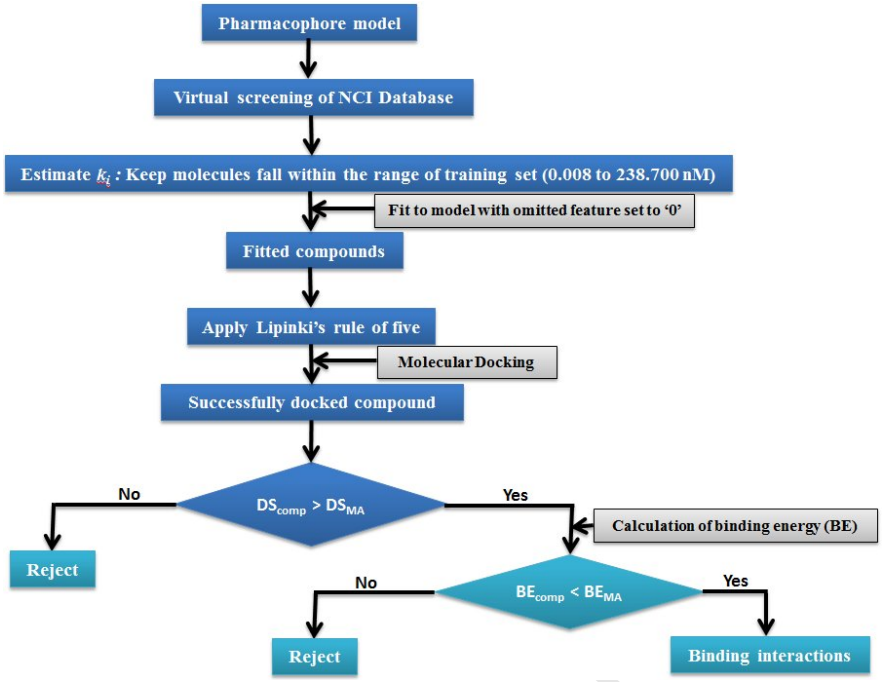



**Figure 2:** Schematic representation of the virtual screening protocol.

*2.5 Molecular docking*

In order to understand how the screened drug-like virtual hits bind to the receptor, potential HIV-protease ligands were analysed using the ligand-receptor interactions by molecular docking. Molecular docking is one of the best filtering methods and a crucial technique in drug design process. The *LigandFit* protocol of the DS was used to dock the retrieved compounds by virtual screening. The *LigandFit* protocol first detects the cavity to identify and select the region of the protein as the active site, and secondly dock the ligands to the selected site. 3D regular grids of points are employed for site detection and also for estimating the interaction energy of the ligand with the protein during docking. The protein receptor of the HIV protease was selected from RCSB Protein Data Bank (RCSB-PDB) for the molecular docking study. Among several HIV protease inhibitors PDB ID: 1T3R was selected based on the receptor size, resolution and deposited date. Both receptor and ligands were prepared using standard tools *Prepare Protein* and *Prepare Ligand* of DS respectively. Both protein and ligand were minimized using CHARMm force field. The 'Build Loop' and 'Protonate' parameters were set to 'True' while, dielectric constant, pH, ionic strengths and energy cut-off were considered as default value for the protein preparation. In case of ligand preparation 'Change ionization', 'Generate Tautomers' and 'Generate isomers' were set to 'False', and 'Generate Coordinates' was set to '3D'. The binding site was identified after protein preparation based on the volume occupied by the ligand in protein-ligand complex. The validation of docking protocol is essential to avoid the false positive results of molecular docking. In order to

ACCEPTED MANUSCRIPT

validate docking parameters, the co-crystal from PDB was initially sketched and docked into the active site HIV protease. The docked pose was checked to see whether it was able to produce the hydrogen-bonding interactions with the critical amino acids. The RMSD between the docked pose and the co-crystal was calculated to determine if the docking parameters were able to reproduce a conformation comparable to that of the co-crystal at the active site of HIV protease. Then the potential molecules were docked using the same parameters as in the co-crystal docking. During the docking process, the top ten conformations for each ligand based on the dock score after energy minimization using the smart minimizer method (which begins with the steepest descent method and is followed by the conjugate gradient method) were assessed. The docked poses were validated based on the hydrogen-bonding interactions between the candidate molecules and the active site residues.

**3. Results and discussion**

The pharmacophore model was developed based on training set ($n_{tr}$ = 30) compounds by the *HypoGen* algorithm. The training set molecules are listed in the Figure 1 (**H1** – **H30**) and the activity values ($K_i$) is shown within the parentheses. Based on the *Feature mapping* protocol of DS, 'HBA', 'HBD', 'H' and 'R' features were selected for as required chemical features and were used as input to the 3D QSAR pharmacophore generation module. Top ten hypotheses were generated with fixed and null cost values 219.780 and 123.574 respectively. The statistical parameters derived based on the activity of the training compounds are given in the Table 1. Debnath's[31, 57] analysis states that the best pharmacophore model should have the lowest cost value, highest cost difference, smallest *rmsd*, and best correlation coefficient. The predictive power of first hypothesis (*Hypo 1*) was confirmed by Debnath's method[31, 57]. A valid hypothesis should have the overall cost of the hypothesis far from the null cost and close to the fixed cost. It is elicited that the difference in cost (Δcost) is the difference between the null cost and the total cost of the hypothesis; a cost difference of 40–60 bits leads to a predictive correlation probability of 75–90%, and if the difference is greater than 60 bits, the hypothesis is considered to have a correlation probability of greater than 90%[58]. In the current study the cost difference for *Hypo1* (Table 1) was found to be 78.524 that is more than 60 bits, indicating that this hypothesis has a >90% chance of being able to select HIV inhibitors.

ACCEPTED MANUSCRIPT

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

| Hypo No. | Spacing | [1]Unc. | [2]Wt. Var. | [3]*R* | [4]*Rmsd* | Costs | | | | | Output features |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Total | Null | Fixed | [5]Δ | [6]Config. | |
| 1 | 100 | 3 | 1.5 | 0.933 | 1.014 | 141.256 | 219.780 | 123.574 | 78.524 | 16.521 | a, d, p, r |
| 2 | 100 | 3 | 1.5 | 0.886 | 1.304 | 151.428 | 219.780 | 123.574 | 68.352 | 16.521 | a, p, r, r |
| 3 | 100 | 3 | 1.5 | 0.884 | 1.315 | 152.253 | 219.780 | 123.574 | 67.527 | 16.521 | a, a, d, p |
| 4 | 100 | 3 | 1.5 | 0.860 | 1.436 | 155.628 | 219.780 | 123.574 | 64.152 | 16.521 | a, d, p, r |
| 5 | 100 | 3 | 1.5 | 0.867 | 1.407 | 156.226 | 219.780 | 123.574 | 63.554 | 16.521 | a, d, p, r |
| 6 | 100 | 3 | 1.5 | 0.852 | 1.476 | 157.632 | 219.780 | 123.574 | 62.148 | 16.521 | a, d, p, r |
| 7 | 100 | 3 | 1.5 | 0.846 | 1.502 | 159.068 | 219.780 | 123.574 | 60.712 | 16.521 | a, a, p, r |
| 8 | 100 | 3 | 1.5 | 0.837 | 1.542 | 161.584 | 219.780 | 123.574 | 58.196 | 16.521 | a, d, p, r |
| 9 | 100 | 3 | 1.5 | 0.833 | 1.558 | 161.603 | 219.780 | 123.574 | 58.177 | 16.521 | a, d, p, r |
| 10 | 100 | 3 | 1.5 | 0.827 | 1.582 | 161.678 | 219.780 | 123.574 | 58.102 | 16.521 | a, d, p, 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 = R

The correlation coefficient (*R*) of all the hypotheses were determined and it was observed that all of the hypotheses had correlation coefficients of >0.800, but the best hypothesis presented the highest correlation coefficient (0.933), which demonstrates the good predictive ability of the selected hypothesis. The fixed and total cost values for *Hypo1* were found to be 123.574 and 141.256, respectively, whereas the difference between total and null cost found as 78.524. The highest cost difference and good correlation along with low *rmsd* and minimum error values were observed for *Hypo1* (Table 1) when compared with the other hypotheses. Hence, *Hypo1* was selected as the best hypothesis and employed in further analyses.

The best model (Figure 3) consisted one of each 'HBA', 'HBD', 'H', and 'R' feature and depicted in Figure 3 mapped with most active compound of the dataset. Apart from the cost analyses, the merit of hypothesis was justified by its ability to predict the activity of individual compounds within the set. For this purpose, the training set compounds were approximately classified into three different categories: highly active ($K_i$ < 1 nM, +++), moderately active (1 nM ≤ $K_i$ < 66 nM, ++) and least active/inactive ($K_i$ > 66 nM, +). All the compounds in the training set were accurately predicted with low error values between the experimental and estimated $K_i$. 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 activities of the compounds estimated by *Hypo1* were close to the corresponding experimental $K_i$ values, and the error values defined as the ratio between the experimental and estimated activity values which demonstrated remarkable consistency between the estimated and experimental $K_i$ values. It was observed that molecule **H16** and **H20** under estimated,

ACCEPTED MANUSCRIPT

and **H1** overestimated the activity when mapped to the pharmacophore model. Possible reason of such depraved estimation of compound **H16** and **H20** was presence of electronegative three fluorine atoms and one amino functional group respectively in the scaffold that withdrawn electrons towards itself. Overestimation of **H1** might be due to possible 3D orientation of the molecule. The experimental and predicted activities of *Hypo1* for the training set compounds are shown in Table 2 and Figure 4. It was observed that best hypothesis gives $Q^2$ of 0.872 and *se* of 0.485. The high $Q^2$ and low *se* of the selected model suggested that model is robust in nature.

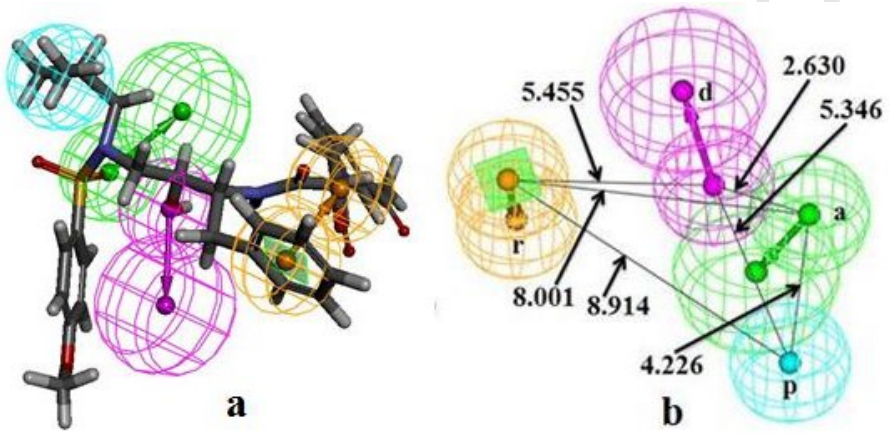



**Figure 3:** a) Mapped pharmacophore features (*Hypo 1*) with most active compound (**H11**); b) Inter-feature distances of *Hypo 1*, a = HBA, d =HBD, p = H, r = R

The best model (Figure 3a, *Hypo 1* in Table 2) mapped with most active compound of the dataset suggested that hydroxyl group present in the molecule behaved as HBD, whereas oxy group attached to sulphur atom revealed as HBA. The bulky group, iso-butane present in the molecular scaffold imparted the hydrophobicity of the molecule. The phenyl ring was critical for the aromatic ring features. The best selected hypothesis was validated to nullify over prediction of the bioactivity for inactive compounds through hyporefine. In this process, the steric interaction of the compounds was considered in hypothesis generation, but this factor was not portrayed in the validated (refine) hypothesis. This indicated the presence of steric hindrance of the molecule has not direct influence on inhibitory constant to HIV protease. Therefore it can be postulated that to design or synthesize new chemical entities of HIV-protease inhibitors HBA, HBD, H and R factors with critical inter-feature distances (Figure 3b) are to be crucial factors.

ACCEPTED MANUSCRIPT

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

| Comp. | Activity ($K_i$ nM) | | Error | Activity scale | | Fit value |
|---|---|---|---|---|---|---|
| | [1]Obs. | [2]Pred. | | [1]Obs. | [2]Pred. | |
| **H1** | 0.8730 | 1.5868 | +1.818 | +++ | ++ | 16.901 |
| **H2** | 0.4260 | 0.7933 | +1.862 | +++ | +++ | 17.203 |
| **H3** | 0.0480 | 0.1710 | +3.563 | +++ | +++ | 17.869 |
| **H4** | 0.0600 | 0.0311 | -0.518 | +++ | +++ | 18.609 |
| **H5** | 31.4400 | 3.1213 | -10.073 | ++ | ++ | 16.608 |
| **H6** | 65.8900 | 33.8476 | -1.947 | ++ | ++ | 15.572 |
| **H7** | 3.3750 | 4.5427 | +1.346 | ++ | ++ | 16.445 |
| **H8** | 0.1000 | 0.2618 | +2.618 | +++ | +++ | 17.684 |
| **H9** | 0.0830 | 0.2133 | +2.570 | +++ | +++ | 17.773 |
| **H10** | 0.0060 | 0.0039 | -1.550 | +++ | +++ | 19.514 |
| **H11** | 0.0008 | 0.0026 | +3.267 | +++ | +++ | 19.685 |
| **H12** | 0.2300 | 0.1342 | -1.714 | +++ | +++ | 17.974 |
| **H13** | 0.1840 | 0.3143 | +1.708 | +++ | +++ | 17.605 |
| **H14** | 0.0160 | 0.0259 | +1.621 | +++ | +++ | 18.688 |
| **H15** | 0.0720 | 0.0044 | -16.414 | +++ | +++ | 19.460 |
| **H16** | 2.0000 | 0.3858 | -5.184 | ++ | +++ | 17.516 |
| **H17** | 238.7000 | 101.8500 | -2.344 | ++ | ++ | 15.094 |
| **H18** | 188.8000 | 128.8150 | -1.466 | ++ | ++ | 14.992 |
| **H19** | 167.7000 | 157.8480 | -1.062 | ++ | ++ | 14.904 |
| **H20** | 7.3600 | 0.8676 | -8.484 | ++ | +++ | 17.164 |
| **H21** | 0.3470 | 0.2271 | -1.528 | +++ | +++ | 17.746 |
| **H22** | 0.1500 | 0.3808 | +2.539 | +++ | +++ | 17.521 |
| **H23** | 0.2120 | 0.8762 | +4.133 | +++ | +++ | 17.159 |
| **H24** | 0.1330 | 0.1103 | -1.206 | +++ | +++ | 18.059 |
| **H25** | 0.3190 | 0.6859 | +2.150 | +++ | +++ | 17.266 |
| **H26** | 0.2390 | 0.2409 | +1.008 | +++ | +++ | 17.720 |
| **H27** | 0.2860 | 0.1568 | -1.824 | +++ | +++ | 17.907 |
| **H28** | 0.0250 | 0.0090 | -2.770 | +++ | +++ | 19.146 |
| **H29** | 0.0080 | 0.0239 | +2.993 | +++ | +++ | 18.723 |
| **H30** | 0.1220 | 0.5399 | +4.425 | +++ | +++ | 17.370 |

[1]Observed; [2]Predicted

*3.1 Test set prediction*

The robust model should have the capability to predict the activities of the compounds which were not involved in model generation. Altogether, 99 compounds (Table S1 in supplementary file and Figure 4) were considered for test set and classified based on their activity values: highly active ($K_i$ < 1 nM, +++), moderately active (1 nM ≤ $K_i$ < 66 nM, ++) and least active/inactive ($K_i$ > 66 nM,



Accepted Manuscript

ACCEPTED MANUSCRIPT

+). The activity of the test set compounds was estimated by *Ligand Pharmacophore Mapping* implemented in DS. Comparison between observed and estimated activity revealed that three active compounds were overestimated as moderately active and remaining compounds classified correctly which suggested that selected model was able to provide accurate estimates for activities of external compounds. The correlation ($R$) between observed and estimated activity of test compounds showed 0.884 and the $R^2_{pred}$ value of 0.768 with error of prediction ($s_p$) of 0.564.

For a better determination of the predictive abilities of the models, the values of $r^2_{m(test)}$ were also calculated. The value of this parameter determines whether the predicted activity values are close to the corresponding observed ones, since a high value of $R^2_{pred}$ may not always indicate a low residual between the observed and predicted activity data. Among all of the hypotheses developed, the largest value of $r^2_{m(test)}$ (0.711) was observed for *Hypo 1*, indicating that this model has acceptable predictive potential. Thus, the results suggested that the selected pharmacophore can reasonably predict the activities of new compounds.

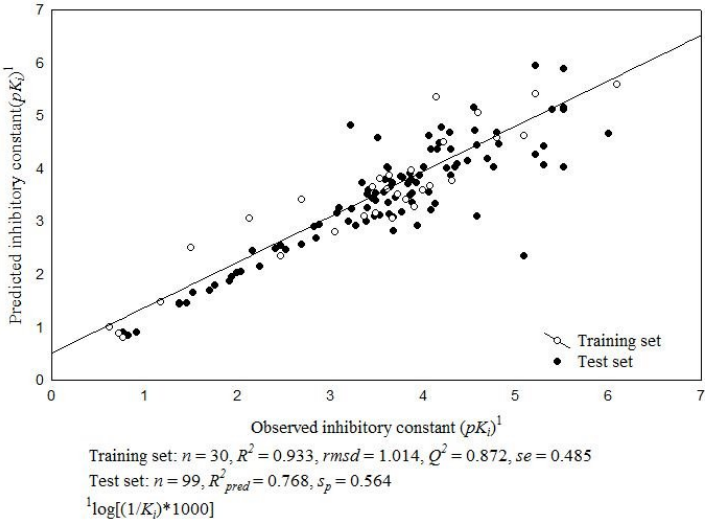



**Figure 4:** Observed and predicted inhibitory constant of HIV protease inhibitors based on *Hypo 1*.

*3.2 Fischer randomization test*

The Fischer randomization test was used to check the statistical relevance of the hypothesis of interest by assigning a particular confidence level. In the present case confidence level was set to 95%, so 19 random spreadsheets were created by shuffling the experimental activity values of the training set compounds, and a hypothesis was generated for each spreadsheet. The significance of the hypothesis was calculated as per equation (1). The correlation of 19 spreadsheets is depicted in the Table 3 which indicates that none of the values generated after randomization produced hypotheses that exhibited predictive powers similar to that of *hypo 1* (Table 1).

ACCEPTED MANUSCRIPT

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

| Validation | Correlation | Total cost |
|---|---|---|
| **random1** | 0.642 | 193.908 |
| **random2** | 0.711 | 185.811 |
| **random3** | 0.792 | 170.670 |
| **random4** | 0.862 | 155.609 |
| **random5** | 0.800 | 169.458 |
| **random6** | 0.694 | 186.072 |
| **random7** | 0.757 | 178.068 |
| **random8** | 0.717 | 183.472 |
| **random9** | 0.737 | 180.122 |
| **random10** | 0.805 | 168.089 |
| **random11** | 0.786 | 174.434 |
| **random12** | 0.796 | 169.971 |
| **random13** | 0.852 | 167.504 |
| **random14** | 0.774 | 172.581 |
| **random15** | 0.768 | 173.457 |
| **random16** | 0.856 | 159.609 |
| **random17** | 0.665 | 192.395 |
| **random18** | 0.777 | 174.247 |
| **random19** | 0.870 | 157.111 |

The average of correlation coefficient for all 19 trials was about 0.772. 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 Figure 5 and Table 3. The above discussion clearly shows that the selected pharmacophore model was not generated by chance.

Accepted Manuscript

ACCEPTED MANUSCRIPT

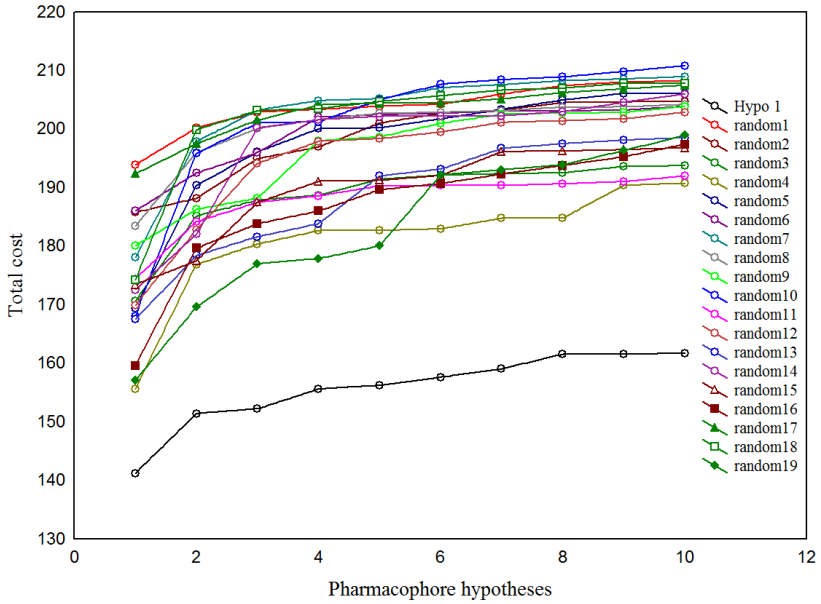



**Figure 5:** Comparison of the total costs of *Hypo 1* and 19 random hypotheses generated in the Fischer's randomization test.

*3.3 Virtual screening*

Virtual screening is the powerful technique for potent molecules identification and an effective alternative to high-throughput screening methodologies. The validated QSAR, pharmacophore or molecular docking models are used to screen the molecular database for lead discovery. In the present research, pharmacophore model (*Hypo 1* in Table 1) was used to search the NCI database (consisting of 265,242 compounds) for potential HIV protease 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 'all' respectively. The flow diagram of the screening protocol is depicted in Figure 2. Initially the *Hypo 1* screening revealed 26548 compounds. The biological activity of all compounds was estimated by *Ligand Pharmacophore Mapping* protocol of DS. It was found that 1268 compounds were in the range of activity 0.008nM to 238.700nM, which is the range of activity of training set molecules. These were further considered for screening. These compounds fitted with *Hypo 1* using omitted feature set to '0' and it was observed that 1241 compound successfully fitted with all the features but 27 compounds failed to map all pharmacophoric features. These molecules were further considered to check Lipinski's rule of five[56] and observed that 493 compounds failed to pass the rule. The remaining 748 compounds were considered for molecular docking study using the *LigandFit* module of DS, where 710 compounds were successfully docked. The most active compound of the training set was also docked and binding energy calculated along with docked 710 compounds. DockScore of 10 molecules out of 710 were found to be greater than DockScore of most active compounds. Among

ACCEPTED MANUSCRIPT

these 10 molecules seven molecules (**NSC70804**, **NSC126464**, **NSC138969**, **NSC163404**, **NSC179371**, **NSC216959** and **NSC351981**) were found to have less binding energy than the most active compound. Consequently these seven molecules (Figure 6) were considered as potential HIV-protease molecules and further subjected to assess the critical interactions with the catalytic amino acid residues present in the active site cavity of the HIV protease. The *Hypo 1* screening identified one potential HIV-protease ligand, **NCS70804** which is already reported as an active molecule in anti-HIV screening confirming the validity of the model.

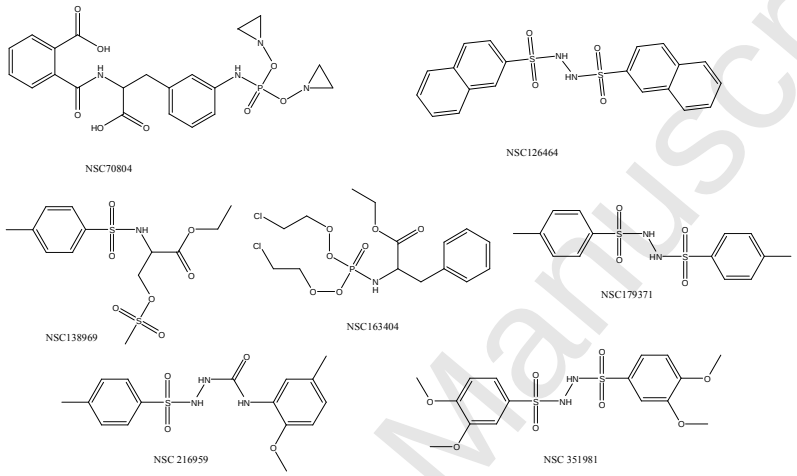



**Figure 6:** Screened lead compounds from NCI database.

*3.4 Molecular docking*

The molecular docking is used in order to find out the accurate and preferred orientations of the ligand at the active site of the protein. The docked complexes of the potential HIV-protease ligands screened from NCI database (Figure 6) and most active compound (**H11** in Figure 1) of the training set were considered to assess the optimal orientation and binding abilities. The crystal structure of HIV protease (PDB ID: 1T3R) was collected from RCSB-Protein Data Bank. Self-docking is one of the approaches to validate the molecular docking method[59] 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 reported that low RMSD (<2 Å) value of original bound ligand validates the docking procedure[59]. In the present study, self-docking approach was considered to validate the docking procedure. In this approach the freshly docked complex of bound compound with the receptor was superimposed with original complex downloaded from RCSB PDB (PDB ID: 1T3R). The RMSD values was found 0.00Å, which indicated that the protocol was selected in the docking method was validated.



Accepted Manuscript

ACCEPTED MANUSCRIPT

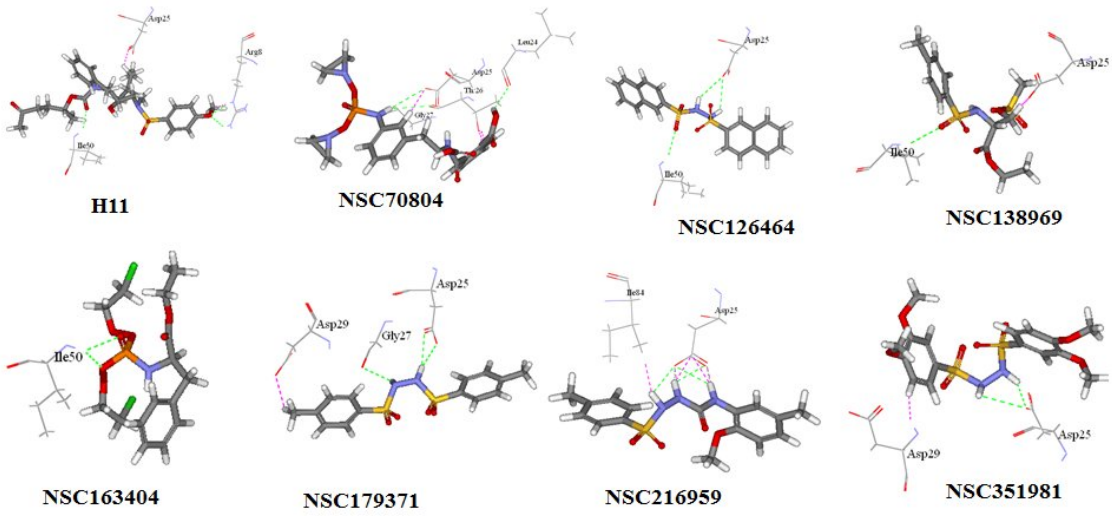



**Figure 7:** Binding modes of most active compound of the training set and NCI database screened potential HIV-protease molecules.

The docked complex (Figure 7) between most active compound of the training set and HIV protease revealed that Arg8, Asp25 and Ile50 were catalytic amino residues. Two hydrogen bonds and one bump interactions were observed between the ligand and Arg8. The Asp25 interacted with the ligand by forming one bump interactions with the ligand. Ile50 clashed with the ligand by one potential hydrogen bonding. All potential HIV-protease compounds showed strong interactions with catalytic residues such as Leu24, Asp25, Thr26, Gly27, Asp29, Ile50 and Ile84 at the active site cavity of HIV protease. As mentioned among these seven potential HIV-protease ligands **NSC70804** was listed as an active compound in anti-HIV screening.

## 4. Conclusion

The work presented here was performed to achieve the structural and orientational factors important for HIV protease inhibitors, to focus on the pharmacophore-based virtual screening of molecular database and propose the potential HIV-protease compounds for the anti-HIV agents. A number of hypotheses were generated based on 30 training set compounds and finally ten hypotheses were retained for further evaluation. Based on Debnath's analysis, Fischer's randomization test, test set prediction and Roy's $r^2_m$ prediction the *Hypo 1* was selected as best pharmacophore model. This model was suggested that one of each HBA, HBD, H and R features were crucial for inhibitory activity. The *Hypo 1* was selected to for the virtual screening of NCI database to retrieve some potential and less toxic anti-HIV agents. Initial screened compounds were passed through several criteria including the range of activity of training set, fit to model with omitted feature set to '0'and comparison of dock score and binding energy with most active compound of the data set to reach the potential compounds. Among the more than twenty five thousand initial hit compounds, seven

ACCEPTED MANUSCRIPT

compounds finally satisfied all these criteria and were used to further evaluate the binding interactions with critical amino residues at the active site of HIV protease. This virtual screening using the *Hypo 1* obtained one potential molecule, **NCS70804** from NCI database which is already confirmed as active in anti-HIV screening. A number of hydrogen bonds and bump interactions were observed between potential HIV-protease compounds and catalytic residues such as Leu24, Asp25, Thr26, Gly27, Asp29, Ile50 and Ile84. Therefore this indirect receptor-independent pharmacophore model can calculate accurately the position and orientation of potential ligand in a binding site and the selected model has enough potential to retrieve the active molecule from database. The potential HIV-protease molecules will be subjected to experimental validation to obtain further confirmation.

**Acknowledgement**

MA Islam and TS Pillay were funded by the University of Pretoria Vice Chancellor's post-doctoral fellowship scheme.

**References**

[1] Report on the Global AIDS Epidemic; UNAIDS. In: The Joint United Nations Program on HIV/AIDS (UNAIDS), Geneva, Switzerland, 2008.
[2] Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983, 220, 868-71.
[3] Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science. 1983, 220, 865-7.
[4] Popovic, M., Sarngadharan, M.G., Read, E., Gallo, R.C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 1984, 224, 497-500.
[5] Sarngadharan, M.G., Popovic, M., Bruch, L., Schupbach, J., Gallo, R.C. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. Science. 1984, 224, 506-8.
[6] Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. Aids. 2006, 20, W13-23.
[7] Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature. 1989, 337, 615-20.
[8] Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., et al. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science. 1989, 245, 616-21.
[9] Bartlett, J.A., DeMasi, R., Quinn, J., Moxham, C., Rousseau, F. Overview of the effectiveness of triple combination therapy in antiretroviral-naive HIV-1 infected adults. Aids. 2001, 15, 1369-77.
[10] Gulick, R.M., Mellors, J.W., Havlir, D., Eron, J.J., Meibohm, A., Condra, J.H., et al. 3-year suppression of HIV viremia with indinavir, zidovudine, and lamivudine. Annals of internal medicine. 2000, 133, 35-9.



Accepted Manuscript

ACCEPTED MANUSCRIPT

[11] Palella, F.J., Jr., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. The New England journal of medicine. 1998, 338, 853-60.
[12] Hogg, R.S., Heath, K.V., Yip, B., Craib, K.J., O'Shaughnessy, M.V., Schechter, M.T., et al. Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. JAMA : the journal of the American Medical Association. 1998, 279, 450-4.
[13] Waters, L., Nelson, M. Why do patients fail HIV therapy? International journal of clinical practice. 2007, 61, 983-90.
[14] Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., et al. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature. 1995, 374, 569-71.
[15] Clavel, F., Hance, A.J. HIV drug resistance. The New England journal of medicine. 2004, 350, 1023-35.
[16] Tomasselli, A.G., Heinrikson, R.L. Targeting the HIV-protease in AIDS therapy: a current clinical perspective. Biochimica et biophysica acta. 2000, 1477, 189-214.
[17] Reddy, G.S., Ali, A., Nalam, M.N., Anjum, S.G., Cao, H., Nathans, R.S., et al. Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres. Journal of medicinal chemistry. 2007, 50, 4316-28.
[18] Ghosh, A.K., Chapsal, B.D., Weber, I.T., Mitsuya, H. Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. Accounts of chemical research. 2008, 41, 78-86.
[19] Wensing, A.M., van Maarseveen, N.M., Nijhuis, M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. Antiviral research. 2010, 85, 59-74.
[20] Ghosh, A.K., Chapsal, B.D., Baldridge, A., Steffey, M.P., Walters, D.E., Koh, Y., et al. Design and synthesis of potent HIV-1 protease inhibitors incorporating hexahydrofuropyranol-derived high affinity P(2) ligands: structure-activity studies and biological evaluation. Journal of medicinal chemistry. 2011, 54, 622-34.
[21] Ganguly, A.K., Alluri, S.S., Caroccia, D., Biswas, D., Wang, C.H., Kang, E., et al. Design, synthesis, and X-ray crystallographic analysis of a novel class of HIV-1 protease inhibitors. Journal of medicinal chemistry. 2011, 54, 7176-83.
[22] Qiu, X., Zhao, G.D., Tang, L.Q., Liu, Z.P. Design and synthesis of highly potent HIV-1 protease inhibitors with novel isosorbide-derived P2 ligands. Bioorganic & medicinal chemistry letters. 2014, 24, 2465-8.
[23] Steindl, T.M., Schuster, D., Laggner, C., Chuang, K., Hoffmann, R.D., Langer, T. Parallel screening and activity profiling with HIV protease inhibitor pharmacophore models. Journal of chemical information and modeling. 2007, 47, 563-71.
[24] Pandit, D., So, S.S., Sun, H. Enhancing specificity and sensitivity of pharmacophore-based virtual screening by incorporating chemical and shape features--a case study of HIV protease inhibitors. Journal of chemical information and modeling. 2006, 46, 1236-44.
[25] Yadav, D., Paliwal, S., Yadav, R., Pal, M., Pandey, A. Identification of novel HIV 1--protease inhibitors: application of ligand and structure based pharmacophore mapping and virtual screening. PloS one. 2012, 7, e48942.
[26] Soliman, M.E.S. A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors. Drug Development Research. 2013, 74, 283–95.
[27] Khedkar, V.M., Ambre, P.K., Verma, J., Shaikh, M.S., Pissurlenkar, R.R., Coutinho, E.C. Molecular docking and 3D-QSAR studies of HIV-1 protease inhibitors. Journal of molecular modeling. 2010, 16, 1251-68.
[28] Saranya, N., Selvaraj, S. QSAR studies on HIV-1 protease inhibitors using non-linearly transformed descriptors. Current computer-aided drug design. 2012, 8, 10-49.
[29] Ibrahim, M., Saleh, N.A., Elshemey, W.M., Elsayed, A.A. Fullerene derivative as anti-HIV protease inhibitor: molecular modeling and QSAR approaches. Mini reviews in medicinal chemistry. 2012, 12, 447-51.
[30] Wermuth, C.G., Langer, T. Pharmacophore identification. In: 3D QSAR in drug design: theory, methods and applications, H, K. Ed., Kluwer, Dordrecht,, 2000, pp. 117–49.
[31] Debnath, A.K. Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. Journal of medicinal chemistry. 2003, 46, 4501-15.

ACCEPTED MANUSCRIPT

[32] Wei, J., Wang, S., Gao, S., Dai, X., Gao, Q. 3D-pharmacophore models for selective A2A and A2B adenosine receptor antagonists. Journal of chemical information and modeling. 2007, 47, 613-25.
[33] Li, H., Sutter, J., Hoffmann, R. Pharmacophore Perception, Development, and Use in Drug Design. La Jolla, CA, International

University Line, 2000.
[34] Discovery Studio. Accelrys Inc., San Diego, USA., 2013.
[35] Middha, S.K., Goyal, A.K., Faizan, S.A., Sanghamitra, N., Basistha, B.C., Usha, T. In silico-based combinatorial pharmacophore modelling and docking studies of GSK-3beta and GK inhibitors of Hippophae. Journal of biosciences. 2013, 38, 805-14.
[36] Al-Balas, Q.A., Amawi, H.A., Hassan, M.A., Qandil, A.M., Almaaytah, A.M., Mhaidat, N.M. Virtual lead identification of farnesyltransferase inhibitors based on ligand and structure-based pharmacophore techniques. Pharmaceuticals. 2013, 6, 700-15.
[37] Huang, D., Zhu, X., Tang, C., Mei, Y., Chen, W., Yang, B., et al. 3D QSAR pharmacophore modeling for c-Met kinase inhibitors. Medicinal chemistry. 2012, 8, 1117-25.
[38] Chhabria, M.T., Brahmkshatriya, P.S., Mahajan, B.M., Darji, U.B., Shah, G.B. Discovery of novel acyl coenzyme a: cholesterol acyltransferase inhibitors: pharmacophore-based virtual screening, synthesis and pharmacology. Chemical biology & drug design. 2012, 80, 106-13.
[39] Ali, A., Reddy, G.S., Cao, H., Anjum, S.G., Nalam, M.N., Schiffer, C.A., et al. Discovery of HIV-1 protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel P2 ligands. Journal of medicinal chemistry. 2006, 49, 7342-56.
[40] Ali, A., Reddy, G.S., Nalam, M.N., Anjum, S.G., Cao, H., Schiffer, C.A., et al. Structure-based design, synthesis, and structure-activity relationship studies of HIV-1 protease inhibitors incorporating phenyloxazolidinones. Journal of medicinal chemistry. 2010, 53, 7699-708.
[41] Parai, M.K., Huggins, D.J., Cao, H., Nalam, M.N., Ali, A., Schiffer, C.A., et al. Design, synthesis, and biological and structural evaluations of novel HIV-1 protease inhibitors to combat drug resistance. Journal of medicinal chemistry. 2012, 55, 6328-41.
[42] Li, H., Sutter, J., Hoffman, R. An automated system for generating 3D predictive pharmacophore models. In: Pharmacophore

Perception, Development, and Use in Drug Design, Guner, O.F. Ed., International University Line, La Jolla, CA, 1999, pp. 173-89.
[43] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry. 1983, 4, 187–217.
[44] Momany, F.A., Rone, R. Validation of the general purpose QUANTA ®3.2/CHARMm® force field. Journal of Computational Chemistry. 1992, 13, 888–900.
[45] Smellie, A., Teig, S.L., Towbin, P. Poling: Promoting conformational variation. Journal of Computational Chemistry. 1995, 16, 171-87.
[46] Kristam, R., Gillet, V.J., Lewis, R.A., Thorner, D. Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. Journal of chemical information and modeling. 2005, 45, 461-76.
[47] Sadler, B.R., Cho, S.J., Ishaq, K.S., Chae, K., Korach, K.S. Three-dimensional quantitative structure-activity relationship study of nonsteroidal estrogen receptor ligands using the comparative molecular field analysis/cross-validated r2-guided region selection approach. Journal of medicinal chemistry. 1998, 41, 2261-7.
[48] Li, H., Sutter, J., Hoffman, R. Pharmacophore Perception, Development, and Use in Drug Design. California, International University Line, 2000.
[49] Kubinyi, H., Hamprecht, F.A., Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. Journal of medicinal chemistry. 1998, 41, 2553-64.
[50] Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H. Comparative studies on some metrics for external validation of QSPR models. Journal of chemical information and modeling. 2012, 52, 396-408.
[51] Ojha, P.K., Mitra, I., Das, R.N., Roy, K. Further exploring rm2 metrics for validation of QSPR models. Chemometrics and Intelligent Laboratory Systems. 2011, 107, 194–205.
[52] Golbraikh, A., Tropsha, A. Beware of q2! Journal of molecular graphics & modelling. 2002, 20, 269-76.

ACCEPTED MANUSCRIPT

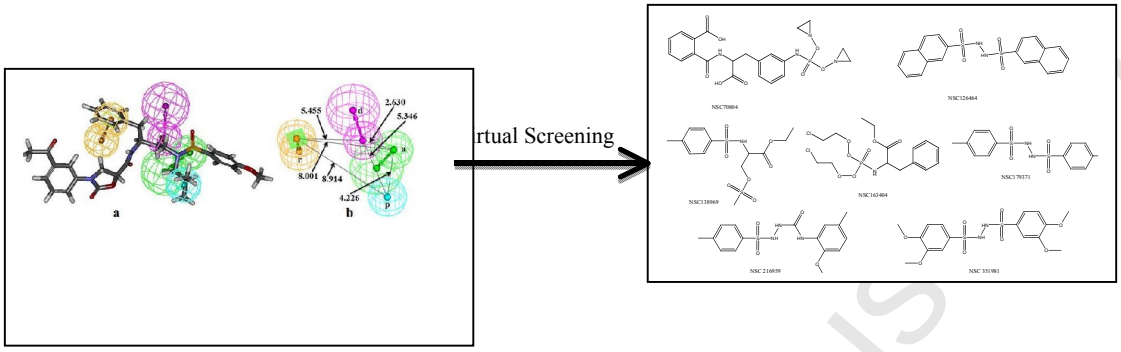
[53] Mitra, I., Saha, A., Roy, K. Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. Journal of molecular modeling. 2010, 16, 1585-96.
[54] Roy, P.P., Paul, S., Mitra, I., Roy, K. On two novel parameters for validation of predictive QSAR models. Molecules. 2009, 14, 1660-701.
[55] Roy, P.P., Roy, K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. QSAR & Combinatorial Science. 2008, 27, 302–13.
[56] Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews. 2001, 46, 3-26.
[57] Debnath, A.K. Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. Journal of medicinal chemistry. 2002, 45, 41-53.
[58] Sakkiah, S., Thangapandian, S., John, S., Kwon, Y.J., Lee, K.W. 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. European journal of medicinal chemistry. 2010, 45, 2132-40.
[59] Taha, M.O., Habash, M., Al-Hadidi, Z., Al-Bakri, A., Younis, K., Sisan, S. Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. Journal of chemical information and modeling. 2011, 51, 647-69.

Accepted Manuscript

ACCEPTED MANUSCRIPT

**Graphical abstract**

The pharmacophore model was developed with a training set of HIV protease inhibitors and further validated using Fischer's randomization, test set prediction and cost function analysis methods. The validated model was used to search the NCI database followed by screening of hit molecules by applying several criteria and finally seven compounds were proposed as potential HIV-protease molecules for the HIV/AIDS therapy.




Accepted Manuscript