

Scientific paper

# Combined Pharmacophore Modeling, 3D-QSAR, Molecular Docking and Molecular Dynamics Study on Indolyl-aryl-sulfone Derivatives as New HIV1 Inhibitors

Mebarka Ouassaf,<sup>1,\*</sup> Faizan Abul Qais,<sup>2</sup> Salah Belaidi,<sup>1,\*\*</sup> Mohamed Bakhouch,<sup>3</sup>  
Ahmed Said Mohamed<sup>4</sup> and Samir Chtita<sup>5</sup>

<sup>1</sup> Group of Computational and Medicinal Chemistry, LMCE Laboratory, University of Biskra, BP 145 Biskra 707000, Algeria

<sup>2</sup> Department of Agricultural Microbiology, Faculty of Agricultural Sciences, Aligarh Muslim University, Aligarh, UP-202002, India

<sup>3</sup> Laboratory of Bioorganic Chemistry, Department of Chemistry, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco

<sup>4</sup> Institut de Recherche Médicinale, Centre d'Etude et de Recherche de Djibouti, Djibouti

<sup>5</sup> Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, B.P 7955, Casablanca, Morocco

\* Corresponding author: \*\*E-mail: s.belaidi@univ-biskra.dz

E-Mail: nouassaf@univ-biskra.dz

Received: 02-22-2022

## Abstract

The present study deals with the *in silico* of 45 indolyl-aryl-sulfones known as anti-HIV1. The data were collected from recent previously reported inhibitors and divided into a sub-set of 33 compounds as the training set and the remaining 12 compounds were kept in the test set. The selected pharmacophore-ADRRR-yielded a statistically significant 3D-QSAR model containing high confidence scores ( $R^2 = 0.930$ ,  $Q^2 = 0.848$ , and  $RMSE = 0.460$ ). The predictive power of the established pharmacophore model was validated with an external test ( $r^2 = 0.848$ ). A systematic virtual screening workflow shows an enrichment factor and has revealed a high predictive power. Then the model was used to screen the filtered PubChem database mapping all chemical features of model pharmacophore. The recognized hits were further assessed by *in silico* ADMET studies. Molecular dynamics also used to explore the stability of obtained complexes. Finally, these selected compounds are probably to become a good lead molecule for the development of effective anti-HIV-1 drugs.

**Keywords:** Indolyl-aryl-sulfone, HIV-1 inhibitor, Pharmacophore, 3D-QSAR, Molecular Docking, Molecular Dynamics.

## 1. Introduction

AIDS is one of the most destructive of human immune system pandemic in the world, caused by human immunodeficiency virus infection (HIV).<sup>1</sup> It continues to be a critical global public health concern, 1.5 million people were newly infected with HIV in 2020, and around 38 million HIV-infected persons are estimated to be dealing with it to date.<sup>2</sup> Unfortunately, there is no effective treatment for HIV infection. Luckily, available antiretroviral drugs are used to control the proliferation of the virus.

Therefore, persons having HIV can lead healthy and productive lives.<sup>3</sup>

Most of the drugs designed and licensed have been classified as Nucleoside Reverse Transcriptase inhibitors (NRTI's), Non-nucleoside Reverse Transcriptase Inhibitors (NNRTI's), Protease Inhibitors (PI's), Fusion Inhibitors, HIV integrase strand transfer inhibitors and Inhibitors-CCR5 co-receptor antagonist.<sup>4</sup> Protease is an important factor for viral maturation within the HIV life cycle.<sup>5,6</sup> The HIV protease is a homodimeric aspartyl pro-

tease and each monomer consists of 99 amino acid residues with a catalytic Asp at position 25. At nine processing sites, the main structural component of HIV-1 is the Gag polyprotein. HIV-1 protease cleaves polyprotein precursors Gag and Gag-Pol encoded by the HIV-1 virus genome to create mature active proteins.<sup>7</sup> Gag-Pol is incorporated into virions via interactions with the Gag precursor Pr55gag. The protease (PR) incorporated into Gag-Pol mediates proteolytic processing of both Pr55gag and Gag-Pol during or shortly after release of viral particles from cells. Since efficient viral incorporation of Gag-Pol depends on interaction with Pr55gag through its N-terminal Gag domain, prevention of premature Gag cleavage may attenuate Gag-Pol packaging deficiencies associated with enhancement of the PR cleavage.<sup>8</sup>

The vital role of HIV protease in viral maturation makes it a popular drug design target; there are 10 FDA-approved HIV protease inhibitors, namely: Saquinavir, Indinavir, Ritonavir, Nelfinavir, Amprenavir, Fosamprenavir, Lopinavir, Atazanavir, Tipranavir, and Darunavir. The FDA-approved HIV protease inhibitors have structural similarities and a similar binding pattern, which might explain some of the protease inhibitor-related adverse effects such as dyslipidaemia, hyperglycaemia, and body-fat distribution. It is possible to optimize the chemical structure of HIV protease inhibitors to avoid side effects.<sup>9,10</sup>

The computer-aided drug design CADD approach has played a crucial role in the search and optimization of potential lead compounds with a substantial benefit in time and expense; it has been used during different phases of drug discovery: target identification, validation, molecular design, and interactions of drug candidates with targets of interest.<sup>11,12</sup>

Pharmacophores are a set of methods related to QSAR: they produce 3-dimensional arrangements of functional group that are required for activity.<sup>13,14</sup>

A well-developed pharmacophore model may be used to design novel and more active molecules, such pharmacophore models are also the starting point for 3D-QSAR analysis, and can allow quantitative predictions. In the very early stages of the drug development process, the use of 3D pharmacophore models will potentially anticipate unwanted side effects and thereby reduce the probability of late failure of drug candidates.<sup>15</sup>

Docking simulations are widely used to screen a library of compounds rapidly and to identify new drug leads employing a simple model. Docking simulations are also useful for lead enhancement using more detailed models to analyze the atomic interactions between inhibitors and target macromolecules.<sup>16</sup>

In order to take a forward step for prediction and guidance of more effective drug, we have utilized state of the art techniques in drug design for the development of a three-dimensional pharmacophore model using a dataset of indolyl-aryl-sulfone derivatives from literature. We have used also a comprehensive approach involving vir-

tual screening-based pharmacophore modeling, molecular docking and Molecular Dynamics (MD) simulations to identify potential HIV1 inhibitors. The studied compounds were consequently analyzed for ADMET properties and were found to be potential drug-like candidates that can effectively bind to the HIV protease enzyme.

Taken together the specifics of the current study could provide important insights needed for the production of next-generation of inhibitors that could theoretically reduce the function of HIV protease.

## 2. Materials and Methods

### 2.1. At a Set Preparation

*In vitro* biological data of a series of 45 indolyl-aryl-sulfones as anti-HIV-1 were collected from literature.<sup>17</sup> the observed anti-HIV-1 activity was represented as EC<sub>50</sub> and converted into logarithmic scale pEC<sub>50</sub> = -logEC<sub>50</sub> (μM).

It is essential to examine the structures of the molecules in the data set before starting molecular modeling. For that, the 3D-structures of the 45 inhibitors were prepared using the builder panel in Maestro 12.0 and were generated for all ligands with LigPrep.<sup>18</sup> Partial atomic charges were ascribed and possible ionization states were generated at a pH equal to 7.0. The OPLS3e force field was used to optimize and to produce low energy conformer of the ligand.<sup>19,20</sup> Energy minimization was performed with OPLS3e force field till root mean square mean deviation (RMSD) of 0.01 Å was attained. The so-prepared ligands were used to generate pharmacophore and to build QSAR model.

### 2.2. Generation of Pharmacophore Model

A pharmacophore describes the arrangement of molecular or functional group's characteristics that a ligand must contain in order to produce a given biological response. Pharmacophore models are developed to identify new compounds that meet the requirements of the pharmacophore, which could have the high probability to be biologically active. Often, such pharmacophore models are the starting point for 3D-QSAR analysis.<sup>21</sup>

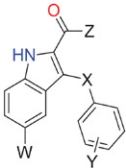
Each compound structure was represented by a set of points in 3D space that correspond to different chemical features, which help the compound to bind with the target receptor non-covalently. The data set was created by assigning pEC<sub>50</sub> > 8.6 as active and pEC<sub>50</sub> < 6.5 as inactive to the threshold. Inactive compounds can be used to screen the hypothesis because they do not provide an explanation for the activity. However, they give signals of the inactive function.

Six pharmacophore features defined the chemical features of the ligands: H-bond Acceptor (A), H-bond Donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P) and Aromatic Ring (R). The consistency of each alignment is calculated by: (1) the

vector score, the average cosine of the angles generated in the aligned structures by the corresponding pairs of vector characteristics (acceptors, donors and aromatic rings); (2) the volume score that is based on the overlap of Van der

Waals non-hydrogen atom models in each pair of structures (3); the site score which is the degree to which site points in the alignment are applicable; (4) The final score, (5) and the function-survival score.

**Table 1.** Various substituents attached to basic structure of indolyl-aryl-sulfones.



| N° | Structural features |                         |                   |          |
|----|---------------------|-------------------------|-------------------|----------|
|    | X                   | Y                       | Z                 | W        |
| 1  | SO <sub>2</sub>     | H                       | NH <sub>2</sub>   | Cl       |
| 2  | S                   | H                       | OE <sub>t</sub>   | H        |
| 3  | S                   | 2-NH <sub>2</sub>       | OE <sub>t</sub>   | Cl       |
| 4  | S                   | 2-NH <sub>2</sub> -5-Cl | OE <sub>t</sub>   | Cl       |
| 5  | SO <sub>2</sub>     | H                       | OE <sub>t</sub>   | H        |
| 6  | SO <sub>2</sub>     | 2-NH <sub>2</sub> -5-Cl | OE <sub>t</sub>   | H        |
| 7  | SO <sub>2</sub>     | 2-NH <sub>2</sub> -5-Cl | OE <sub>t</sub>   | Cl       |
| 8  | S                   | H                       | NH <sub>2</sub>   | H        |
| 9  | S                   | 2-NH <sub>2</sub> -5-Cl | NH <sub>2</sub>   | H        |
| 10 | S                   | H                       | NH <sub>2</sub>   | Cl       |
| 11 | S                   | 2-Me                    | NH <sub>2</sub>   | Cl       |
| 12 | S                   | 4-F                     | NH <sub>2</sub>   | Cl       |
| 13 | S                   | 4-Cl                    | NH <sub>2</sub>   | Cl       |
| 14 | S                   | 4-iPr                   | NH <sub>2</sub>   | Cl       |
| 15 | S                   | 4-tBu                   | NH <sub>2</sub>   | Cl       |
| 16 | S                   | 3,5-Me <sub>2</sub>     | NH <sub>2</sub>   | Cl       |
| 17 | S                   | 2,6-Cl <sub>2</sub>     | NH <sub>2</sub>   | Cl       |
| 18 | S                   | 2-NH <sub>2</sub> -5-Cl | NH <sub>2</sub>   | Cl       |
| 19 | SO <sub>2</sub>     | H                       | NH <sub>2</sub>   | H        |
| 20 | SO <sub>2</sub>     | 2-NH <sub>2</sub> -5-Cl | NH <sub>2</sub>   | H        |
| 21 | SO <sub>2</sub>     | 2-Me                    | NH <sub>2</sub>   | Cl       |
| 22 | SO <sub>2</sub>     | 3-Me                    | NH <sub>2</sub>   | Cl       |
| 23 | SO <sub>2</sub>     | 4-Me                    | NH <sub>2</sub>   | Cl       |
| 24 | SO <sub>2</sub>     | 4-F                     | NH <sub>2</sub>   | Cl       |
| 25 | SO <sub>2</sub>     | 4-Cl                    | NH <sub>2</sub>   | Cl       |
| 26 | SO <sub>2</sub>     | 4-iPr                   | NH <sub>2</sub>   | Cl       |
| 27 | SO <sub>2</sub>     | 4-tBu                   | NH <sub>2</sub>   | Cl       |
| 28 | SO <sub>2</sub>     | 2,4 Me <sub>2</sub>     | NH <sub>2</sub>   | Cl       |
| 29 | SO <sub>2</sub>     | 3,5-Me <sub>2</sub>     | NH <sub>2</sub>   | Cl       |
| 30 | SO <sub>2</sub>     | 2,6-Cl <sub>2</sub>     | NH <sub>2</sub>   | Cl       |
| 31 | SO <sub>2</sub>     | 2-NH <sub>2</sub> -5-cl | NH <sub>2</sub>   | Cl       |
| 32 | SO <sub>2</sub>     | 3,5-Me <sub>2</sub>     | NH <sub>2</sub>   | Br       |
| 33 | SO <sub>2</sub>     | 3,5-Me <sub>2</sub>     | NH <sub>2</sub>   | COMe     |
| 34 | SO <sub>2</sub>     | 3,5-Me <sub>2</sub>     | NH <sub>2</sub>   | CH(OH)Me |
| 35 | S                   | H                       | NHNH <sub>2</sub> | Cl       |
| 36 | S                   | 4-Me                    | NHNH <sub>2</sub> | Cl       |
| 37 | S                   | 4-F                     | NHNH <sub>2</sub> | Cl       |
| 38 | S                   | 4-Cl                    | NHNH <sub>2</sub> | Cl       |
| 39 | SO <sub>2</sub>     | H                       | NHNH <sub>2</sub> | H        |
| 40 | SO <sub>2</sub>     | H                       | NHNH <sub>2</sub> | Cl       |
| 41 | SO <sub>2</sub>     | 4-Me                    | NHNH <sub>2</sub> | Cl       |
| 42 | SO <sub>2</sub>     | 4-F                     | NHNH <sub>2</sub> | Cl       |
| 43 | SO <sub>2</sub>     | 4-Cl                    | NHNH <sub>2</sub> | Cl       |
| 44 | SO <sub>2</sub>     | 3,5-Me <sub>2</sub>     | NHNH <sub>2</sub> | Cl       |
| 45 | SO <sub>2</sub>     | 2-NH <sub>2</sub> -5-Cl | NHNH <sub>2</sub> | Cl       |

### 2. 3. Building 3D-QSAR Model

QSAR modeling was performed using the selected hypothesis by dividing randomly the data set into training set (60%) and test set (40%). This phase presents two options for the alignment of the 3D-structure of molecules: pharmacophore-based alignment and atom-based alignment.

In this study, the selected 45 compounds from the chemical dataset were used to develop an atom-based 3D-QSAR model based on previously developed pharmacophoric maps as a backbone with a default grid space of 1 Å via partial least-square (PLS) regression.<sup>22</sup>

### 2. 4. Model Validation

Validation is a critical aspect of pharmacophore design, particularly when the model is constructed for predicting molecular activity in external test series.<sup>23</sup> the intensity of the defined pharmacophore hypotheses was internally validated by statistical parameters, squared coefficient of correlation ( $R^2$ ) and the ratio of variance (F). Validation on chemicals was not used in the model development, the so-called external validation, is particularly important in the context of using QSAR models for the prediction of new data in virtual screening.<sup>24</sup> The approach demonstrated by Golbraikh and Tropsha, in 2000<sup>25</sup> and Roy and al., 2008<sup>26</sup> was used to evaluate the predictive potential of the current QSAR model. Further, the best hypothesis selected was validated by enrichment

studies using the decoy test. More than 1000 decoy test set compounds retrieved from the PubChem database, were taken to evaluate the predictive power of the built model<sup>27</sup> and were taken to evaluate some parameters, such as: Enrichment Factor (EF), Robust Initial Enhancement (RIE), Receiver Operating Characteristic (ROC) and Boltzmann-Enhanced Discrimination of ROC (BEDROC). These parameters were used to benchmark the reliability of the model and for the accurate ranking of compounds.<sup>28</sup>

### 2. 5. Virtual Screening of PubChem Database

In pharmaceutical research, computational screening of databases has become incredibly popular. Based on biological structures, virtual screening uses computer-based methods to discover new ligands.<sup>29,30</sup> the aim of virtual screening, in this work, is to detect potential leads to anti-HIV with various scaffolds and high inhibitory activity. To identify inhibitors of PR HIV, we have screened the PubChem database<sup>31</sup> by searching compounds having more than 80% similarity instead of compound that have the most fitness score (compound 40, Table 1). All PubChem drug-like compounds (459926) were filtered by Canvas's property filter utility to pick compounds with low-dimensional properties similar to the anti-HIV compounds. We used the following property filters: AlogP  $\geq 1$ , AlogP  $\leq 5$ , HBA  $\geq 2$ , HBA  $\leq 3$ , HBD  $\geq 1$ , HBD  $\leq 3$ , MW  $\geq 250$ , MW  $\leq 500$ , Num rings  $\geq 3$ , Num rings  $\leq 5$ , Polar  $\geq 45$ , Polar  $\leq 60$ , RB  $\geq 3$  and RB  $\leq 5$ . As a result,

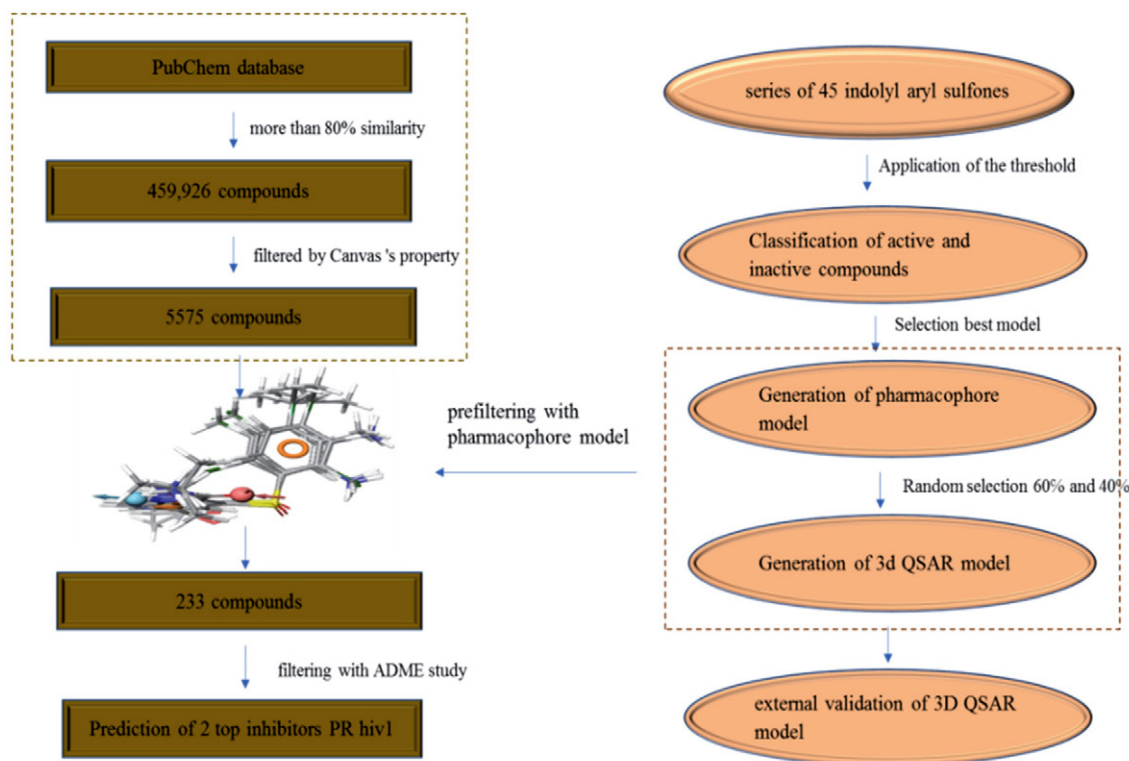


Fig. 1. Schematic representation of the methods followed in the current study.

5575 compounds were obtained and mounted as a series of decoys. Each compound must fit a minimum of four sites and a distance of 2.0 Å matching between sites. After the pharmacophore-based screening according to the fitness score, the top 230 hits were selected for molecular docking analysis. Figure 1 represents the schema of the methods followed in the current study.

## 2. 6. Molecular Docking Analysis

The top 230 compounds were selected for molecular docking analysis based on the pharmacophore model following virtual screening. X-ray crystal structure of wild type HIV-1 protease in complex with GRL-09510 (PDB ID: 5v4y) were obtained from the Protein Data Bank PDB.<sup>32</sup> The protein structure was preprocessed using the Protein Preparation Wizard,<sup>33</sup> available in Schrödinger suite 2021-2, by eliminating crystallographic water molecules (water molecules without H-bonds), inserting all missing side chain atoms, and pH 7.0 corresponding hydrogen bonds, taking into account the necessary ionization states for the residues of both acid and basic forms of the amino acid.

Finally, energy minimization (up to 3.0 Å RMSD value) was performed using OPLS-2005 force field after the assignment of charge and protonation state. In fact, it was minimized to alleviate the steric clashes between the residues due to the addition of hydrogen atoms. The active site was defined with a radius of 10 Å around the ligand present in the crystal structure and a grid was generated at the center of gravity of the active site for docking. All studied compounds were docked into the catalytic pocket of the Protease protein (PDB-ID: 5v4y) using Grid-Based Ligand Docking with Energetics<sup>34</sup> with default parameters. Finally, the docking results were analyzed using Biovia Discovery Studio 4.5.12 (Dassault Systèmes 2018).<sup>35</sup>

## 2. 7. Analyzing ADMET

ADMET (Absorption, Distribution, Metabolism, Elimination, Toxicity) analysis is important in drug design. These properties were calculated using the QikProp module<sup>36</sup> of Schrödinger suite for assessing the drug ability and to filter the ligand molecules at an early stage of identifying the new inhibitors.

Toxicity is the degree to which a substance can damage an organism or substructure of the organism. The predictions of toxicity of the compounds are essential to re-

duce the cost and labor of a drug's preclinical and clinical trials. The toxicity evaluation was performed also using the ProTox platform.<sup>37</sup> It gives predicted toxicity values, cytotoxicity, mutagenicity, carcinogenicity, immunotoxicity and LD50 values of selected compounds.

## 2. 8. Molecular Dynamics Simulations

Two compounds showing highest binding affinity towards HIV-1 protease were selected for Molecular Dynamics (MD) simulation studies. The molecular docked complexes with lowest binding energy were used as initial point for the MD simulations performed using Gromacs-2018.1 packages with amber99sb-ILDN force field.<sup>38,39</sup> The protein alone and their complexes with ligands (11630770 and 55868948) were solvated in triclinic-boxed using TIP3P water model. Each structure was neutralized using counter chlorine ions. The topology of both ligands was prepared using antechamber packages in Amber Tools 19.<sup>40</sup> For the removal of weak Van der Waals contacts; each system was minimized using the steepest descent minimization. The systems were then equilibrated for NVT using V-rescale thermostat for 1 ns at 300 K temperature followed by NPT equilibration using Parrinello-Rahman barostat at 1.0 bar for 1 ns.<sup>41,42</sup> MD simulation of each system was carried out for 100 ns and the trajectories were recorded at 10 ps intervals. Each trajectory was subjected to PBC (periodic boundary conditions) corrections before analysis. All calculations except MM-PBSA was done using Gromacs utilities. MM-PBSA calculation was performed for the calculation of various binding energies the ligands with HIV-1 protease.<sup>43</sup>

## 3. Results and Discussions

### 3. 1. Pharmacophore Modeling

Our work is focused on the identification of new compounds with potential antiviral activity anti-HIV-1. To fulfill the objective, a ligand-based pharmacophore model was built using previously reported inhibitors, with a different combination of pharmacophoric features, 920 pharmacophore hypotheses have been produced.

Its vector, volume, sites survival score, and the number of matches measured the quality of each hypothesis. Table 2 represents the different scoring parameters for best hypothesis. The best fitted Model ADRRR1 with the

**Table 2.** Different parameter scores of the generated hypothesis ADRRR1.

|        | Survival score | Site  | Vector | Volume | Bedroc | Matches |
|--------|----------------|-------|--------|--------|--------|---------|
| ADRRR1 | 6.067          | 0.944 | 0.977  | 0.917  | 0.843  | 5       |
| ADRRR2 | 6.063          | 0.695 | 0.922  | 0.613  | 0.843  | 5       |
| ADRRR3 | 6.062          | 0.705 | 0.919  | 0.620  | 0.844  | 5       |
| ADRRR4 | 6.061          | 0.711 | 0.922  | 0.619  | 0.843  | 5       |
| ADRRR5 | 6.061          | 0.681 | 0.932  | 0.622  | 0.844  | 5       |

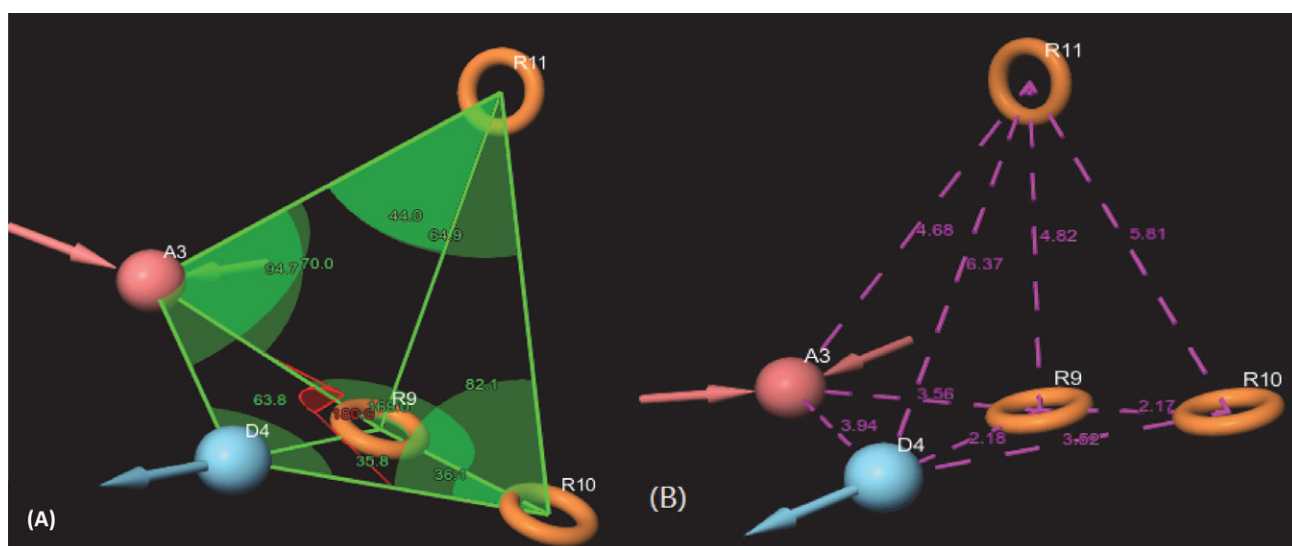
highest survival score (6.067) and site score (0.944) consist five-point hypothesis one hydrogen acceptors, one hydrogen donor, and three ring group. The spatial arrangement of the best pharmacophore hypothesis, ADRRR1 with their distance between the five-pharmacophore features is shown in figure 2 and tables S1-S2.

### 3. 2. 3D-QSAR Model

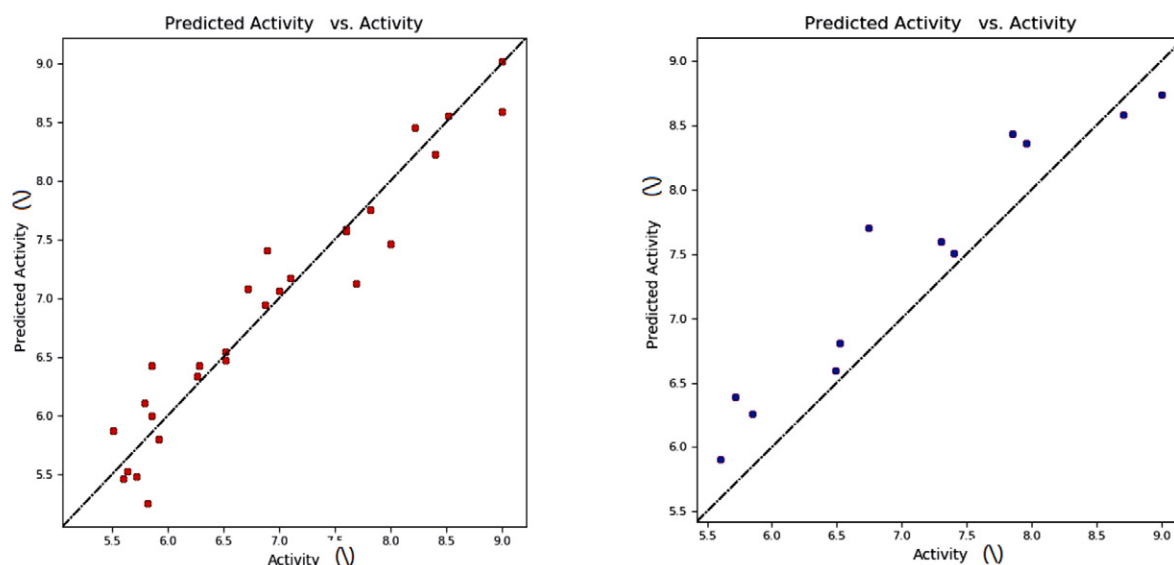
The previously developed pharmacophore hypothesis ADRRR1 was used to build an atom-based 3D-QSAR with the phase program.<sup>44</sup> Based on the training set molecules for the chosen hypothesis, the pharmacophore model, that is statistically significant, was created through partial least-square (PLS) regression. The partial least-squares

factor has been raised to five, as there is a gradual improvement in the model's predictive power and statistical significance until the fifth factor.

A statistically significant 3D-QSAR model was obtained using this pharmacophore hypothesis with a strong correlation coefficient ( $R^2 = 0.929$ ) and a high Fisher ratio ( $F = 57$ ) for the training set. The predictive power of the developed model was also found to be important, verified by the high value of the coefficient of cross-validated correlation ( $Q^2 = 0.848$ ) and Pearson's  $R$  (0.926) for the test set. The plots between the observed and the predicted activities were made for both the training and test sets (Figure 3). The higher values of  $R^2$  and  $Q^2$  in the training and test sets, respectively, are clearly indicated by the points lying extremely near to the best-fit line.



**Fig. 2.** (A) Pharmacophore model ADRRR1 interstice angles in (°) unit between the pharmacophoric points and (B) Pharmacophore model ADRRR1 interstice distances in Å unit.



**Fig. 3.** The plot of the correlation between the experimental and predicted activity of based indolyl aryl sulfones inhibitors using pharmacophore-based QSAR model of training and test set.



### 3. 3. Model Validation

All the external validation results were above the threshold values for the various parameters presented in table S3. The squared correlation coefficient values between the observed and predicted values of the test set compounds ( $r^2$ ) and ( $r^2_0$ ), respectively, were observed and the model had satisfied the requirement of the term  $(r^2 - r^2_0)/r^2$ . This was in agreement with a previous study reported by Golbraikh and co-workers, which states that the value  $(r^2 - r^2_0)/r^2$  exhibits less than 0.1. In case of good external prediction, predicted values will be very close to observed activity values. Therefore,  $r^2$  value will be very near to  $r^2_0$  value. In the best case  $r^2_m$  will be equal to  $r^2$ , whereas in the worst-case  $r^2_m$  value will be zero, including values of  $r^2_m < 0.6$  indicate these models are useless for external predictivity. In the present study  $r^2_m$  value of the model is acceptable (Table S3). This developed model passed all the Golbraikh and Tropsha criteria for the acceptability of the model. The screening results were evaluated by an enrichment factor at the top 1% of the ranked database (EF1) and are summarized in table S4.

reported by Golbraikh and co-workers, which states that the value  $(r^2 - r^2_0)/r^2$  exhibits less than 0.1.

In case of good external prediction, predicted values will be very close to observed activity values. Therefore,  $R^2$  value will be very near to  $R^2_0$  value. In the best case  $r^2_m$  will be equal to  $r^2$ , whereas in the worst-case  $r^2_m$  value will be zero, including values of  $r^2_m < 0.6$  indicate these models are useless for external predictivity. In the present study  $r^2_m$  value of the model are acceptable (Table S3).

This developed model passed all the Golbraikh and Tropsha criteria for the acceptability of the model. The screening results were evaluated by an enrichment factor at the top 1% of the ranked database (EF1) and are summarized in table S4.

The enrichment factor (EF) of this screening protocol was calculated to be 13.012, which indicated that selected model has 13 times more stability to identify active molecules than inactive.

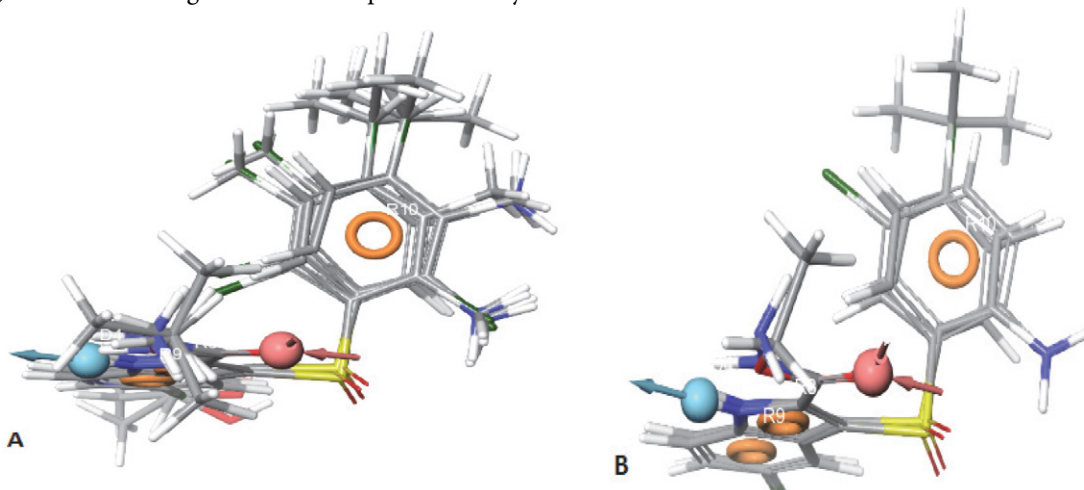
Visualization of the validation was presented in ROC analysis to show how effectively the pharmacophore mod-

**Table 3.** 3D-QSAR PLS statistical results of the selected Pharmacophore model ADRRR1.

| ID     | PLS factors | SD    | R <sup>2</sup> | F  | P                      | RMSE  | Q <sup>2</sup> | Pearson-R |
|--------|-------------|-------|----------------|----|------------------------|-------|----------------|-----------|
| ADRRR1 | 1           | 0.650 | 0.722          | 39 | 4.52 10 <sup>-9</sup>  | 0.600 | 0.742          | 0.869     |
|        | 2           | 0.509 | 0.835          | 49 | 1.78 10 <sup>-11</sup> | 0.620 | 0.724          | 0.865     |
|        | 3           | 0.425 | 0.889          | 56 | 5.83 10 <sup>-13</sup> | 0.490 | 0.826          | 0.914     |
|        | 4           | 0.404 | 0.903          | 50 | 7.10 10 <sup>-13</sup> | 0.530 | 0.799          | 0.904     |
|        | 5           | 0.351 | 0.930          | 57 | 4.30 10 <sup>-14</sup> | 0.460 | 0.848          | 0.925     |

All the external validation results were above the threshold values for the various parameters presented in table S3. The squared correlation coefficient values between the observed and predicted values of the test set compounds ( $r^2$ ) and ( $r^2_0$ ), respectively, were observed and the model had satisfied the requirement of the term  $(r^2 - r^2_0)/r^2$ . This was in agreement with a previous study

els distinguished between active and inactive compounds (Figure 4). Sensitivity (in other words, true positive rate, recall, hit rate) and specificity (in other words, true negative rate) are general indices to show the predictive power of a validated model and is indicated by the area under the curve. The area under the curve (AUC) was calculated as 0.8157.



**Fig. 4.** (A) Mapping of the active compounds onto the pharmacophore. (B) Mapping of inactive compounds on to the pharmacophore.

Thus, we can conclude that our model is not randomly classified. Considering area under the ROC curve, it is statistically significant from those obtained by random classifier (Area = 0.5).

The result revealed that in 1% of the total database the generated model screened 28 decoys and five active compounds overall while in the top 2% it was able to get 21 decoys and 8 active compounds hits, with an enrichment factor of 19. The detailed results are given in table S5.

### 3. 4. Contour Plot Analysis

Contour plot analysis was conducted at spatial locations of the system to interpret and understand the distinct vital pharmacophoric criteria. Positive and negative activity coefficients of different properties are described in the map, including (a) donor hydrogen bond, (b) hydrophobic/non-polar and (c) ionizable negative properties. The blue cubes show their individual positive contribution, and the red cubes reflect the negative contribution (Figure 5).

#### 3. 4. 1. H-Bond

Red region near and around position X and W indicates that the substitutions at these positions by groups having more hydrogen bond donor property is unfavorable to anti-HIV activity for example in compounds 3 and 5.

The blue cubes around the position Z suggests that the presence of a donor substitution (e.g N, O, P, or S) at this position may favor the formation of H bond interaction. Almost all the compounds containing Sulfur Dioxide (e, g Compound 1) were found to have better activity

profile in comparison with the ones with Sulfur atom (e, g Compound 8).

#### 3. 4. 2. Hydrophobic

Another significant component that affects the anti-HIV activity is the hydrophobic character. In figure 5B, the contour map for hydrophobic characteristics displays blue cubes highly distributed proximal to the R9, R10 and R8 regions of indolyl-aryl-sulfones. This result reveals that the multiple rings R8, R9, and R10 of the indolyl-aryl-sulfones may enhance the hydrophobicity, and might play a major role in its higher activity. The presence of red cubes at W position of phenyl ring directly attached to the Cl group indicates that hydrophobic groups are unfavorable at this position. This assumption is supported by the low activity of Cl substituted compounds when compared to their unsubstituted derivatives. This is evident while comparing the compounds 6 with 7 and 32.

#### 3. 4. 3. Negative

In contour plot of compound 25 (Figure 5C), the presence of red cubes at *para* position X indicates that the presence of electron withdrawing groups is undesirable at this position. This is evident while comparing the compounds 2 (X: S) with 5 (X: SO<sub>2</sub>). On the contrary, the presence of blue cubes at *para* position of W indicated the preference of electron withdrawing groups at this position. This is explained by the significant anti-HIV activity of compounds with para halogen substitution (30, 35 and 40) (pEC<sub>50</sub> = 8.70, 7.82 and 7.60) in the order of Br > COMe > CHOHMe.

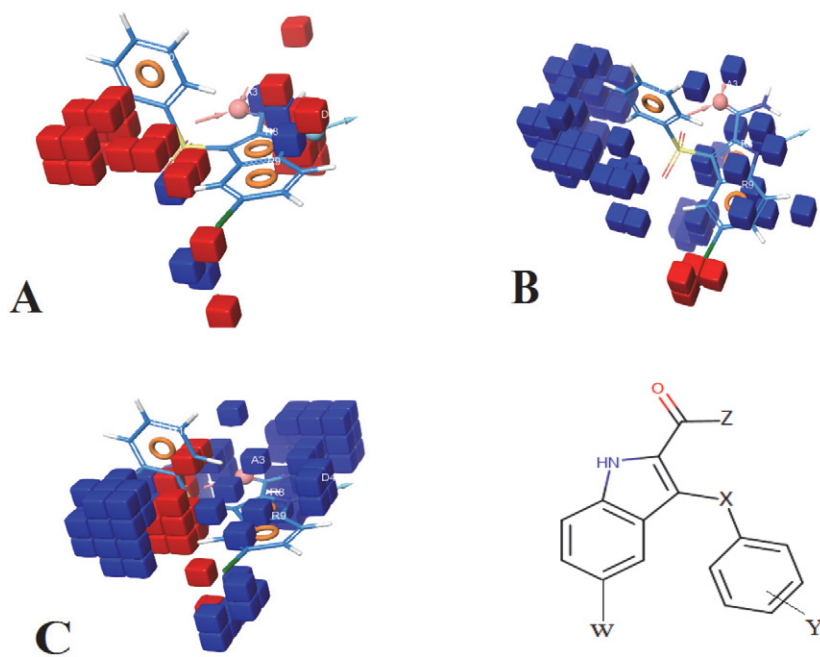


Fig. 5. QSAR model visualized in the context of favorable and unfavorable effects in compound: (A) hydrogen bond donor, (B) hydrophobic/non-polar and (C) negative ionizable properties.



### 3. 5. Identifying Novel Inhibitors

The created five-site pharmacophore hypothesis, ADRRR-1, was used to identify new inhibitors with a new scaffold that corresponds to the 3D-QSAR model's predicted molecular properties. Lead-like compounds from PubChem have been used to obtain new inhibitors, which could potentially target the HIV protease receptor. There was a total of 5575 hits compounds found from the PubChem datasets was similar 80% for most active compound of indolyl-aryl-sulfones composes. Pharmacophore model was used to identify the molecules that satisfy the hypothesis. Pharmacophore pre-filtering with ADRRR-1 hypothesis reduced initial 233 hits.

### 3. 6. Virtual Docking Screening

Molecular docking simulations for the selected set of hits were performed using the Glide (Grid-Based Ligand Docking with Energetics) program available in Schrödinger 2021. The 3D structures were prepared using the Maestro LigPrep module. This module generates possible 3D conformations for each ligand with various ionization states at pH 7.0 ± 2.0. The docking screening process was conducted in two steps: (i) Glide/SP is performed for 230 ligands accelerated docking simulations; (ii) Glide/XP docking has chosen the top-docking ligands for more detailed analyses. The scores of docking studies Glide/SP are shown in the (Table S5).

The key residues involved in substrate binding, including Asp25 Gly27 Ala28 Asp29 Asp30 Thr31 Val32 Ile47 Gly48 Gly49 Ile50 Gly51 Gly52 Phe53 Ile54 Leu76 Thr80 Pro81 Val82 Ile84, the structural analysis using the X-ray crystallographic data of PR complexed with GRL-09510 (8FM) showed that the P2-Crwn-THF of GRL-09510 forms a strong hydrogen-bond network with the backbone atoms of Asp 29 and Asp 30.

Validation of the docking process was done by docking of the compounds Nelfinavir and co-crystallized ligand 8FM ((3S,3aR,5R,7aS,8S)-hexahydro-4H-3,5-methanofuro[2,3-b] pyran-8-yl [(2S,3R)-3-hydroxy-4-[(4-methoxyphenyl) sulfonyl] (2-methylpropyl) amino]-1-phenylbutan-2-yl) carbamate) at the active site of the target. Whereas, Nelfinavir is one of many protease inhibitors currently available, used to limit viral replication and improve immune function in people with HIV infection.<sup>45</sup> The docking complex and binding interactions of Nelfinavir with HIV-1 protease are given in (Figure S1) with binding affinity –5.88

kJ/mol. This interaction is favored by the formation of the H-bond and non-hydrophobic interactions. The H-bonds are supported by the amino acids, Asp29 and Asp25 with the active site of protease and salt bridge interaction with Asp25, whereas pi-alkyl stacking with Val82 and Ile54. By comparing the docked energy of all the molecules studied, it is noteworthy that six bonds have better energy scores than Nelfinavir, knowing that the energy value for the reaction of the indicated molecule is –5.88 kcal/mol (Table 4).

The top-scored hit molecule identified is PubChem 11560933 with a binding energy of –7.55 kcal/mol. The second top-scored hit molecule is PubChem 11654778 with a binding energy of –7.227 kcal/mol and the third top-scored hit molecule is PubChem 11710411 with an average binding energy of –6.655 kcal/mol. The molecular structure of top-scored compounds can be seen in table 5.

Visualization of the docking results revealed that all the ligands adopted a very similar orientation in the active site. The nitrogen atoms with amide groups are oriented to the two aspartic acids as shown in figure 6. All of them form H-bond, while their large hydrophobic groups were often orientated to the main hydrophobic site, which distinguishes the PR active site. Moreover, the literature was mentioned that the catalytic triad Asp-Thr-Gly that is where the ligand binds determines the active site of the enzyme.<sup>46,47</sup>

Simplifying the docking results for the compounds, we have taken the 2D representative ligands; the binding mode of the most active compound CID 11560933 is shown in table S13. In its binding mode, the Diamino Hexanoyl amino fragment is observed to be inserted deeply in the cavity, interacting with Gly27 Asp29 Ala28 and Val32 through H-bond and pi-H contacts, respectively.

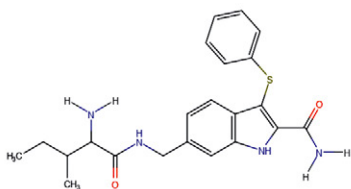
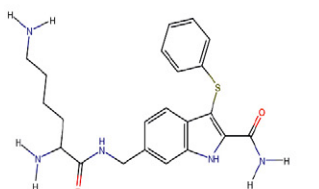
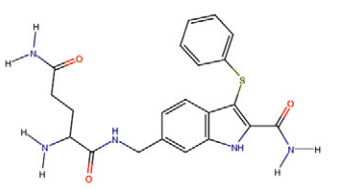
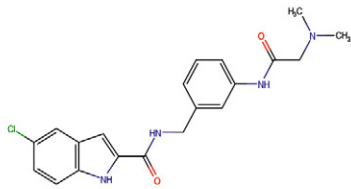
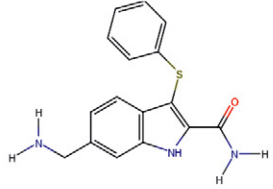
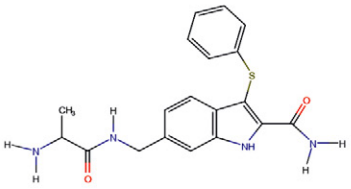
In ligand 11654778, the two oxygen of the pentane diamide have shown strong hydrogen bonding acceptor interaction with aspartate (residue number 29 and 30) was also mapped on HBA features on ADRRR1 pharmacophore model.

Whereas the nitrogen atom of the carboxamide group in all the compounds shows H-bonds with Gly27 are observed explaining why the H-donor is beneficial for activity in the pharmacophore 3D-QSAR model. It should be noted that favorable interactions of hydrophobic type are observed between the indole rings and the residues Ala28, Ile54 and Val32; we can say that the aromatic ring was important for binding and stability of ligand with the HIV-1 PR active site complex.

**Table 4.** Docking scores of the selected hits using GLIDE module.

| Compound Names CID | Xp score | Glid E model | Compound Names | Xp score | Glid E model |
|--------------------|----------|--------------|----------------|----------|--------------|
| CID 11560933       | –7.554   | –56.274      | CID 55868948   | –6.615   | –45.144      |
| CID 11654778       | –7.227   | –53.648      | CID 11567743   | –6.454   | –45.532      |
| CID 11710411       | –6.655   | –48.052      | CID 11630770   | –6.449   | –47.750      |
| Nelfinavir.        | –5.882   | –51.263      | MF8            | –4.851   | –48.632      |

Table 5. Details of virtual screening compounds.

| Structure                                                                           | CID      | Molecular formula                                               | PubChem IUPAC name                                                                     | Molecular mass (g/mol) |
|-------------------------------------------------------------------------------------|----------|-----------------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------|
|    | 11567743 | C <sub>22</sub> H <sub>26</sub> N <sub>4</sub> O <sub>2</sub> S | 6-[[[(2-amino-3-methylpentanoyl)amino]methyl]-3-phenylsulfanyl-1H-indole-2-carboxamide | 410.5                  |
|    | 11560933 | C <sub>22</sub> H <sub>27</sub> N <sub>5</sub> O <sub>2</sub> S | 6-[(2,6-diaminohexanoylamino)methyl]-3-phenylsulfanyl-1H-indole-2-carboxamide          | 425.5                  |
|    | 11654778 | C <sub>21</sub> H <sub>23</sub> N <sub>5</sub> O <sub>3</sub> S | 2-amino-N-[(3-[(2-carbamoyl-3-phenylsulfanyl-1H-indol-6-yl)methyl]pentanediamide       | 425.5                  |
|   | 55868948 | C <sub>20</sub> H <sub>21</sub> ClN <sub>4</sub> O <sub>2</sub> | 5-chloro-N-[[3-[[2-(dimethylamino)acetyl]amino]phenyl]methyl]-1H-indole-2-carboxamide  | 384.9                  |
|  | 11630770 | C <sub>16</sub> H <sub>15</sub> N <sub>3</sub> OS               | 6-(aminomethyl)-3-phenylsulfanyl-1H-indole-2-carboxamide                               | 297.4                  |
|  | 11710411 | C <sub>19</sub> H <sub>20</sub> N <sub>4</sub> O <sub>2</sub> S | 6-[(2-aminopropanoylamino)methyl]-3-phenylsulfanyl-1H-indole-2-carboxamide             | 368.5                  |

One of the most important characteristics of the HIV-1 protease is that all amino acids of the active site are hydrophobic except for hydrophilic aspartic acids (Asp25).<sup>48</sup> Due to these HIV-1 protease active site characters, the hits are considered a good inhibitor of the activity of HIV-1 protease because its hydrophobic surface provides strong Van der Waals interaction between hits and HIV-1 protease active site, which are beneficial to activity. This supports the proposed model pharmacophore ADRRR1 as it consists of three rings.

The compounds CID 11560933, CID 11654778,

CID 11710411, CID 55868948, CID 11567746, and CID 11630770 make good interaction with HIV-1 protease by forming hydrogen bonds, hydrophobic interactions and non-bonding interaction with catalytic residues such as Asp30, Thr80, Gly27, Asp29, Ile54 and Ile84 at the active site cavity of HIV protease. These interactions systems have the lowest total energies. According to this study, the type and spatial location of the hit compounds agree perfectly with the pattern of enzyme inhibitor interactions identified from Nelfinavir. In future, optimality of the compounds should be confirmed experimentally and

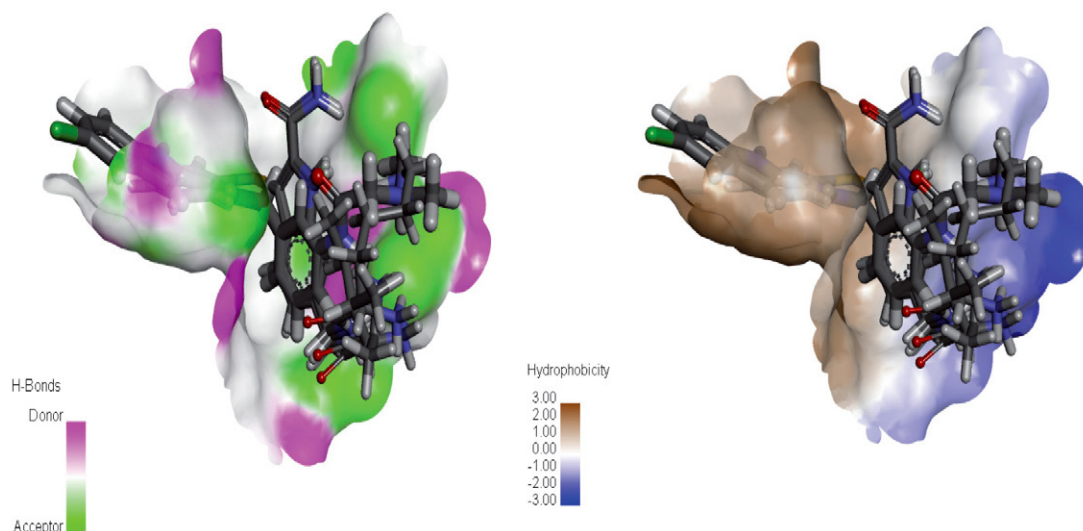


Fig. 6. The orientation of HIV 1 PR inhibitors in the active site and hydrogen and hydrophobic surface

compared its binding modes with a number of HIV-1 protease.

### 3. 7. Analyzing Absorption, Distribution, Metabolism and Excretion (ADME)

For additional validation purpose, The QikProp module of Schrodinger is a quick, accurate, and easy-to-use to describe absorption, distribution, metabolism and elimination results listed in table 6.

The aqueous solubility plays an essential role in the bioavailability of the candidate. The aqueous solubility parameter (QPlog S) of the test entities was assessed and the compounds were also found to be in the permissible range ( $< 0.5$ ) (Table 6).

One of the essential factors to be studied concerning the absorption of the drug molecule is also intestinal absorption or permeation, which was further supported by the expected permeability of Caco-2 cells (QPPCaco). The estimation of the test compounds for Caco-2 cell permeability the compounds CID 55868948 shows excellent

results, predicting strong intestinal absorption. The other compounds have poor permeability across the gut–blood barrier. The parameters of the brain/blood partition coefficient (QPlogBB) define the drug's ability to pass through the blood-brain barrier, which is important for ADME to investigate drug performance. The QikProp descriptor for blood/brain partition coefficient QPlogBB has shown reliable prediction for all the test compounds and reference drugs. Predictions related to skin permeability ( $K_p$ ), showed that these parameters for the active analogs fall within the standard ranges normally observed for drugs (Table 6). In addition, four out of the six compounds, shown to be more than 25% human oral absorption. Human Ether-a-go-go Related Gene (hERG) parameter is used to determine the potential cardiac toxicity of the compounds. The hERG encodes a potassium ion ( $K^+$ ) channel that plays a role during systolic and diastolic activities of the heart. The blockage of hERG  $K^+$  channels can lead to cardiac arrhythmia. All the six compounds have  $\log IC_{50}$  (hERG) values less than the acceptable range for the blockage of hERG  $K^+$  channels ( $\log IC_{50}$  (hERG)  $< -5$ ); however,

Table 6. Estimated physicochemical and pharmacokinetic parameters by QikProp.

| Compound Names CID | QPlogs | QPPCaco | QplogHEGd | Qlog BB | Percent human oral absorption | Metabolism | Logkp  |
|--------------------|--------|---------|-----------|---------|-------------------------------|------------|--------|
| 11560933           | −2.474 | 0.822   | −4.316    | −3.024  | 7.540                         | 7          | −7.518 |
| 11654778           | −0.446 | 1.018   | −3.709    | −2.655  | 8.635                         | 7          | −7.374 |
| 11710411           | −3.067 | 15.627  | −5.325    | −1.665  | 40.552                        | 5          | −5.837 |
| 55868948           | −4.648 | 211.95  | −7.225    | −0.459  | 86.524                        | 4          | −4.064 |
| 11567743           | −3.223 | 25.586  | −5.525    | −1.540  | 50.499                        | 5          | −5.302 |
| 11630770           | −3.231 | 64.203  | −5.923    | −0.929  | 65.168                        | 4          | −5.479 |

**QPlogS** is the predicted aqueous solubility,  $\log S$ : S in moles/l is the concentration of the solute in a saturated solution that is inequilibrium with the crystalline solid; **QPPCaco** is the predicted apparent Caco-2 cell permeability in nm/s; Caco-2 cells are a model for the gut–blood barrier Recommended values **QPPCaco** =  $<25$  POOR,  $>500$  great;; **QplogHEGd**: Predicted  $IC_{50}$  value for blockage of hERG  $K^+$  channels QplogHEGd concern below  $-5$ . **QPlog BB** is the predicted brain–blood partition coefficient; **Percent Human-OralAbsorption** =  $<20$  POOR,  $>80$  great; **QPlogBB** =  $-3.0$  to  $1.2$ ; \*Recommended values – As per the guidelines given in Schrodinger's Maestro software suite manual

the values are near the borderline. All the six compounds were predicted to possess non-blocker to hERG channel (Table 6). An estimated number of possible metabolic reactions has also been predicted by QikProp and used to determine whether the molecules can easily gain access to the target site after entering the blood stream. The compounds have the most elevated QPlogP values. A number of likely metabolic responses of the compounds are in the range of 4–7. The *In silico* ADMET results revealed that the top six of Mpro inhibitors are virtually safe and active.

### 3. 8. Prediction of Toxicity

The computational prediction of toxicities was based on 5 different targets linked to adverse drug-reactions. The hepatotoxicity, carcinogenicity, mutagenicity and cytotoxicity of the compounds were predicted. It was found that 6 compounds have shown no toxicity.

The LD50 has been also predicted, the obtained results have shown that compounds 11630770 and 5586948 present a LD50 of 650 and 1000 mg/kg, respectively, as well as class four of toxicity. The other compounds show moderate toxicity with a LD50 value of 200 mg/kg and class 3 of toxicity.

Table 7. Toxicity prediction of the selected compounds

| Compound Names CID | Hepatotoxicity | Carcinogenicity | Mutagenicity | Cytotoxicity | LD50 mg/Kg |
|--------------------|----------------|-----------------|--------------|--------------|------------|
| 11567743           | Inactive       | Inactive        | Inactive     | Inactive     | 200        |
| 11560933           | Inactive       | Inactive        | Inactive     | Inactive     | 200        |
| 11654778           | Inactive       | Inactive        | Inactive     | Inactive     | 200        |
| 55868948           | Inactive       | Inactive        | Inactive     | Inactive     | 625        |
| 11630770           | Inactive       | Inactive        | Inactive     | Inactive     | 1000       |
| 11710411           | Inactive       | Inactive        | Inactive     | Inactive     | 200        |

### 3. 9. Molecular Dynamics Simulations

To further obtain the insights regarding the interaction of the ligands (CID 11630770 and CID 55868948) with 5v4y HIV-1 protease, Molecular Dynamics Simulation was performed. The docked complex was used as initial conformations for MD simulations.

#### 3. 9. 1. Analysis of RMSD and RMSF

The initial analysis of MD simulation was performed by calculating the root-mean square deviations (RMSD) with respect to their respective backbone of initial structure to assess the stability of the 5v4y HIV-1 protease and complexes under physiological conditions. The RMSD of 5v4y HIV-1 protease and complexes is shown in figure 7 and listed in table 8. The RMSD of 5v4y HIV-1 protease alone showed some variations initial time till 20 ns then it became stable for entire simulation period. The average RMSD of was found to be  $0.290 \pm 0.047$  nm. The RMSD of 5v4y HIV-1 protease CID 11630770 complex was found

to be more stable during simulation competed to 5v4y HIV-1 protease alone. However, 5v4y HIV-1 protease 55868948 complex showed more variations with some spikes in RMSD. The RMSD average of both complexes was statistically insignificant ( $p$ -value  $< 0.05$ ) with respect to 5v4y HIV-1 protease alone. The RMSD analysis of CID 55868948 alone revealed that these variations were due to the fluctuations in RMSD of the ligand (CID 55868948). The fluctuation in structures was assessed by calculating the root mean square fluctuation (RMSF) of  $C_{\alpha}$  of 5v4y HIV-1 protease in the absence and the presence of ligands (Figure 8A). As evident from data, the RMSF of most of the residues of 5v4y HIV-1 protease alone was found to be less than 0.2 nm. A similar fluctuation was recorded for 5v4y HIV-1 protease CID 55868948 complex where most of the fluctuation in most of residues was below 0.2 nm. However, there was comparatively more fluctuations in  $C_{\alpha}$  atoms of 5v4y HIV-1 protease CID 55868948 complex. The RMSF of each atom of both ligands (CID 11630770 and CID 55868948) was also calculated (Figure 8B). The RMSF value of the atoms of both ligands varied from their respective initial values indicating that the ligands exhibited dynamical shift from its initial position in the binding region.

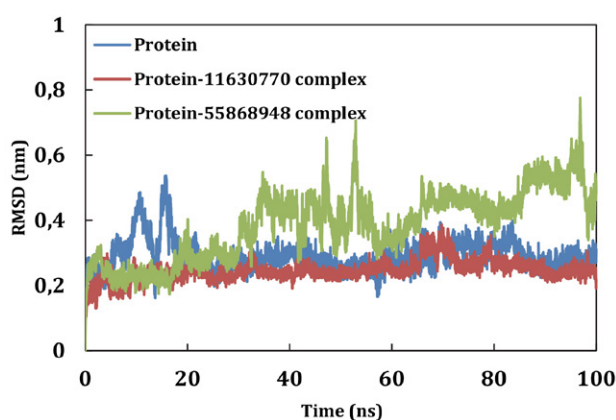


Fig. 7. Root mean square deviation (RMSD) of HIV-1 protease in the absence and presence of 11630770 and 55868948.

#### 3. 9. 2. Assessment of Rg, SASA and Energies

The mass-weighted root mean square distance of a collection of atoms from their common center of mass is

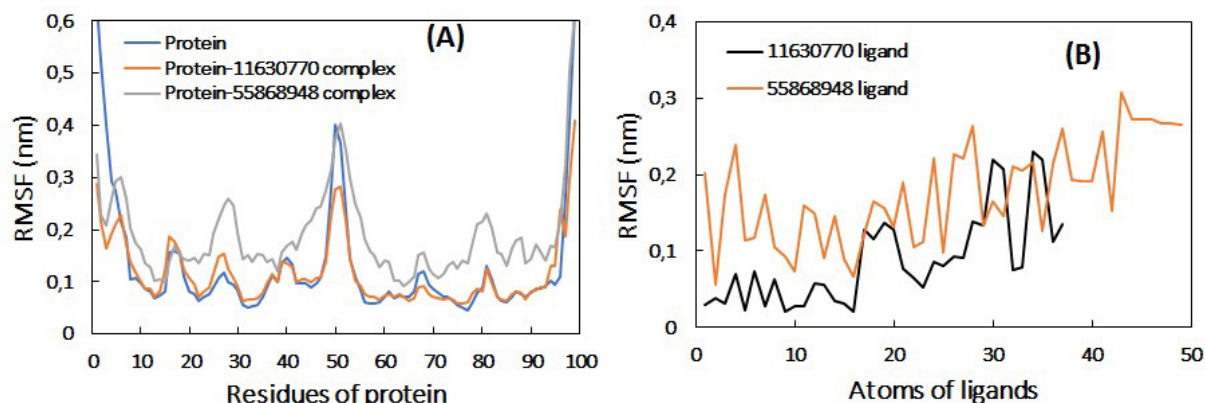


Fig. 8. Root mean square fluctuation (RMSF) of HIV-1 protease in the absence and presence of 11630770 and 55868948. (B) The average RMSF value of each atom of ligands during the MD simulation.

defined as radius of gyration ( $R_g$ ). The  $R_g$  is also considered as indicator of the stability of proteins or their complexes during MD.<sup>49</sup> Generally, compact or globular proteins exhibit lesser variations in their  $R_g$  compared to the expanded form of proteins (5V4Y).

The changes in  $R_g$  of 5v4y HIV-1 protease alone and in complex with ligands are shown in figure 9A. The  $R_g$  of 5v4y HIV-1 protease alone and 5v4y HIV-1 protease CID 11630770 complex was found to be stable during entire simulation period with negligible variations. On contrary, the  $R_g$  of 5v4y HIV-1 protease CID 55868948 complex showed relatively more variations that may be due to the dynamic behavior of the ligand. The average RMSD of HIV-1 protease alone, 5v4y HIV-1 protease CID 11630770 complex, and 5v4y HIV-1 protease CID 55868948 complex was found to be  $1.278 \pm 0.021$ ,  $1.265 \pm 0.016$ , and  $1.301 \pm 0.027$  nm, respectively. These values were statistically insignificant with respect to the control (HIV-1 protease alone).

Solvent accessible surface area (SASA) of proteins is also taken into account while studying the stability of pro-

teins during MD simulation.<sup>50</sup> SASA of HIV-1 protease in the absence and the presence of ligands over simulation time is presented in figure 9B. The average SASA of HIV-1 protease alone, HIV-1 protease CID 11630770 complex, and HIV-1 protease CID 55868948 complex was found to be  $64.318 \pm 2.388$ ,  $63.049 \pm 2.194$ , and  $65.833 \pm 1.995$  nm<sup>2</sup>, respectively. The negligible variations in SASA of these structures further confirm their stable nature under physiological conditions.

Further verification of the stability of HIV-1 protease and complexes was performed by calculating the physicochemical parameters such as potential energy and total energy (Figure S2). The straight line with negligible fluctuations in potential energy and total energy shows that the system reached equilibrium and remained stable during the entire simulation period.<sup>51</sup> The RMSD,  $R_g$  and SASA values of both complexes showed statistically insignificant variations with respect to the potential and total energies HIV-1 protease alone (Table 8).

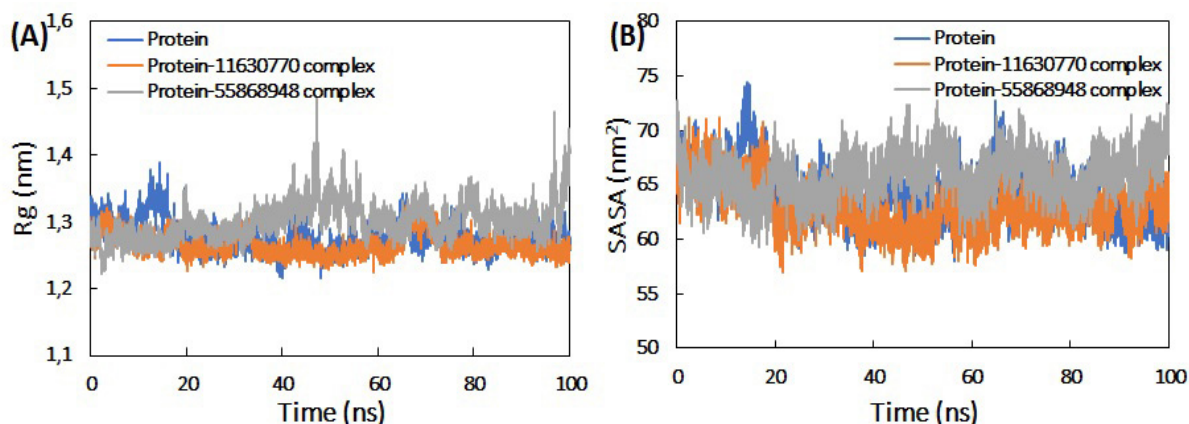


Fig. 9. (A) Radius of gyration ( $R_g$ ) of 5v4y HIV-1 protease in the absence and presence of CID 11630770 and CID 55868948 as a function of simulation time. (B) Solvent accessible surface area (SASA) of HIV-1 protease in the absence and presence of CID 11630770 and CID 55868948 as a function of simulation time.



**Table 8.** Average RMSD,  $R_g$  and SASA of HIV-1 protease alone, and their complexes calculated over 10000 frames of 100 ns MD simulation.

| Parameters                  | Proteins/complexes              | Values             |
|-----------------------------|---------------------------------|--------------------|
| RMSD (nm)                   | HIV-1 protease only             | $0.290 \pm 0.047$  |
|                             | HIV-1 protease 11630770 complex | $0.246 \pm 0.031$  |
|                             | HIV-1 protease 55868948 complex | $0.386 \pm 0.105$  |
|                             | 11630770 only                   | $0.096 \pm 0.030$  |
|                             | 55868948 only                   | $0.191 \pm 0.060$  |
| $R_g$ (nm)                  | HIV-1 protease only             | $1.278 \pm 0.021$  |
|                             | HIV-1 protease 11630770 complex | $1.265 \pm 0.016$  |
|                             | HIV-1 protease 55868948 complex | $1.301 \pm 0.027$  |
|                             | 11630770 only                   | $0.342 \pm 0.003$  |
|                             | 55868948 only                   | $0.520 \pm 0.033$  |
| SASA (nm <sup>2</sup> )     | HIV-1 protease only             | $64.318 \pm 2.388$ |
|                             | HIV-1 protease 11630770 complex | $63.049 \pm 2.194$ |
|                             | HIV-1 protease 55868948 complex | $65.833 \pm 1.995$ |
|                             | 11630770 only                   | $05.227 \pm 0.195$ |
|                             | 55868948 only                   | $06.869 \pm 0.244$ |
| Potential energy (kcal/mol) | HIV-1 protease only             | $-6.360 \pm 0.013$ |
|                             | HIV-1 protease 11630770 complex | $-6.366 \pm 0.014$ |
|                             | HIV-1 protease 55868948 complex | $-6.358 \pm 0.014$ |
| Total energy (kcal/mol)     | HIV-1 protease only             | $-5.096 \pm 0.018$ |
|                             | HIV-1 protease 11630770 complex | $-5.101 \pm 0.018$ |
|                             | HIV-1 protease 55868948 complex | $-5.092 \pm 0.018$ |

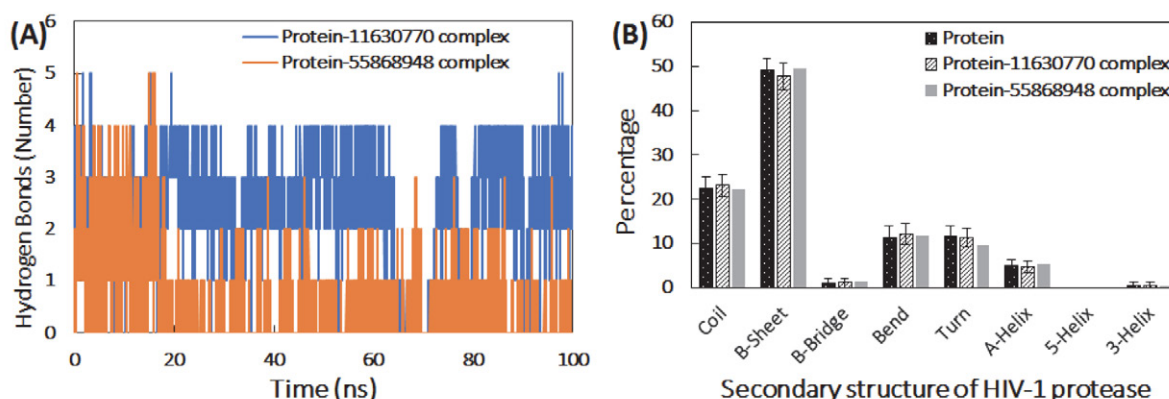
RMSD: Root-mean square deviation,  $R_g$ : Radius of gyration, SASA: Solvent accessible surface area.

### 3. 9. 3. Analysis of Hydrogen Bonds and Secondary Structure

The interaction of ligands with HIV-1 protease was studied by calculating the hydrogens bond profiles for 10000 frames of the MD simulation (Figure 10A). The average number of hydrogen bonds between CID 11630770 and HIV-1 protease was found to be  $2.437 \pm 1.050$ . Lesser number of average hydrogens binds was between CID 55868948 and HIV-1 protease. The hydrogen bond existence was also calculated for both ligands. As observed, the hydrogen bond existence between CID 11630770 and HIV-1 protease was found over the entire simulation period. However, the hydro-

gen bond existence between CID 55868948 and HIV-1 protease showed some breaks as the MD simulation progressed.

The effect of binding of the ligands on the secondary structure of HIV-1 protease was studies by calculating the average secondary structure of all frames of the respective trajectories (Figure 10B). The coil,  $\beta$ -sheet, bends, and turns in HIV-1 protease alone was found to be 22.34, 49.03, 11.18, and 11.55% respectively. These secondary structural components were insignificantly altered in the presence of both ligands. However, there was some increment in  $\alpha$ -helix, which increased from 4.81% to 5.27% due to binding of 55868948



**Fig. 10.** (A) Number of hydrogen bonds the ligands (11630770 and 55868948) and HIV-1 protease over simulation time. (B) Percentage of secondary structure in HIV-1 protease in the absence and presence of 11630770 and 55868948.



### 3. 9. 4. Calculation of Binding Energies and Identification of Key Residues Involved in Binding

The detailed investigation of the various binding energies involved in the interaction of both ligands with HIV-1 protease was done using MM-PBSA calculation. Usually in protein-ligand interactions, non-covalent interactions are predominant. The forces include hydrophobic forces, hydrogen bonds, electrostatic interactions, Van der Waals force. Each of these forces contributes either positively or negatively to the overall binding energy.<sup>52</sup> The MM-PBSA binding energies were calculated by extracting 100 frames from the entire MD simulation trajectories at uniform intervals (Table 9). In the binding of 11630770 to HIV-1 protease, electrostatic interactions were most prominent followed by Van der Waals interactions. The interaction of 55868948 with HIV-1 protease was mostly favored by Van der Waals forces. Additionally, there was also small contribution of SASA energy in the overall binding of both leads. However, polar solvation energy impaired the interaction of both ligands with HIV-1 protease. The overall binding energy for 11630770 and 55868948 were found to be  $-7.067 \pm 0.509$  and  $-7.218 \pm 2.080$  kcal/mol respectively.

From MM-PBSA calculations, the binding energies of all residues can be calculated. The polar, a polar and total binding energy contribution of the key residues of HIV-1 protease in the interaction is presented in figure 10 B. Glu-21, Asp-25, Asp-29, Asp-30, Glu-34, Glu-35, Asp-60, Glu-65, Ile-84, and Phe-99 were the major contributor to overall binding energy in interaction of 11630770 with HIV-1 protease. Similarly, Glu-21, Asp-25, Asp-29, Asp-30, Glu-34, Glu-35, Ile-50, Asp-60, Glu-65, Pro-81, Val-82, Phe-99 of HIV-1 protease contributed maximally in the binding of 55868948 to the protein. It is interesting to note that polar energy of some key residues contributed negatively towards the total binding process.

**Table 9.** Binding free energy (kcal mol<sup>-1</sup>) for the interaction of protein with 11630770 and 55868948 ligands using MMBSA analysis.

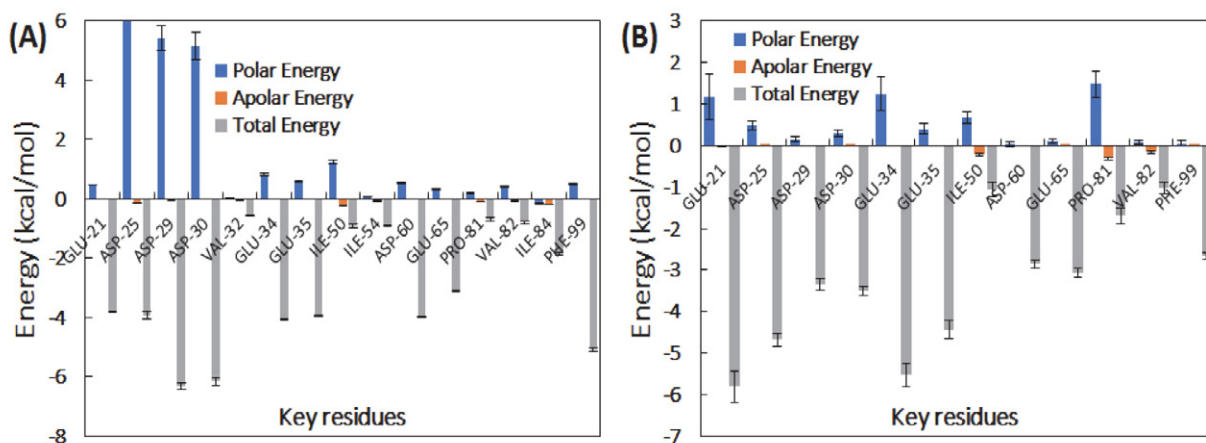
| Type of energy    | Ligands             |                     |
|-------------------|---------------------|---------------------|
|                   | 11630770 ligand     | 55868948 ligand     |
| $\Delta E_{vdW}$  | $-24.258 \pm 0.533$ | $-20.163 \pm 1.340$ |
| $\Delta E_{ele}$  | $-45.156 \pm 1.261$ | $-3.787 \pm 1.817$  |
| $\Delta E_{PSE}$  | $66.022 \pm 1.583$  | $19.443 \pm 1.997$  |
| $\Delta E_{SASA}$ | $-3.650 \pm 0.046$  | $-2.701 \pm 0.163$  |
| $\Delta E_{BE}$   | $-7.067 \pm 0.509$  | $-7.218 \pm 2.080$  |

$\Delta E_{vdW}$ : Van der Waal energy,  $\Delta E_{ele}$ : Electrostatic energy,  $\Delta E_{PSE}$ : Polar solvation energy,  $\Delta E_{SASA}$ : Solvent accessible surface area energy and  $\Delta E_{BE}$ : Binding energy.

### 3. 9. 5. Principal Component Analysis

Principal component analysis (PCA) is the standard statistical procedure used for the investigation of large-scale motion in protein, which is performed by reducing the dimensionality of data set without losing important information, which is characterized by eigenvectors.<sup>53</sup> PCA was done to assess the differences in the flexibility parameters between the HIV-1 protease alone and complexes. Using PCA analysis, a set of eigenvectors and eigenvalues were projected (Figure 12A). HIV-1 protease alone and HIV-1 protease 11630770 complex occupied larger conformational space compared to HIV-1 protease 55868948 complex. These observations denote the presence of more structural stability HIV-1 protease 11630770 complex than HIV-1 protease 55868948 complex.

Moreover, the free energy landscapes for the protein alone and both complexes were plotted to decipher the variations in the protein folding patterns (Figure 13). Variations in the projection of free energy were recorded where alone energetically favorable and relatively stable was compared to both the complexes. The observations show that the binding of ligands partly perturbed the conformation of HIV-1 protease.



**Fig. 11.** (A) Polar, apolar and total energy contributions of the key residues of HIV-1 protease for binding of 11630770. (B) Polar, apolar and total energy contributions of the key residues of HIV-1 protease for binding of 55868948.

To gain further insights into the conformational transition of HIV-1 protease, the lowest energy minima structures were extracted. The Ramachandran plots were made for the energy minima structures of HIV-1 protease alone and in complex with the ligands (Figure 12B). The phi ( $\phi$ ) and psi ( $\psi$ ) angles for HIV-1 protease alone were found to be  $-82.31$  and  $93.60$ , respectively. The  $\phi$  and  $\psi$  angles for HIV-1 protease 11630770 complex were obtained as  $-85.15$  and  $91.30$ , respectively. Similarly,  $\phi$  and  $\psi$  angles for HIV-1 protease CID 55868948 complex were recorded as  $-85.14$  and  $84.72$ , respectively. In both the complexes, remarkable variations in the dihedral angles with respect

Donor (HBD), Hydrogen-Bond Acceptor (HBA), hydrophobic and aromatic groups, which may be responsible for the HIV1 inhibition. Both generated pharmacophore model was validated for its quality to identify new reliable chemical compounds. The validation procedure included two methods: test set validation, and decoy set validation. Based on the 3D-QSAR model ADRRR we have selected hits from the PubChem database. Initial screened compounds were passed through several criteria including the range of activity of training set, fitness score more than 2.5 and comparison of dock score and binding energy with Nelfinavir to reach the potential compounds. Docking into

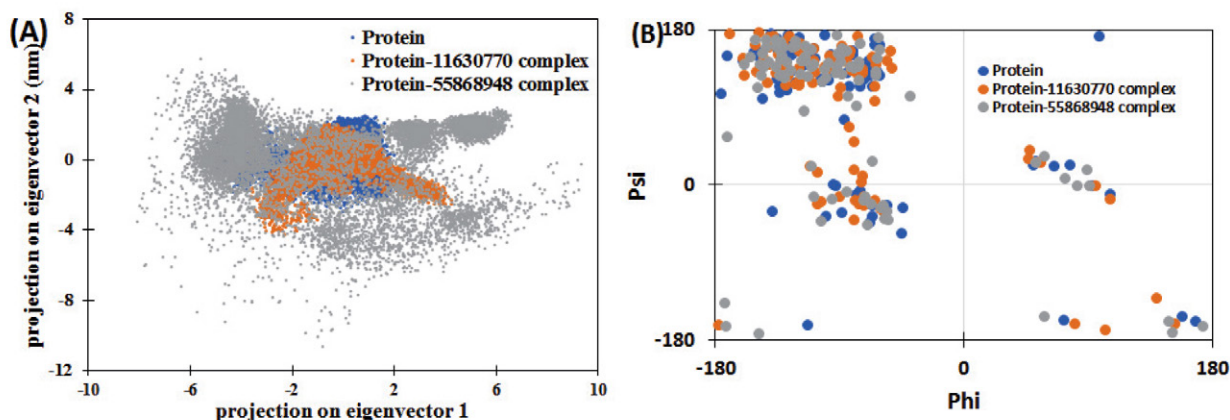


Fig. 12. (A) Principal component analysis (PCA) of HIV-1 protease in the absence and presence of 11630770 and 55868948. (B) Ramachandran plot of the energy minima of HIV-1 protease in the absence and presence of 11630770 and 55868948.

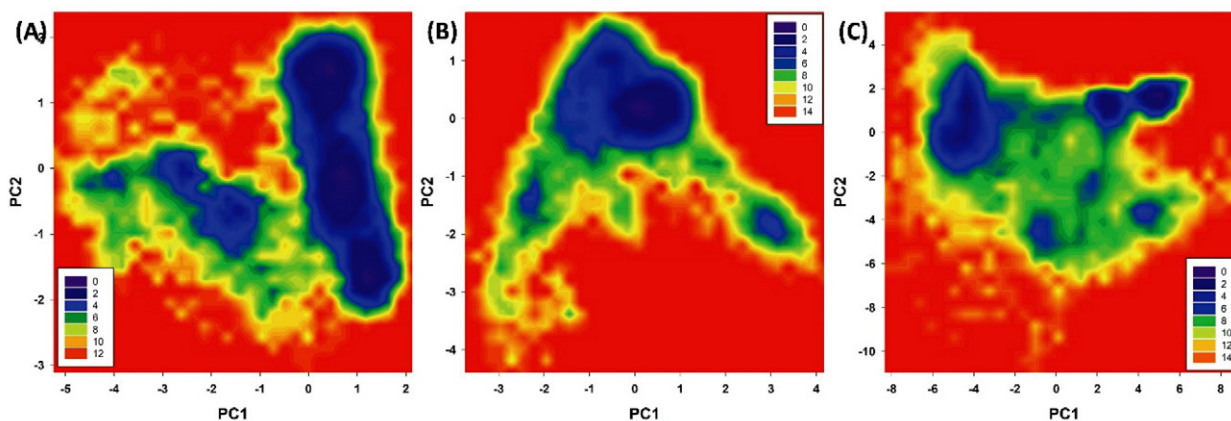


Fig. 13. Free energy landscape plot of (A) HIV-1 protease alone (B) HIV-1 protease 11630770 complex and (C) HIV-1 protease 5586894 complex.

to HIV-1 protease alone were observed, indicating the structural transition in the presence of ligands.

## 4. Conclusion

In this study, a five-featured (ADRRR) pharmacophore model for the HIV1 inhibitors was developed. The generated model revealed the importance of Hydrogen-Bond

predicted active site conclusively infers those hydrophobic contacts possess more dominance compared to other interactions. Subsequently, amide group play an essential role in hydrogen binding. ADME properties of the six hits were found to be in accordance with known chemically and biologically active compounds. Conclusively, the hits obtained on virtual screening of the database have provided new chemical starting points for design and development of novel HIV-1 inhibitory agents.

## 5. References

1. B. Gazzard, J. Anderson, A. Babiker, M. Boffito, G. Brook, G. Brough, D. Churchill, B. Cromarty, S. Das, M. Fisher, A. Freedman, A. M. Geretti, M. Johnson, S. Khoo, C. Leen, D. Nair, B. Peters, A. Phillips, D. Pillay, A. Pozniak, J. Walsh, E. Wilkins, I. Williams, M. Williams, M. Youle, *HIV Med.* **2008**, *9*, 563–608. DOI:10.1111/j.1468-1293.2008.00636.x
2. Global HIV & AIDS statistics-Fact sheet. **ONUSIDA**. <https://www.unaids.org/en/resources/fact-sheet>
3. S. J. Pocock, D. R. Elbourne, *N. Engl. J. Med.* **2000**, *342*, 1907–1909. DOI:10.1056/NEJM200006223422511
4. C. F. Gilks, S. Crowley, R. Ekpini, S. Gove, J. Perriens, Y. Souteyrand, D. Sutherland, M. Vitoria, T. Guerna, K. De Cock, *the Lancet*. **2006**, 368, 505–510. DOI:10.1016/S0140-6736(06)69158-7
5. A. B. Kleinpeter and E. O. Freed, *Viruses*, **2020**, *12*, E940. DOI:10.3390/v12090940
6. E. Urano, U. Timilsina, J. A. Kaplan, S. Ablan, D. Ghimire, P. Pham, N. Kuruppu, R. Mandt, S. R. Durell, T. J. Nitz, D. E. Martin, C. T. Wild, R. Gaur, E. O. Freed, *J. Virol.* **2019**, *93*, 2017–18. DOI:10.1128/JVI.02017-18
7. S. Mattei, A. Tan, B. Glass, B. Müller, H.G. Kräusslich, J. A. G. Briggs, *Proc. Natl. Acad. Sci. USA*, **2018**, *115*, 9401–9410. DOI:10.1073/pnas.1811237115
8. Y. R. Lin, S. M. Chu, F. H. Yu, K.-J. Huang, C. T. Wang, *BMC Microbiol.*, **2022**, *22*, 94. DOI:10.1186/s12866-022-02503-3
9. C. Voshavar, *Curr. Top. Med. Chem.*, **2019**, *19*, 1571–1598. DOI:10.2174/1568026619666190619115243
10. Z. Lv, Y. Chu, Y. Wang, *HIV AIDS (Auckl.)*. **2015**, *7*, 95–104. DOI:10.2147/HIV.S79956
11. A. V. Veselovsky, A.S. Ivanov, *Curr. Drug Targets Infect. Disord.* **2003**, *3*, 33–40. DOI:10.2174/1568005033342145
12. C. M. Song, S. J. Lim, J. C. Tong, *Brief. Bioinform.* **2009**, *10*, 579–591. DOI:10.1093/bib/bbp023
13. S. Y. Yang, *Drug Discov. Today*, **2010**, *15*, 444–450. DOI:10.1016/j.drudis.2010.03.013
14. S. A. Khedkar, A. K. Malde, E. C. Coutinho, S. Srivastava, *J. Med. Chem.* **2007**, *3*, 187–197. DOI:10.2174/157340607780059521
15. H. Kubinyi, *3D QSAR in Drug Design: Volume 1, Theory Methods and Applications*, Springer Science & Business Media, Berlin, Germany, **1993**.
16. R. Dias, W. F. De Azevedo, *Curr. Drug Targets*, **2008**, *9*, 1040–1047. DOI:10.2174/138945008786949432
17. R. Ragno, A. Coluccia, G. La Regina, G. De Martino, F. Piscitelli, A. Lavecchia, E. Novellino, A. Bergamini, C. Ciapini, A. Sinistro, G. Maga, E. Crespan, M. Artico, R. Silvestri, *J. Med. Chem.* **2006**, *49*, 3172–3184. DOI:10.1021/jm0512490
18. LigPrep, Schrödinger Release 2021-4, Schrödinger, LLC, New York, NY, **2021**. <https://www.schrodinger.com/products/lig-prep>
19. E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J.Y. Xiang, L. Wang, D. Lupyran, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, E. D. Harder, *J. Chem. Theory Comput.*, **2016**, *12*, 281–296. DOI:10.1021/acs.jctc.5b00864
20. K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu, C. Lu, M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R. A. Friesner, E. D. Harder, *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874. DOI:10.1021/acs.jctc.8b01026
21. R. Kunal, K. Supratik, N. D. Rudra, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Elsevier, Amsterdam, Pays-Bas, **2015**; DOI:10.1016/C2014-0-00286-9
22. W. J. Dunn, D. Rogers, *Genetic Partial Least Squares in QSAR. Genetic Algorithms in Molecular Modeling*. Academic Press, London, **1996**. Pp. 109–130. DOI:10.1016/B978-012213810-2/50006-2
23. A. R. Leach, V. J. Gillet, R. A. Lewis, R. Taylor, *J. Med. Chem.* **2010**, *53*, 539–558. DOI:10.1021/jm900817u
24. P. Gramatica, *QSAR Comb. Sci.*, **2007**, *26*, 694–701. DOI:10.1002/qsar.200610151
25. A. Golbraikh, A. Tropsha, *Mol. Divers.* **2000**, *5*, 231–243. DOI:10.1023/A:1021372108686
26. K. Roy, *Expert. Opin. Drug Discov.*, **2007**, *2*, 1567–1577. DOI:10.1517/17460441.2.12.1567
27. J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, *J. Comput. Aided Mol. Des.* **2008**, *22*, 213–228. DOI:10.1007/s10822-007-9163-6
28. M. D. Mackey, J. L. Melville, *J. Chem. Inf. Model.* **2009**, *49*, 1154–1162. DOI:10.1021/ci8003978
29. M. Lill, *Virtual Screening in Drug Design, Methods Mol. Biol.* **2013**, *993*, 1–12. DOI:10.1007/978-1-62703-342-8\_1
30. E. Lionta, G. Spyrou, D. K. Vassilatis, Z. Cournia, *Curr. Top. Med. Chem.* **2014**, *14*, 1923–1938. DOI:10.2174/1568026614666140929124445
31. S. Kim, P. A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, *Nucleic Acids Res.* **2016**, *44*, 202–213. DOI:10.1093/nar/gkv951
32. M. Amano, P. Miguel Salcedo-Gómez, R.S. Yedidi, N. S. Delino, H. Nakata, K. Venkateswara Rao, *Sci. Rep.* **2017**, *7*, 12235. DOI:10.1038/s41598-017-12052-9
33. Schrödinger Suite 2021 Protein Preparation Wizard; Epik Schrödinger, LLC: New York, NY, USA, **2021**.
34. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, *J. Med. Chem.*, **2006**, *49*, 6177–6196. DOI:10.1021/jm051256o
35. Dassault Systèmes BIOVIA Discovery Studio Modeling Environment, Release **2017**
36. L. Ioakimidis, L. Thoukydidis, A. Mirza, S. Naeem, J. Reynison, *QSAR Comb. Sci.* **2008**, *27*, 445–456. DOI:10.1002/qsar.200730051
37. P. Banerjee, A. O. Eckert, A. K. Schrey, R. Preissner, *Nucleic Acids Res.* **2018**, *46*, 257–263. DOI:10.1093/nar/gky318
38. H. J. C. Berendsen, D. Van der Spoel, R. van Drunen, *Comput. Phys. Commun.*, **1995**, *91*, 43–56. DOI:10.1016/0010-4655(95)00042-E
39. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. DOI:10.1021/acs.jctc.5b00255
40. J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph.*

- Model.* **2006**, 25, 247–260. DOI:10.1016/j.jmngm.2005.12.005
41. M. Topel, A. L. Ferguson, *J. Chem. Phys.* **2020**, 153, 194–102. DOI:10.1063/5.0024732
42. M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, 52, 7182–7190. DOI:10.1063/1.328693
43. R. Kumari, R. Kumar, A. Lynn, *J. Chem. Inf. Model.* **2014**, 54, 1951–1962. DOI:10.1021/ci500020m
44. Schrödinger Release 2021-2: phase; Schrödinger, LLC: New York, NY, USA, **2021**.
45. Bardsley-Elliott, G. L. Plosker, *Drugs.* **2000**, 59, 581–620. DOI:10.2165/00003495-200059030-00014
46. B. Jarvis, D. Faulds, *Drugs.* **1998**, 56, 147–167. DOI:10.2165/00003495-199856010-00013
47. P. P. Mager, *Med. Res. Rev.* **2001**, 21, 348–353. DOI:10.1002/med.1012
48. L. K. Nicholson, T. Yamazaki, D. A. Torchia, S. Grzesiek, A. Bax, S.J. Stahl, *Nat. Struct. Mol. Biol.* **1995**, 2, 274–280. DOI:10.1038/nsb0495-274
49. F. A. Qais, T. Sarwar, I. Ahmad, R. A. Khan, S. A. Shahzad, F. M. Husain, *Int. J. Biol. Macromol.* **2021**, 169, 143–152. DOI:10.1016/j.ijbiomac.2020.12.096
50. B. Rath, F. Abul Qais, R. Patro, S. Mohapatra, T. Sharma, *Bioorg. Med. Chem. Lett.* **2021**, 41, 128029. DOI:10.1016/j.bmcl.2021.128029
51. P. V. Coveney, S. Wan, *Chem. Phys.* **2016**, 18, 30236–30240. DOI:10.1039/C6CP02349E
52. S. Siddiqui, F. Ameen, T. Kausar, S. M. Nayeem, S. Ur Rehman, M. Tabish, *Spectrochim. Acta A*, **2021**, 249, 119296. DOI:10.1016/j.saa.2020.119296
53. S. Siddiqui, F. Ameen, I. Jahan, S. M. Nayeem, M. Tabish. *New J. Chem.* **2019**, 43, 4137–4151. DOI:10.1039/C8NJ05486J

## Povzetek

Ta raziskava obravnava *in silico* 45 indolil-aril-sulfonov, znanih kot anti-HIV1. Podatki so bili zbrani iz nedavnih predhodno prijavljenih zaviralcev in razdeljeni na podskupino 33 spojin, namenjenih nizu preizkušanj, preostalih 12 spojin pa je bilo shranjenih v namen testnega niza. Izbrani farmakofor –ADRRR– je podal statistično pomemben model 3D-QSAR, ki vključuje visoke ocene zaupanja ( $R^2 = 0,930$ ,  $Q^2 = 0,848$  in  $RMSE = 0,460$ ). Napovedna moč uveljavljenega modela farmakofora je bila potrjena z zunanjim testom ( $r^2 = 0,848$ ). Sistematičen navidezni potek dela je pokazal visok faktor uporabnosti in razkril visoko moč napovedovanja. Nato je bil model uporabljen za pregled filtrirane baze podatkov PubChem in označil vse kemijske značilnosti modelne farmakofore. Ustrezni zadetki so bili dodatno ocenjeni z *in silico* ADMET študijami. Poleg tega se molekularna dinamika lahko uporablja tudi za raziskovanje stabilnosti dobljenih kompleksov. V končni fazi bodo te izbrane komponente lahko postale dobra začetna molekula za razvoj učinkovitih učinkovin proti HIV-1.



Except when otherwise noted, articles in this journal are published under the terms and conditions of the Creative Commons Attribution 4.0 International License