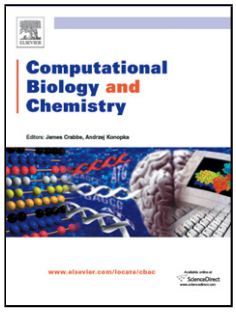
# Journal Pre-proof




*De novo* designed based identification of potential HIV-1 integrase inhibitors: A pharmacoinformatics study

Pooja Balasaheb Shinde (Data curation) (Formal analysis) (Investigation) (Methodology), Shovonlal Bhowmick (Data curation) (Formal analysis) (Investigation) (Methodology) (Writing - original draft)<ce:contributor-role>Writing – review and editing), Etidal Alfantoukh (Conceptualization)<ce:contributor-role>Writing – review and editing), Pritee Chunarkar Patil (Conceptualization)<ce:contributor-role>Writing – review and editing), Saikh Mohammad Wabaidur (Conceptualization)<ce:contributor-role>Writing – review and editing), Rupesh V Chikhale (Conceptualization)<ce:contributor-role>Writing – review and editing), Md Ataul Islam (Conceptualization) (Investigation)<ce:contributor-role>Super-vision) (Data curation) (Writing - original draft)<ce:contributor-role>Writing – review and editing)

PII: S1476-9271(20)30241-3

DOI: https://doi.org/10.1016/j.compbiolchem.2020.107319

Reference: CBAC 107319

To appear in: *Computational Biology and Chemistry*

Received Date: 10 March 2020

Revised Date: 10 May 2020

Accepted Date: 22 June 2020

Please cite this article as: Balasaheb Shinde P, Bhowmick S, Alfantoukh E, Patil PC, Wabaidur SM, Chikhale RV, Islam MA, *De novo* designed based identification of potential HIV-1 integrase inhibitors: A pharmacoinformatics study, *Computational Biology and*

*Chemistry* (2020), doi: https://doi.org/10.1016/j.compbiolchem.2020.107319

# *De novo* designed based identification of potential HIV-1 integrase inhibitors: A pharmacoinformatics study

Pooja Balasaheb Shinde[1#], Shovonlal Bhowmick[2#], Etidal Alfantoukh[3], Pritee Chunarkar Patil[1], Saikh Mohammad Wabaidur[4], Rupesh V Chikhale[5], Md Ataul Islam[6,7,8*]

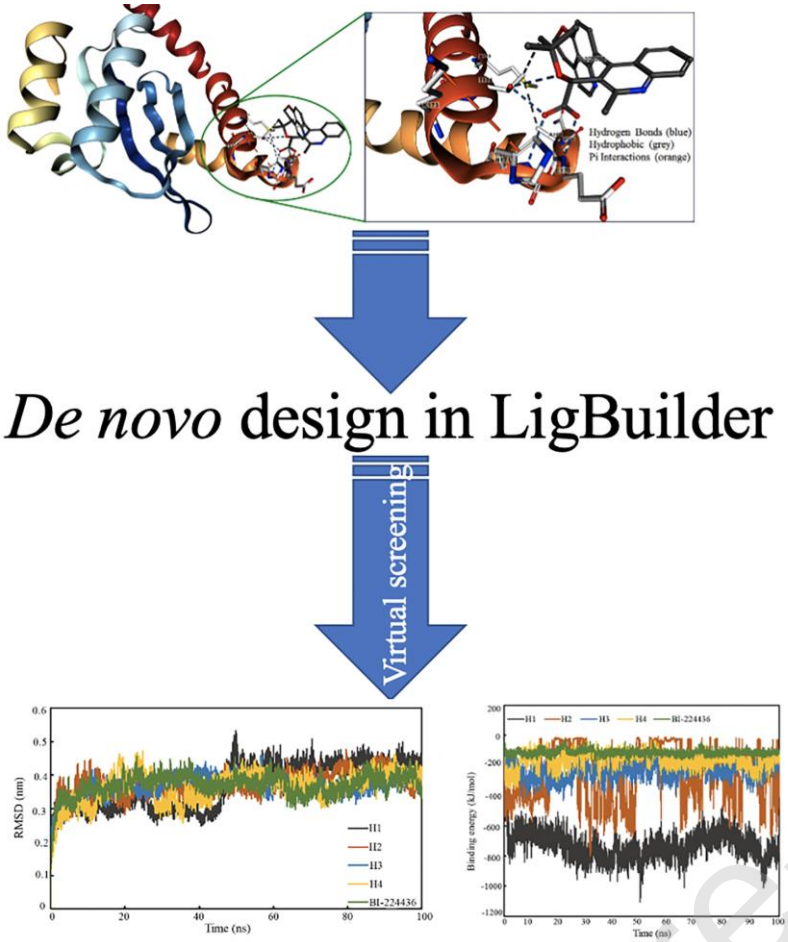
[1]*Department of Bioinformatics, Rajiv Gandhi Institute of IT and Biotechnology, Bharati Vidyapeeth Deemed University, Pune-Satara Road, Pune, India*
[2]*Department of Chemical Technology, University of Calcutta, 92, A.P.C. Road, Kolkata - 700009, India*
[3]*Health Sciences Research Center, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia*
[4]*Department of Chemistry P.O. Box 2455, College of Science, King Saud University, Riyadh, 11451, Saudi Arabia*
[5]*School of Pharmacy, University of East Anglia, Norwich, United Kingdom*
[6]*Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom*
[7]*School of Health Sciences, University of Kwazulu-Natal, Westville Campus, Durban, South Africa*
[8]*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa*

[*]To whom correspondence should be addressed: ataul.islam80@gmail.com (MA Islam)

[#]Contributed equally and can be considered as the first author.



**Graphical abstract**




Highlights

- Complex between BI-224436 and HIV-1 integrase was used to design novel anti-HIV inhibitors.
- The *de novo* approach available in LigBuilder was used.
- A number of criteria was used to narrow down the chemical space of initial designed 80000 inhibitors.
- Different pharmacokinetics parameters were checked for final proposed molecules.
- MD simulation was performed of HIV-1 integrase complex with proposed molecules.
- The binding energy was calculated using MM-PBSA approach.



**Abstract**

In the present study, pharmacoinformatics paradigms include receptor-based *de novo* design, virtual screening through molecular docking and molecular dynamics (MD) simulation are

implemented to identify novel and promising HIV-1 integrase inhibitors. The *de novo* drug/ligand/molecule design is a powerful and effective approach to design a large number of novel and structurally diverse compounds with the required pharmacological profiles. A crystal structure of HIV-1 integrase bound with standard inhibitor BI-224436 is used and a set of 80000 compounds through the *de novo* approach in LigBuilder is designed. Initially, a number of criteria including molecular docking, *in-silico* toxicity and pharmacokinetics profile assessments are implied to reduce the chemical space. Finally, four *de novo* designed molecules are proposed as potential HIV-1 integrase inhibitors based on comparative analyses. Notably, strong binding interactions have been identified between a few newly identified catalytic amino acid residues and proposed HIV-1 integrase inhibitors. For evaluation of the dynamic stability of the protein-ligand complexes, a number of parameters are explored from the 100ns MD simulation study. The MD simulation study suggested that proposed molecules efficiently retained their molecular interaction and structural integrity inside the HIV-1 integrase. The binding free energy is calculated through the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach for all complexes and it also explains their thermodynamic stability. Hence, proposed molecules through *de novo* design might be critical to inhibiting the HIV-1 integrase.

**Keywords:** HIV-1 Integrase; *De novo* design; Virtual screening; Molecular docking; Molecular dynamics; MM-PBSA

## 1. Introduction

The Human Immune deficiency Virus (HIV) infections can lead to the acquired immune deficiency syndrome (AIDS) which is a medical condition and can be simply labeled as a curse upon the human race. The progression of HIV infection is highly variable and depends on the individual's immune system [1]. HIV infection continues to be a major global public health concern, as estimated by approximately 37.9 million people (according to the report of World Health Organization) living with HIV till the end of 2018 [2]. Among all, approximately 25.7 million people are living with HIV in the African Region. Moreover, at the end of the year 2018, 62% of HIV infected people received antiretroviral therapy (ART). Most recent data indicates that about 24.5 million people accessed ART during June 2019. However, due to long-term extensive gaps in HIV management and awareness services, approximately 0.7 million people died from HIV-associated causes in 2018 [2]. As of now, there is no definite curative measure for HIV infection. However, antiretroviral drugs (ARVs)

can only control or suppress the HIV virus and subsequently help to prevent onward transmission to other individuals [3, 4]. There are several numbers of ARVs, and they can be combined with different ways for treating HIV without any restrictions of CD4 counts [1].

In the current study, an attempt has been taken to identify potential anti-retroviral modulator or inhibitor hits against the HIV-1 integrase protein using computational multiple molecular modeling approaches. The HIV-1 integrase protein is one of the essential enzymes which drives viral replication processes and also responsible for the obligatory integration of viral DNA into the host chromosome via covalent insertion [5-8]. Thus for long decades, the HIV-1 integrase has been considered to be an experimentally well-validated target for clinically effective anti-AIDS therapeutics developments. It can also be effective to manage against drug-resistant mutants HIV infections [6, 9-13]. In general, HIV-1 integrase is comprised of three distinct functional domains: the C-terminal domain (CTD), the catalytic core domain (CCD), and the N-terminal domain (NTD) [14, 15]. Each domain has distinct characteristics such as NTD region (residues 1-50) contains high conserved His and Cys amino acid residues (or HHCC motif) and known to help in chelation of zinc ion. The CCD region of HIV-1 integrase, spanning amino acids 51–212, contains endonuclease site as well as highly conserved acidic amino acid residues such as Asp64, Asp116 and Glu152 (known as DDE motif). The CDD is considered to be an important instrumental part that is conserved among all retroviral integrase proteins. Majorly, this CCD region of HIV-1 integrase contains the enzyme catalytic site(s) [16]. Any mutation in these residues diminishes the catalytic activity. Indeed, all three domains cooperate for the variety of the enzymatic functions catalyzed by the integrase protein [17]. On the other hand, CTD region (amino acid residues: 213–288) specifically plays a role in DNA binding and is responsible for stabilizing the DNA-protein complex once bonded with DNA [15]. In the present scenario, comparison to understanding the role of CCD region, the biological role of NTD and CTD regions are less understood. Although, it is known that all three domains are essential and responsible for 3’-p processing and strand transfer activity. In this study, comprehensively emphasis has been given on CCD region of integrase which plays an indispensable role for demonstrating integrase-mediated enzymatic activities [18]. Moreover, the HIV integrase protein can exists in dimer or in higher oligomeric states but their structural and functional aspects are yet under the research and developmental phase. Moreover, for the last two decades, the CCD region of integrase has been under special attention for assisting and expediting the future drug design efforts [19, 20]. Although, integrase strand transfer inhibitors (INSTIs) target the HIV-1 integrase active site. However, increasing genesis of drug resistance circumstances and numerous side

effects pose striking challenges to INSTI-based therapies moving ahead [21]. Among various clinically available INSTIs, such as Raltegravir, Elvitegravir, and Dolutegravir are now being used in clinical practice [22-25]. These drugs also show adverse or toxic, neuropsychiatric, osteomuscular or digestive complications [26, 27]. Moreover, INSTI-resistant mutants can also emerge during treatment, that can cause resistance to the approved first-generation INSTIs [28-31]. Resistance mutations usually take place in proximal to the HIV-1 integrase active site, primarily in the β4α2 loop of the CCD [9]. Such instances motivate the researchers to continue to study the mechanisms of HIV drug resistance and develop novel therapeutics, a top priority for HIV research. Previously, a number of experimental as well as computational approaches were undertaken precisely targeting the HIV-1 integrase to identify promising chemical entities that can provide therapeutic and preventive regimes or can act as antiretroviral therapy and vaccines for managing HIV/AIDS infection [20, 32-46]. However, complete eradication of HIV infection is a daunting objective that requires definitive treatment strategies which are still missing.

In the present era of pharmaceutical development for novel drug discovery, besides the high-throughput virtual screening, the *de novo* drug/ligand design technique is getting much faster attention as one of the finest complementary tools [47-50]. Most importantly, the significant elicitation of the *de novo* design methodology is a novel molecular structure generation method with very low cost and less time taking in comparison to the traditional drug discovery techniques [51, 52]. Herein, in order to find few potential novel HIV-1 integrase inhibitors, the receptor-based *de novo* drug design technique is employed using the LigBuilder tool followed by virtual screening carried out through molecular docking and molecular dynamics (MD) simulations studies. Finally, the binding free energies of the final molecules are estimated. Additionally, the important residues of HIV-1 integrase protein have already been reported which participate in hydrogen bond (H-bond) and non-H-bond interactions crucial for blocking the integrase activity. Finally, the study has reported four structurally diverse compounds designed through the *de novo* technique capable to decipher strong molecular interactions with viral HIV-1 integrase protein. Hence, reported chemical entities identified through extensive *de novo* design techniques can optimally block the integrase activity and therefore capable to opt for developing a next-generation newer class of HIV-1 integrase inhibitors subjected to experimental validation.

## 2. Materials and methods

The crystal structure of the HIV-1 integrase was obtained from the RCSB-Protein Data Bank (PDB) having PDB ID 6NUJ (53). This is the most recently resolved crystalized HIV-1 integrase (CCD) with resolution and R-value of 2.1 Å and 0.244 respectively. In order to select the HIV-1 integrase few parameters were checked included resolution, date of deposition in the PDB and bound small molecule. The selected HIV-1 integrase was found to be consisting of 166 amino acid residues. The standard HIV-1 integrase inhibitor, BI-224436 was bound inside the CCD of the selected protein. The quinoline-based compound, BI-224436 is a first non-catalytic site HIV-1 integrase inhibitor having an effective antiviral activity (54). It is reported that the BI-224436 inhibits HIV replication through association at the conserved allosteric pocket of the HIV integrase enzyme. The molecular weight and exact mass of the BI-224436 are reported as 442.50 and 442.18 g/mol, respectively. Throughout the study, BI-224436 was selected as a control compound and used for a comparative perspective. The position of catalytic and allosteric sites of HIV-1 integrase protein were identified and it is given in Figure S1 (Supplementary data).

**2.1 Molecular docking**

Prior to molecular docking, it is essential to prepare the protein and ligands. The HIV-1 integrase enzyme was prepared in the graphical user interface program AutoDock Tool (ADT) (55). The missing atoms and side chains were checked and subsequently repaired in protein structure. A number of required polar hydrogens were added to the protein structure. The Gasteiger charge was assigned. The prepared protein was saved in .pdbqt file format for further uses. For the preparation of the standard ligand, BI-224436 was first re-drawn using MarvinSketch (v. 17.8.0) and followed by the addition of hydrogen atoms and a required number of active torsions were assigned. Re-drawing of the standard compound was carried out to accurately get the optimized chemical structure which must be needed in docking study. Similar to the protein structure, the Gasteiger charge was also adjusted to the standard compound. Finally, the standard compound BI-224436 was saved as .pdbqt file format. Such a specific file format is required for the input file in Autodock Vina during docking execution. Simultaneously, for the large number of generated HIV-1 integrase inhibitors designed through *de novo* technique were also converted into .pdbqt format using the OpenBabel (56), a freely available file format conversion tool. During the structural conversion, the appropriate number of hydrogen atoms were added and charge adjusted to each ligand. The grid was generated by confining the bound ligand BI-224436. The grid coordinate was assigned as (-14.162Å) x (43.427Å) x (-14.494Å) along the X-, Y- and Z-

axes, respectively. The size of the grid box dimension was set to 60Å x 60Å x 60Å. The prepared protein and ligands along with the grid information were kept in a configuration file that was considered for molecular docking study in AutoDock Vina (ADV) tool (57) installed in the CHPC server at Cape Town, South Africa (https://www.chpc.ac.za/). During the execution of molecular docking, the protein structure was retained as rigid, while the ligand structures were considered as flexible. However, firstly to validate the molecular docking protocol, the self-docking approach was used. In this method, the bound standard ligand (BI-224436) was re-docked in the same active site cavity of HIV-1 integrase where it was originally attached. It is reported that the root means square deviation (RMSD) less than 2 Å between best docked pose and crystalized ligand can predict the perfect orientation of any small molecule (58). After self-docking, the re-docked best pose of BI-224436 was extracted and superimposed with the original crystalized pose of the same and subsequently RMSD was calculated. In the present study, generated docked poses for all the designed compounds and their binding energies were explored to analyze the molecular docking outcomes. Thereafter, binding interactions in 2D and 3D-space were explored through the Maestro interface of Schrodinger and PyMol.

**2.2 *De novo* design** using LigBuilder tool

Among several structure-based drug design (SBDD) and ligand-based drug design (LBDD) approaches, the *de novo* design technique is one of the crucial and pivotal methods to build small molecules with desired pharmacological properties. In the *de novo* design technique, the 3D structure of the receptor was used to design a large number of newer molecules with high potential and therapeutic value. A common phenomenon of *de novo* design technique is the generation of large ligands as the greater molecular size inclines to accomplish the higher binding affinity (59). However, the fact is, sometimes it may produce false-positive outcomes as well due to size/shape and negligence of molecular interactions complementarity between protein and ligand. Keeping that view in consideration, in the current study, widely accepted LigBuilder v2.0 - an open-source *de novo* library designed tool (60) which follows genetic algorithm to construct ligands library, was used to design the HIV-1 integrase inhibitors. Precisely, LigBuilder is a highly acknowledged multi-purposed tool used for structure-based *de novo* drug design and optimization. LigBuilder can automatically build-up ligand compounds within the user defined binding pocket and subsequently screen them based on the 3D structure of the target protein. Herein, the 3D crystal structure of the HIV-1 integrase bound with the standard inhibitor of the same was used for the LigBuilder. There are mainly

three features to follow and the execution of LigBuilder to generate the *de novo* molecules. In general, LigBuilder required a material library as input. In this purpose the large chemical database, ZINC was used as the material library. The LigBuilder has two main functional modules for execution, i.e. 'cavity' and 'build'. The 'cavity' module takes the receptor and best scoring small molecules in .pdb and .mol2 format respectively. Herein, the 'cavity' module was used to analyze the binding pocket of the given HIV-1 integrase protein and prepared the data which are necessary to execute the 'build' module. The 'build' is the main functional module for *de novo* design and results analysis. The major function of 'build' module is the construction of ligand molecules for the target protein by applying the fragment-based design strategy. Although 'build' module relies on the binding-site describing in the file generated by 'cavity' module, however, it can take care of most inclusive parameters such lead optimization, fragment linking, design mimicking, binding-affinity estimation, ligand filtering, ligand recommendation, synthesis analysis, substructure search, and molecule clustering, etc. In the current study, out of available three modes available in the LigBuilder program such as 'growing', 'linking' and 'exploring', the 'exploring' mode was used to automatically generate the compounds. Firstly, the 'cavity' module was executed which derives few key interaction sites inside the binding site and, based on that, it suggests potential pharmacophore features. Further, the 'build' module was executed to generate the molecules with a given number of parameters including the number of generation, population size and the maximal number of outputs. In the current effort, the design number of generation was given as 80000. Hence, after successful completion, the LigBuilder was generated a total of 80000 new chemical components as HIV-1 integrase inhibitors. All generated compounds were checked carefully in terms of structural redundancy and similarity or dissimilarity before further assessment.

**2.3 Virtual screening strategies**

Virtual screening is an *in-silico* approach to find out the promising chemical entities from a large dataset for a specific target. The molecules designed in the *de novo* design approach targeting the HIV-1 integrase were further screened through the molecular docking, drug-likeness and *in-silico* pharmacokinetics analyses. Initially, all generated molecules were checked for Lipinski's rule of five (LoF) (61) and Veber's rule (62). The LoF explains the drug-likeness characteristics of the molecules having the hydrophobicity less than 5, molecular weight less than 500, hydrogen bond (HB) donor and acceptor less than 5 and 10 respectively. The flexibility and surface area of promising molecules can be explained

through Veber's rule. According to this rule, being a potential molecule the total polar surface area (TPSA) and the number of rotatable bonds should not be more than 140Å$^2$ and 10 respectively. Further, the binding energy of the BI-224436 was used as a threshold to reduce the chemical space of remaining molecules from the previous step. Molecules having binding energy value higher than BI-224436 were discarded for further study. The toxic and environmental effects of the molecules were checked through the TOxicity Prediction by Komputer Assisted Technology (TOPKAT) module available in Discovery Studio. The TOPKAT predicts the toxic and environmental effects of the chemical substances through cross-validated quantitative structure toxicity relationship (QSTR) models. Molecules showed toxic, mutagenic, irritant and carcinogenic were removed for further analysis. Finally, molecules remained after TOPKAT assessment were used to calculate a number of pharmacokinetics parameters including *n*-octanol and water (log $P_{o/w}$) partition coefficient or lipophilicity, molar solubility in water, blood-brain barrier (BBB) permeability, skin permeation and human gastrointestinal absorption (HIA) capability. On detained analysis, the molecules showed good absorption, distribution, metabolism and excretion were proposed as promising HIV-1 integrase inhibitors.

A number of drug-likeness properties included ligand efficiency (LE), ligand efficiency scale (LE_scale), fit quality (FQ), and LE-dependent lipophilicity(LELP) were explored to assess the potentiality of the molecules. First of all, the LE given by Hopkins *et al*. (63) was calculated as per equation (1). This parameter is the negative ratio between binding energy (Kcal/mol) and the number of heavy atoms (NHA).

$$LE = \frac{-BE}{NHA} \tag{1}$$

The LE_Scale is the size-dependent comparison of the small molecules given by Reynolds *et al*. (64) known as LE_Scale and can be calculated using equation (2).

$$LE\_Scale = 0.873 \times e^{-0.026 \times NHA} - 0.064 \tag{2}$$

Another crucial parameter, the fit quality score was calculated by equation (3). The FQ articulated for good binding characteristics of the small molecule in the receptor.

$$FQ = \frac{LE}{LE\_Scale} \tag{3}$$

Finally, the LELP was calculated by equation (4). This parameter was proposed by Keseru and Makara (65) and can be calculated by getting the ratio between logP and LE.

$$LELP = \frac{\log p}{LE} \tag{4}$$

### 2.4 Molecular dynamics simulation and binding free energy through MM-PBSA

It is a matter of worth exploring the physical basis of the small molecule inside the protein and the dynamic evolution of the protein-ligand complex through MD simulation study. Each of the final proposed inhibitors bound with HIV-1 integrase was considered for MD simulation study for a time of span of 100ns. The MD simulation was carried out in the Gromacs 2018-2 available at the Lengau CHPC server with the time step, constant pressure and constant temperature value of 2 fs, 1 atm and 300 K respectively. The ligand topology was generated through the online freely available SwissParam tool (66). In the case of protein topology, the all-atom CHARMM36 force field (67) was used. The simulation system was built confining the protein-ligand complex within a cubic box with a diameter of 1Å from the center of the system. The entire simulation was performed in the water solvent and the TIP3P water model was used to solvate the system. A sufficient number of required ions, either $Na^+$ or $Cl^-$ were added to neutralize the system. The system was minimized using the Steepest descent algorithm(68) to overcome the close-contacts or overlaps between the atoms. In order to equal distribution of the solvent and ions around the protein-ligand ligand complex, the entire system was equilibrated with NVT (constant number of particles, volume, and temperature) followed by NPT (constant number of particles, pressure, and temperature) ensemble approaches. To address the long-range interaction parameter the van der Waals and electrostatic the cut off were used to 0.9 and 1.4 nm respectively. The trajectory information was updated in 1ps interval. Different parameters including RMSD, root-mean-square fluctuation (RMSF) and radius of gyration (Rg) were calculated from the entire trajectory of MD simulation to explore the behavior of the molecules in the dynamic states. Further, the binding affinity of the molecules was analyzed by the evaluation of binding free energy obtained through the MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) approach (69). The methodology of the MM-PBSA-based binding free energy calculation can be found somewhere else (70).

## 3. Results and discussion

### 3.1 *De novo* design and virtual screening

In order to design novel chemical entities against the HIV-1 integrase infection, the *de novo* designed based virtual screening study is performed. The flow diagram of the study is given in Figure 1. The best-docked complex between the HIV-1 integrase and BI-224436 at the allosteric site is used to design novel and potential HIV-1 integrase inhibitors through the *de*

*novo* design approach. On successful completion of the *de novo* design in the LigBuilder program [60], a total of 80000 molecules are generated. The valance or representation errors are initially checked in a two-dimensional representation. It is found that very few molecules are generated with any such above errors. It is also found that most of the molecules are structurally diverse in nature. The entire set of the structure has been considered to check the duplicates and redundant molecules. On the deletion of duplicates, a total of 63112 molecules remained as structurally unique. The above compounds are allowed to check LoF and Veber's rule, and it is found that 19927 molecules failed to pass through the above rules and subsequently removed. The remaining 43185 molecules are taken forward to dock with HIV-1 integrase using the ADV tool [57].

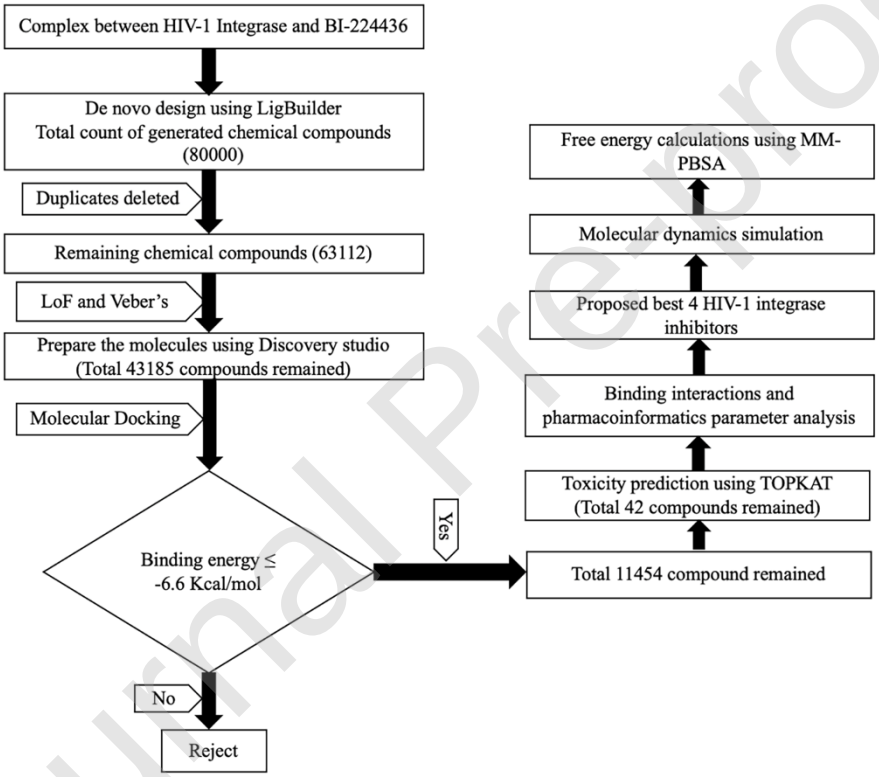



**Figure 1.** Flow diagram of *de novo* design based virtual screening

Additionally, the standard compound BI-224436 has been docked in similar fashion which reveals the binding affinity score to be -6.60 Kcal/mol and subsequently considers it as the threshold value for further molecular screening. The binding energy is explored and 11454 molecules are found to have binding energy less than the threshold (-6.60 Kcal/mol). A number of parameters of the above molecules have been generated through the TOPKAT to analyze the toxic and environmental effects. After the removal of compounds having toxic,

mutagenic, carcinogenic and irritant properties only 42 molecules are retained. Finally, the binding interactions and pharmacoinformatics parameters of each of the 42 molecules have been explored carefully. Particularly, accounting critical analysis, several important parameters such as H-bond energies, glide ligand efficiency scores, ligand binding free energies, the total number of intermolecular interactions present between HIV-1 integrase and designed compounds are recorded for selecting the potential inhibitors compounds. Not only that, different bio- and physicochemical parameters are also obtained in terms of molecular weight, number of heavy atoms, number of aromatic heavy atoms, number of rotatable bonds, molar refractivity, topological polar surface area, solubility class/scores, etc. for all compounds and analyzed extensively for separating the compounds with more suitable drug-like characteristics. Finally, based on the above-mentioned parameters and acceptable pharmacokinetics profile data, four molecules are proposed as promising HIV-1 integrase inhibitors. To adjudge the uniqueness of the selected molecules, publicly available databases such as PubChem and ChemSpider chemical databases are searched through 'Exact Match' or 'Identical Structure' search method. Not a single identical molecule is found from the above databases which substantiate the uniqueness of the molecule. Two-dimensional representation of the final *de novo* designed promising HIV-1 integrase inhibitors is given in Figure 2.

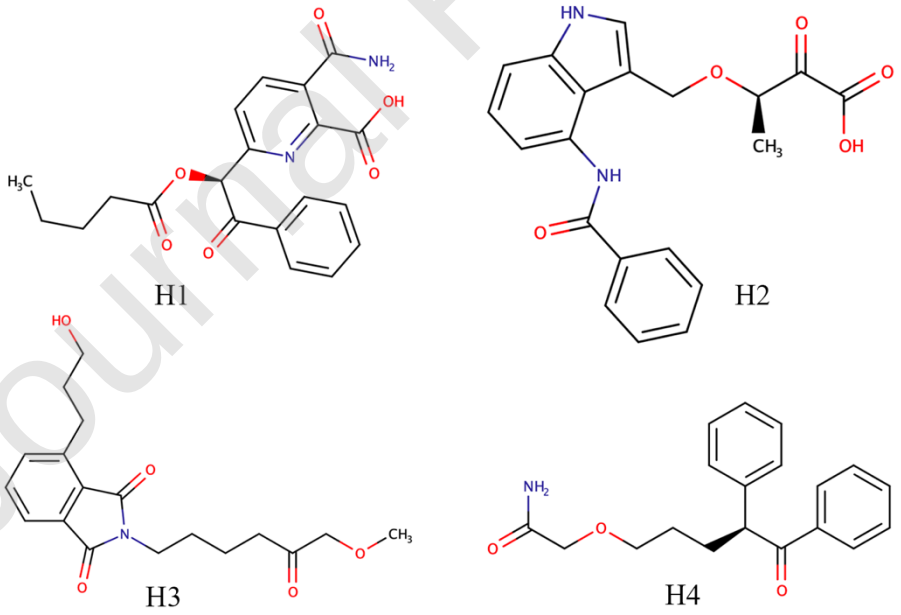



**Figure 2.** Final selected *de novo* designed HIV-1 integrase inhibitors

### 3.2 Molecular interaction analyses through docking study

Firstly, to investigate the docking interaction analysis for the standard compound BI-224436, the interaction profile of original co-crystallized ligand has been extracted in NGL Viewer [71] (Figure 3A), which is directly linked with the PDB. In addition to the above re-docking protocol has also been critically executed. The binding interaction mode of the re-docked ligand is compared with co-crystallized ligand's interaction mode by superimposing the protein-ligand complex crystal structure to each other (Figure 3B). In docking analyses, it is quite interesting to observe that the re-docking protocol successfully generates the exact or alike molecular interaction mode as co-crystallized ligand hold tightly (Figure 3C). For instance, in re-docking, the standard compound shows the similar hydrogen bonding interaction pattern (Figure 3C) which perfectly adjudges with the HIV-1 integrase residues (Glu170, His171 and Thr174) involved in H-bond interactions within the v-shaped pocket of HIV-1 integrase that is experimentally explained [53, 72]. Moreover, the RMSD value of superimposed re-docked ligand has been noted as 0.27Å, which undoubtedly acknowledges the accurate validation of the executed docking protocol employed in the present study. The re-docked superimposed position is displayed in Figure 3D.

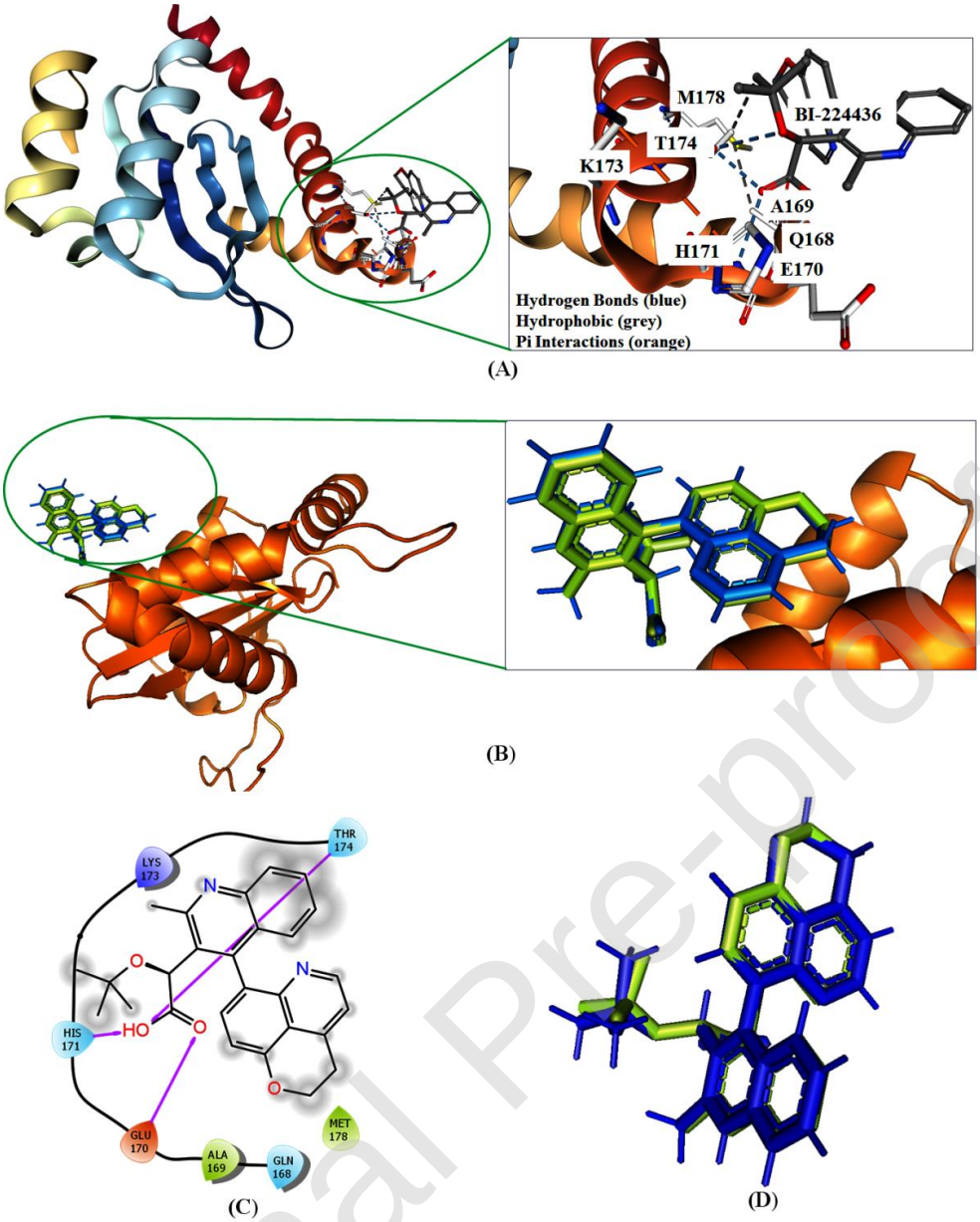



**Figure 3.** Docking interaction analysis of the standard compound BI-224436. **A:** PDB extracted molecular interaction profile for original co-crystallized standard compound BI-224436 and HIV-1 integrase protein. **B:** Illustration of the redocking ligand pose (light blue) and the original ligand pose in the co-crystallized structure (green). The HIV-1 integrase is shown in cartoon representation (light orange). **C:** 2D interaction map obtained in docking for the standard compound (BI-224436) with HIV-1 integrase protein. **D:** Superimposition of the redocking ligand pose (deep blue) and the original ligand pose in the co-crystallized structure (green). Superimposed RMSD value is 0.27.

Docking results of all four designed ligands with HIV-1 integrase protein in 2D interactions map are displayed in Figure 4. Moreover, to understand the better orientation of the docked molecule in the receptor cavity the binding mode in 3D framework is given in Figure 5. The docked complex of the H1 and HIV-1 integrase protein has been established through two H-

bond and one π-π stacking interactions. Amino acid residues Glu87 and Asn184 are involved in H-bond formation and Phe181 formed π-π stacking interaction with H1. The HIV-1 integrase and H1 docking complex exhibits the binding affinity score of -6.90 Kcal/mol. It has been revealed that H2 formed two H-bond interactions with residues Ala86 and Gln177 of HIV-1 integrase protein. Apart from the above involvements other two amino acid residues (Tyr83 and Phe181) are also found which form pi-pi (π-π) stacking interactions with HIV-1 integrase protein. The H-bond interactions with integrase protein are mediated by the association with –OH group of the designed ligand H2; whereas the aromatic ring present in H2 attributes the formation of π-π stacking interactions. In docking analyses, the binding affinity score between HIV-1 integrase and H2 is found to be -7.00 Kcal/mol. The molecular interaction profile between H3 and HIV-1 integrase has been investigated which reveals two H-bond interactions with residues Val113 and Gln138. The terminal oxygen atom and –OH group present in H3 participate to form such interaction profiles with HIV-1 integrase protein. The docking based binding affinity value is found to be -6.60 and -6.60 Kcal/mol for H3 and H4 respectively. The carboxylated oxygen atom of H4 forms H-bond interaction with two consecutive amino acid residues (i.e. Gln137 and Glu138). Further, the list of amino acids interacting with *de novo* designed molecules and BI-224436 are given in Table S1 (Supplementary data). Collectively, the analyses of the docking interaction map provide a possible knowledge of molecular interaction mechanism of designed ligands which may exhibit for inhibiting the HIV-1 integrase protein with better efficacy in comparison to the existing compound. Further, these above-mentioned docking complexes are subjected to all-atom MD simulations analyses in order to understand the dynamic nature of designed compounds as HIV-1 integrase protein inhibitors.

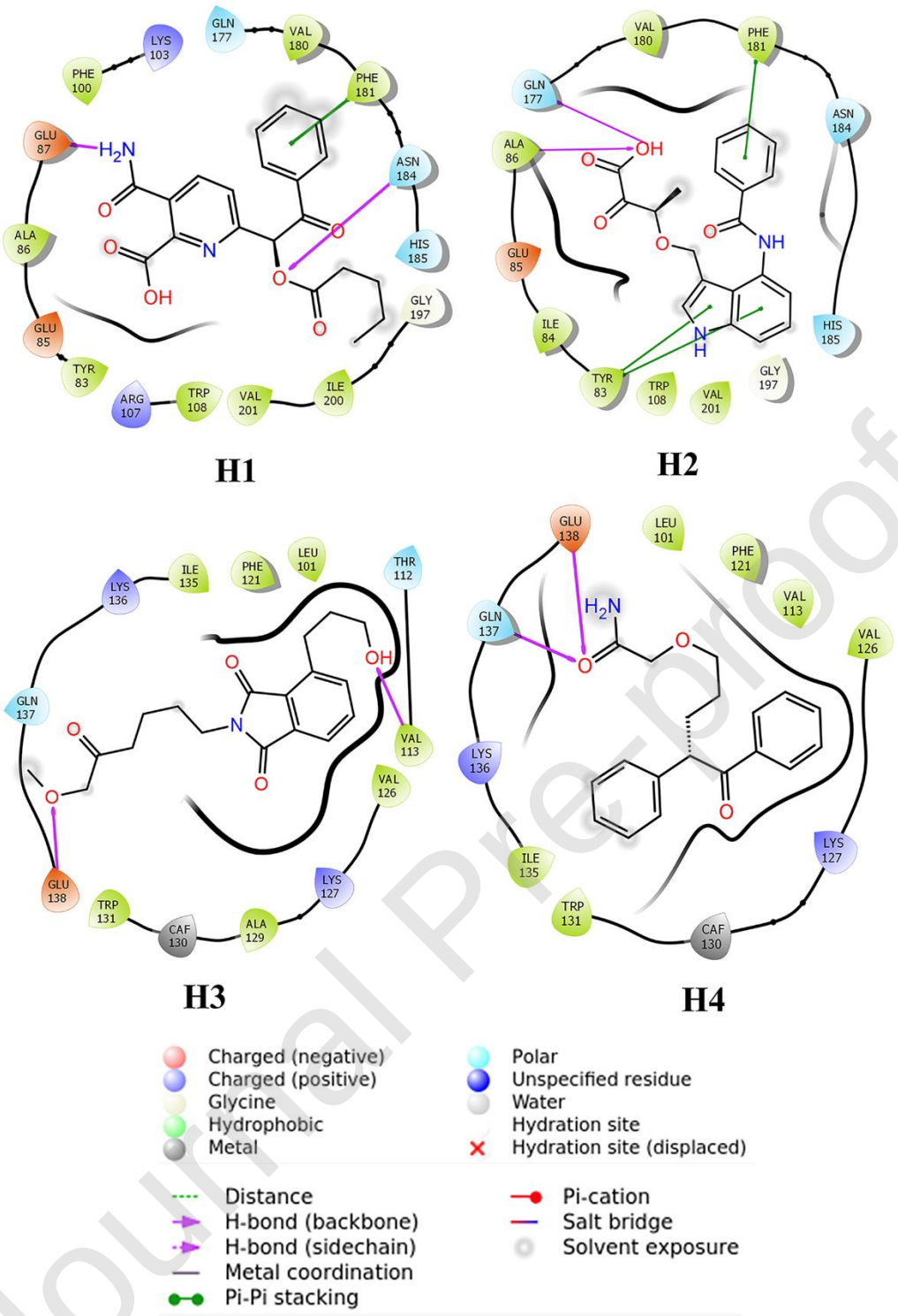



**Figure 4.** Binding interactions between proposed inhibitors and HIV-1 integrase

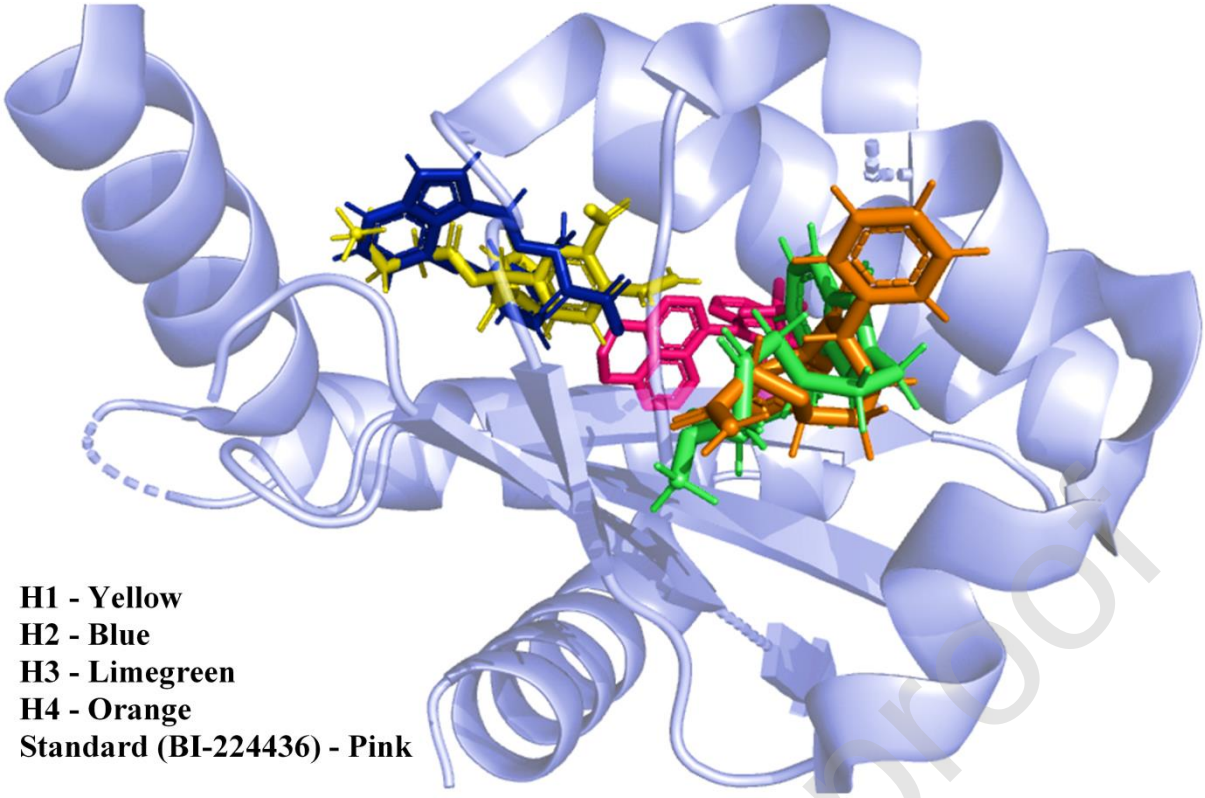



**Figure 5.** Binding orientation and location of identified all 4 proposed compounds and BI-224436 obtained in docking study

### 3.3 Comparative study of binding interactions with existing literature

It is obvious that due to the diversity in the structural framework of all four newly designed compounds (H1, H2, H3 and H4 in Figure 2) demonstrate a somewhat different molecular interaction profile with HIV-1 integrase protein. For instance, the present study finding reveals some novel molecular interaction profiles for the *de novo* designed compounds. However, few observations also noted comparable interaction patterns and residues involvement with four compounds. An earlier study has explained the importance of residue Glu87 in interactions with a small molecule such as tetra-acetylated inhibitor (2E)-3-[3,4-bis(acetoxy)phenyl]-2-propenoate-N-[(2E)-3-[3,4-bis(acetyloxy)phenyl]-1-oxo-2-propenyl]-l-serine methyl ester [73]. The same amino acid (Glu87) seems to be engaged as a central linker to form an integrase dimer [73]. In the present study, H1 and H2 reveal the strong interaction affinity with the residue Glu87 through the formation of H-bond interaction and hydrophobic contact. Additionally, in a computational study, a few HIV-1 integrase inhibitors were studied through 3D QSAR, docking and molecular dynamics simulation which explained the importance of Gln177 residue as a favorable hydrogen bond donor that is present in the CCD of HIV-1 integrase protein [74]. It is also found that the same residue

involved in exhibiting H-bond interaction with H2. Among a number of conserved hydrophobic amino acid residues such as Trp132, Met178, Phe181, Phe185 are present in the CCD or core-core interface of HIV-1 integrase protein usually trusted to participate in the 4-tiered π interaction for multimerization [75, 76]. The compound H2 is found to form H-bond interaction with Phe181 which may contribute to stabilizing the complex as well as play a critical role in DNA integration and viral infectivity through 4-tiered π interaction. In an earlier study, it has been reported the importance of the aromatic amino acid residue (Tyr83) for the involvement in π–π interaction [74], and this study also highlights its importance in terms of creation of similar (pi-pi stacking) interaction with compound H2. Moreover, this region also proves to be crucial due to the aromatic hydrophobic interactions. Amino residue, Val113 is found to be important to form H-bond interaction with H3 which can be substantiated by an earlier study where replacement of the same amino residue found to decrease the integrase catalytic activity [77]. In the present study, H4 exhibit the participation of residue Glu138 in H-bond interaction. Although, the presence of this amino acid in CCD and its importance are highlighted in different ways in different studies [78]. The residue definitive impact has been investigated to suppress the replication of specific drug-resistant (E138K) RT and impairment of HIV-1 integrase SUMOylation [79, 80]. Additionally, in another study, the docking interaction revealed that residue Glu138 is formed similar H-bond interaction profile [81]. Alltogether, the findings of the molecular docking analyses of four compounds establish significant molecular binding interactions as explained earlier or newly formed interaction profiles with the HIV-1 integrase protein complex which may suggest that the *de novo* compounds provide a better opportunity of developing therapeutic activity for HIV-1 integrase inhibition.

### 3.4 Pharmacokinetics and drug-likeness analysis

To explore the pharmacokinetic behavior of the final proposed HIV-1 integrase inhibitors, a number of pharmacokinetics and physicochemical properties have been calculated through the SwissADME web server [66] and given in Table 1. The molecular weight of all four molecules is found to be in the range of 311 to 443 g/mol. The oral activeness can be explained through the total polar surface area (TPSA) and the value is found to be 136.65, 108.49, 83.91 and 69.39 $Å^2$ for H1, H2, H3 and H4 respectively. Above TPSA value undoubtedly favors the oral activeness of the molecules. Solubility class explains that all the molecules are either soluble or moderately soluble in nature. The human intestinal absorption (GI) parameter clearly indicates that each molecule possesses high absorbable characteristics

in the intestine. The flexibility of the molecule can be explained through its number of rotatable bonds and it is found to be 10, 8, 10 and 9 in H1, H2, H3 and H4 respectively.

Moreover, the HIA and BBB are explored through the BOILED-Egg model retrieved from the SwissAdme server which is given in Figure 6. In this model, the molecule present in the albumin (white) and yolk (yellow) regions explain their intention to penetrate the HIA and BBB respectively. It is also reported that yellow and white regions are not mutually exclusive. From Figure 6, it can be seen that all the molecules except H4 show strong absorption in the HIA. Further, substrates (PGP+) and non-substrates (PGP-) characteristics of each molecule are verified and represented by blue and red circles respectively in Figure 6. As reported that on drives back the molecules into the intestinal lumen in the liver the PGP declines the efficiency of PGP+. In Figure 6, H1 and H4 are found to be in the category of PGP- which indicates that they belong to the non-substrate category. The other two molecules, H2 and H3 are found to be in PGP+ category. From the above observations, it is undoubtedly clear that H1, H2, H3 and H4 consist of characteristics to be the promising HIV-1 integrase inhibitors.

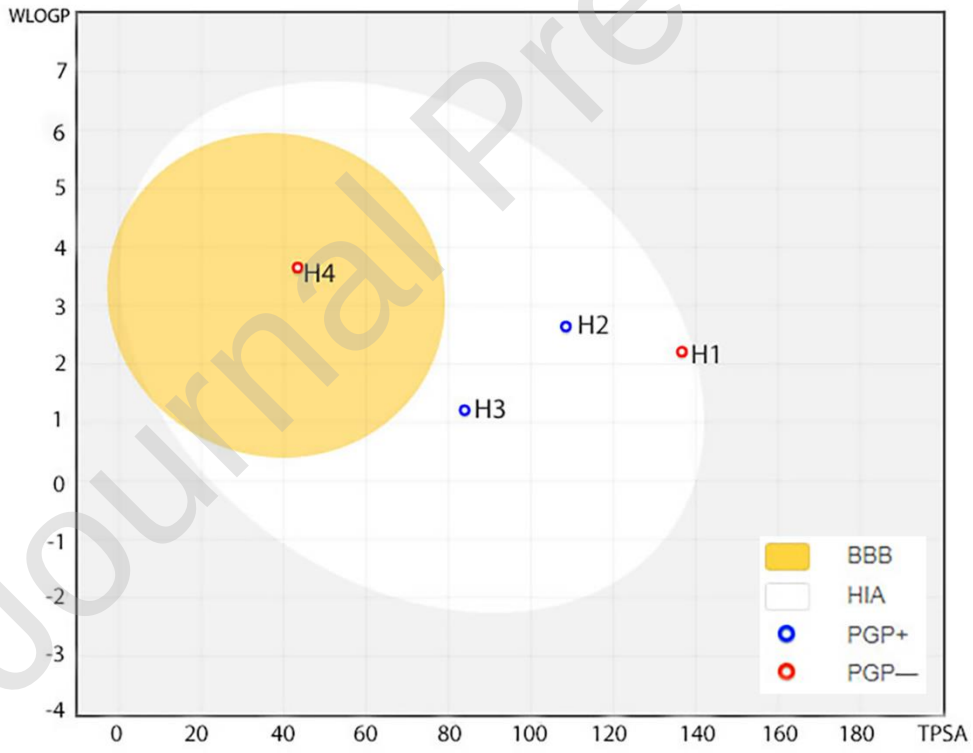



**Figure 6.** BOILED-egg model of final proposed HIV-1 integrase inhibitors

In order to check the drug-likeness of the proposed molecules, a number of parameters such as LE, LE_Scale, FQ and LELP have been calculated which is given in Table 1. The LE is calculated as per equation (1). It is illustrated that the $LE \leq 0.4$ indicates the lead-like nature of the molecule. Not a single proposed molecule is found to be possessing LE more than 0.290 which clearly explains that all the molecules consist of drug-likeness characteristics.

Table 1. Bioactivity and efficiency parameters of proposed HIV-Integrase inhibitors

| Molecule | [1]BE | [2]LE | [3]LE_Scale | [4]FQ | [5]LELP |
|---|---|---|---|---|---|
| **H1** | -7.0 | 0.250 | 0.358 | 0.698 | 4.840 |
| **H2** | -6.9 | 0.255 | 0.369 | 0.692 | 7.278 |
| **H3** | -6.6 | 0.275 | 0.404 | 0.680 | 10.4 |
| **H4** | -6.6 | 0.289 | 0.416 | 0.696 | 7.804 |
| **BI-224436** | -6.6 | 0.200 | 0.307 | 0.651 | 15.95 |

[1]Binding energy; [2]Ligand efficiency; [3]Ligand efficiency scale; [4]Fit quality; [5]Ligand-efficiency-dependant-lipophilicity

The LE_Scale of all four molecules has been calculated through equation (2) and found to be in the range of 0.307 to 0.416. Low LE_Scale value undoubtedly explains that H1, H2, H3 and H4 are potential in nature.

Being a potential molecule, the FQ (calculated using equation (3)) value should be close to 1. Table 1 indicates that the FQ value of H1, H2, H3 and H4 is found to be 0.698, 0.692, 0.980 and 0.696 respectively. The above data clearly indicates that all the molecules have a strong binding capability to the HIV-1 integrase.

Finally, the LELP has been calculated by using equation (4). The recommended value of the LELP is more than 3. The LELP value of H1, H2, H3 and H4 is found to be 4.840, 7.278, 1.0400 and 7.804 respectively. Above LELP value undoubtedly explains that all the proposed molecules possess drug-like properties.

### 3.5 Molecular dynamics simulation analyses

MD simulation strategy of the protein-ligand complex is a crucial and powerful approach to check the dynamic behavior and the stability of the molecules. For this purpose, the proposed molecules and BI-224436 complex with HIV-1 integrase are considered for all-atoms 100ns

MD simulation study. The entire trajectory of each complex is used to calculate the RMSD, RMSF and Rg parameters.

The RMSD of each frame corresponding to the simulation time can give a clear idea about the dynamic stability and the protein backbone conformational perturbations. It has been noticed that all the systems achieve the equilibrium state at 100ns of MD simulations. Backbone RMSD of each complex is calculated which is given in Figure 7. Average, maximum and minimum RMSD are also calculated that is given in Table 2. Manual inspection of the plot (Figure 7) clearly indicates that the backbone of HIV-1 integrase bound with proposed inhibitors and BI-224436 shows consistency during the simulation. There is no significant fluctuation observed in any protein-ligand complex except H1-HIV-1 integrase complex. Frames of each complex are stabilized within the 0.532 nm. The HIV-1 integrase backbone bound with H1 initially deviates within the range of 0.3 nm and finally, it achieves consistency around RMSD value of 0.4 nm. The backbone of HIV-1 integrase bound with remaining molecules is found steady throughout the simulation. The difference between the maximum and average RMSD is found to be 0.161, 0.105, 0.090, 0.145 and 0.087 nm bound with H1, H2, H3, H4 and BI-224436 respectively. The above low value clearly explains the less deviation of the backbone in comparison to the crystal structure. The steadiness of the RMSD throughout the MD simulation is clearly explained that all the proposed HIV-1 integrase inhibitors hold the protein backbone through the binding interactions which lead to restricting the higher deviation.

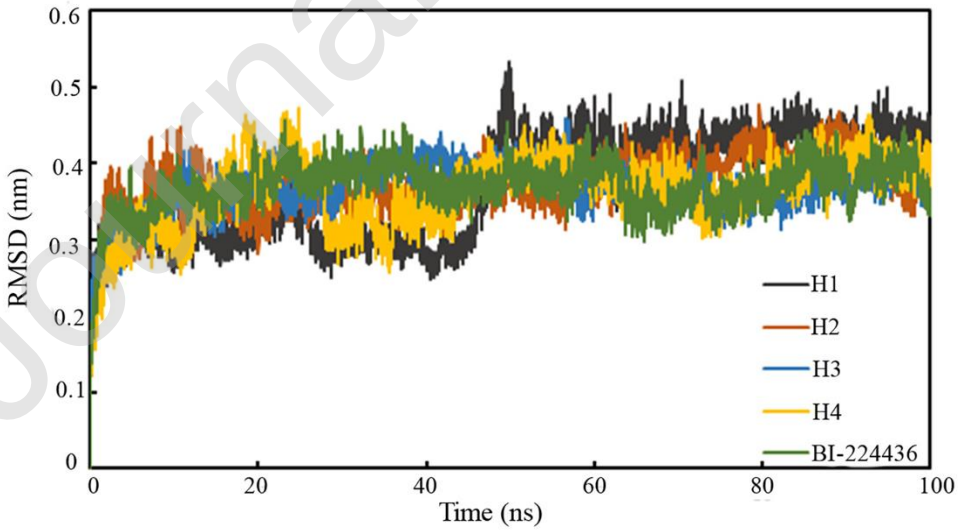



**Figure 7.** Backbone RMSD of HIV-1 integrase bound with H1, H2, H3, H4 and BI-224436

In the protein-ligand complex stability, the individual amino acid plays a critical role. The fluctuation of the singular amino acid of HIV-1 integrase has been calculated in terms of RMSF and depicted in Figure 8. The RMSF plot clearly indicates that all the amino acids of HIV-1 integrase bound with proposed molecules fluctuate in a similar pattern with few exceptions. It can be seen that amino residues approximately range from 135 to 155 are fluctuated with higher value in comparison to others. The distant fluctuation of amino acids in the above range might be due to their role in the binding of ligand differently. Maximum and minimum RMSF value is found to be 0.634 and 0.067; 0.538 and 0.066; 0.497 and 0.060; 0.674 and 0.076; and, 0.732 and 0.067 nm for the HIV-1 integrase bound with H1, H2, H3, H4 and BI-224436 respectively. Above low RMSF values clearly indicate that there is no significant fluctuation of HIV-1 integrase amino residues observed on bound with the proposed molecules.

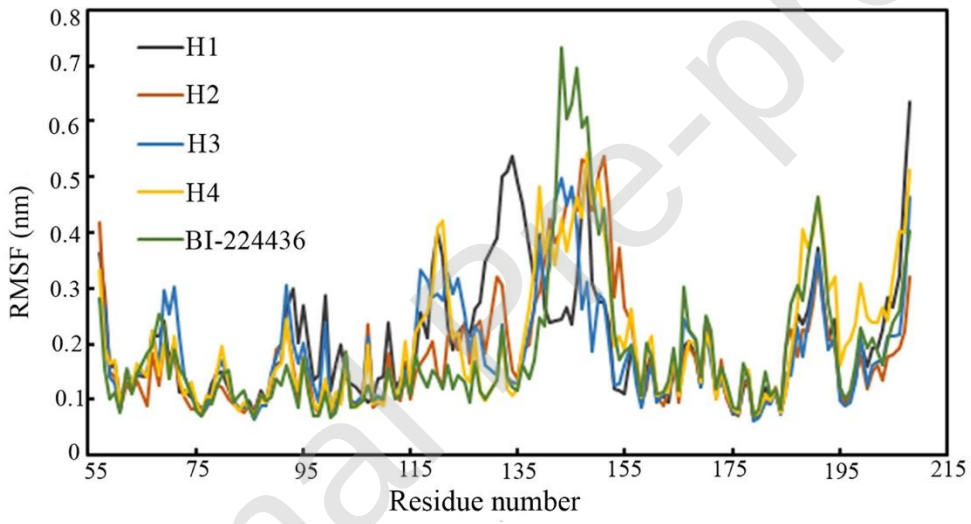



**Figure 8**. RMSF of HIV-integrase amino residues bound with H1, H2, H3, H4 and BI-224436

Moreover, per residue contribution in binding energy is an important and essential approach to check the important amino residues for the ligand binding. It has been observed that all the interacting amino residues with proposed HIV-integrase inhibitors reveal high binding energy with few exceptions. The average contributing binding energy of each amino residue has been calculated and plotted in Figure S2 (Supplementary data). In the case of HIV-1 integrase bound with H1 and H2, the binding energy of Phe181 has found to be negative. Another catalytic amino acid, Glu138 is found with negative binding energy when bound with H3 and H4. The catalytic amino residues Glu87 and Asn184 in HIV-1 integrase-H1; Tyr83, Ala86

and Gln177 in HIV-1 integrase-H2; Val113 in IN-H3; and, Gln137 in IN-H4 are found to have negative binding energy. Hence above amino residues are undoubtedly major interacting residues with the proposed molecules. Moreover, other amino residues are found vicinity to the proposed molecules also showed negative binding energy with few exceptions. In the case of RMSF analysis, the amino residues 115 to 155 of IN fluctuate differently when bound with proposed IN inhibitors. Per residue contribution analysis is clearly shows that the amino acids of the above range having very low range binding energy that might be the reason for fluctuation differently.

Table 2. RMSD and RMSF of HIV-1 integrase bound with proposed inhibitors

| Complex | | | H1 | H2 | H3 | H4 | BI-224436 |
|---|---|---|---|---|---|---|---|
| Protein-ligand complex | RMSD (nm) | Minimum | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Maximum | 0.532 | 0.476 | 0.459 | 0.506 | 0.4567 |
| | | Average | 0.371 | 0.371 | 0.369 | 0.361 | 0.370 |
| | RMSF (nm) | Minimum | 0.067 | 0.066 | 0.060 | 0.076 | 0.067 |
| | | Maximum | 0.634 | 0.538 | 0.497 | 0.674 | 0.732 |
| | | Average | 0.203 | 0.184 | 0.184 | 0.215 | 0.188 |

The rigidness of the system can be addressed through the inspection of the radius of gyration value. For each protein-ligand system, the Rg value has been calculated which is given in Figure 9. Except for the HIV-1 integrase bound with H1 all systems are found steady during the MD simulation. Not a single system is found with a significant deviation in comparison to the crystal structure. HIV-1 integrase bound with H1 is found to show a little bit deviation around 50ns of MD simulation. For deep inspection about the steadiness of each system, the difference of maximum and minimum with average Rg value checked and found to be 10.106 and 0.170 nm; 0.086 and 0.139 nm; 0.051 and 0.096 nm; 0.053 and 0.107 nm; and, 0.054 and 0.102 ns for H1, H2, H3, H4 and BI-224436, respectively. All the above values are found to be less 0.6 nm which undoubtedly explained that no observable deviation of the system found throughout the simulation. Therefore, a detailed analysis of RMSD, RMSF and Rg parameters above has indisputably indicated that the proposed molecule retained inside the receptor cavity of HIV-1 integrase in the dynamic states.

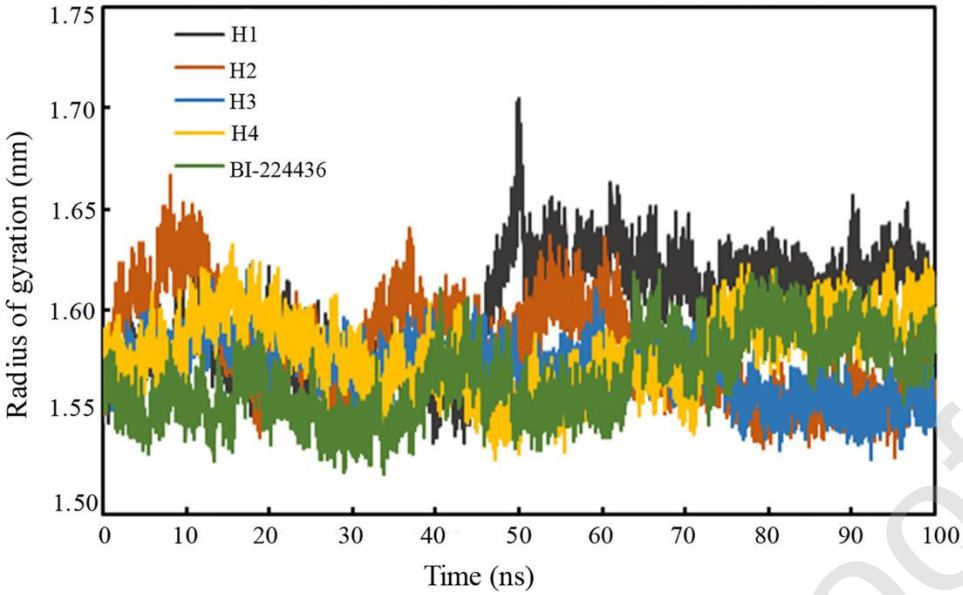



**Figure 9.** Radius of gyration of HIV-1 integrase bound with H1, H2, H3, H4 and BI-224436

### 3.6 Post MD simulations based conformation and interactions analyses

The binding interactions obtained in molecular docking are not always preferred because some molecular interactions found in the initial docking complex might lose its interaction integrity during the MD simulation run, or sometimes new molecular interactions also might appear as well. Therefore, instead of depending on molecular docking data, MD ensemble data are much richer and reliable, and it is proven to be a more accurate and effective approach to explore molecular conformations analysis. In this study, a longitudinal analysis of protein-ligand interactions based on the MD simulations data have been critically explored to understand the sustainability of the key interactions found in the MD simulation. The last frame of each protein-ligand complex after all-atoms 100ns MD simulation extracted and binding interaction profile is given in Figure 10. The molecular interactions profile analysis reveals that at the end of the simulation, H1 consists of two newer H-bond interactions with amino acid residues Tyr83 and Ala86. Few common amino acids Trp108, Phe181, and Val201 remain to be firmly interacted via the formation of hydrophobic contacts with H1. Interestingly, a new type of molecular interaction *viz.* salt bridge interaction has appeared during MD simulation for H1 through the participation of residues Lys103 and Arg107.

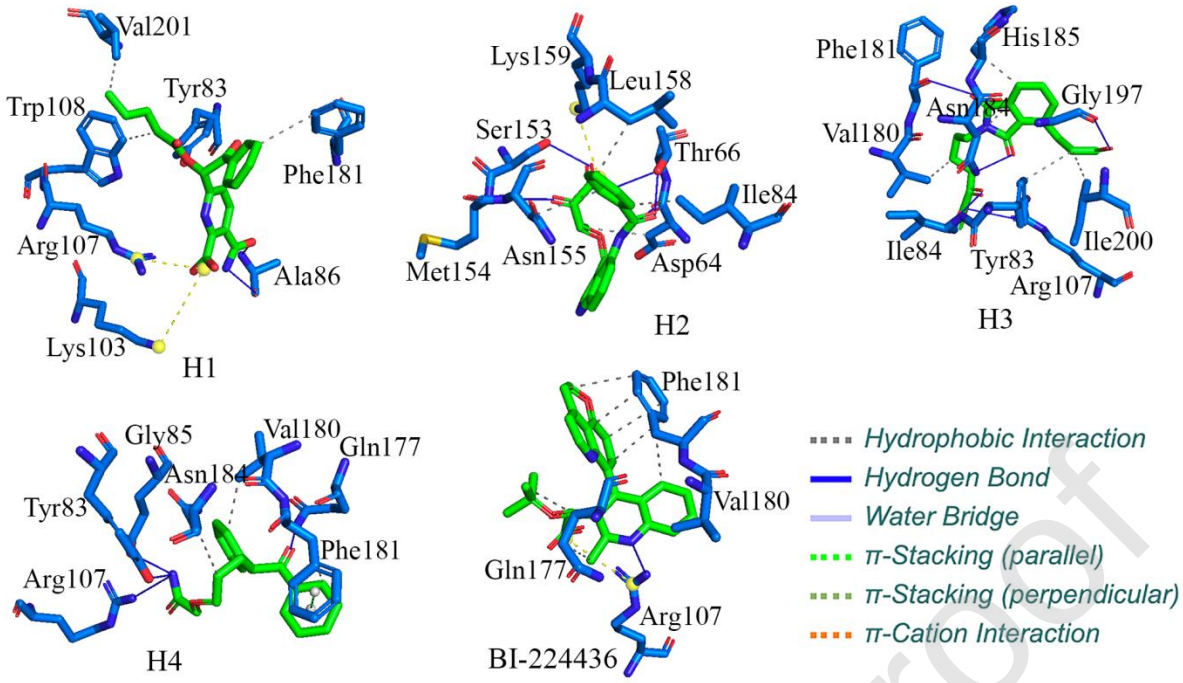



**Figure 10.** Post-MD simulated interaction analyses for all proposed inhibitors including the standard compound BI-224436

The MD simulation interactions profile of H2 shows very surprising and crucial observations. Most of the interactions found are new in comparison to molecular docking findings. In particular, the amino acid residues such as Asp54, Thr66, Ser153, Met154, and Asn155 are observed to form H-bond interactions during MD simulation with the help of induce fit energy effects that might help to create such H-bond interactions with H2. Additionally, a few other residues such as Asp64, Ile84, Asn155, and Leu158 establish the hydrophobic contacts with H2. A salt bridge formation is also observed with basic amino acid residue, Lys159 with H2 during MD simulation. The ligand binding interaction in the dynamic state reveals that H3 also forms newer types of interactions profile in respect to that found in docking analysis. The amino residues Ile84, Arg107, Phe181, Asn184 and Gly197 form H-bond interactions, and Tyr83, Val180, Asn184, His185 and Ile200 found to participate in hydrophobic contacts. It is important to note that the new binding interactions observed in MD simulation are not far away from the initial binding pose / active site pocket. It can be said that during MD simulation, flexibility exerted to the ligand as well as protein might develop such a conformational state to exhibit new conformation/state of the protein-ligand complex for H3-integrase protein. At the end of the simulation, the molecular binding interactions of H4 with integrase protein have been checked. A slight shift is also found in

binding/interaction orientation in comparison to docking orientation. Precisely, residues Tyr83, Glu85, Arg107, and Gln177 form H-bond interactions with H4. Beside above residues involvement in interactions, hydrophobic contacts also form with hydrophobic residues *viz.* Val180, Phe181, and Asn184 during MD simulations. In addition, one pi-stacking interaction is observed between residue Phe181 and H4 based on MD simulation interaction analysis. The binding orientation of standard compound BI-224436 is checked at the end of the MD simulation which reveals that residue Arg107 form H-bond interaction and salt bridge with compound BI-224436, and residues Gln177, Val180, and Phe181 are attributed to participate in hydrophobic contacts. Although, there are some new findings in terms of participation of amino acids involvement based on MD simulations interaction analyses. However, few common amino acid residues are also found to be anchored with the identified compound as observed in molecular docking analysis. The above mentioned observations for all protein-ligand complexes suggest that the difference in the structural features of identified inhibitors can lead to govern different internal dynamics in HIV-1 integrase protein, specifically at the CCD region. Overall, all the aforementioned molecular interactions obtained in MD simulation study indicate that binding of proposed inhibitors generates an obvious impact on the conformational changes in CCD region of HIV-1 integrase protein which is induced to bring forth new intermolecular interactions during MD simulation. Furthermore, such observations suggest that newly *de novo* designed inhibitors might act as allosteric inhibitors because identified compounds (H1-H4) are essentially found to interact with residues at the allosteric site or surrounding residues present in the allosteric site of HIV-1 integrase protein. As the reference standard compound BI-224436 binds to a large allosteric region extending from amino acid residues 168 to 178 for inhibiting 3'-end processing activity. Hence, it is highly expected that identified compounds also interact with amino residues at the allosteric region or its close proximity. Post-MD simulation complexes are found to exert the binding interactions with critical amino residues present in the allosteric site HIV-1 integrase.

### 3.7 Binding energy through MM-PBSA approach

The binding affinity of the proposed molecules and BI-224436 towards the HIV-1 integrase has been assessed through the calculation of binding free energy using the MM-PBSA approach. For this purpose, the entire trajectory of each system is considered. It is illustrated that binding free energy obtained through MM-PBSA approach is more authentic in comparison to the binding energy derived from any other static frame such as molecular docking. Binding free energy is plotted against the time span of the MD simulation which is

given in Figure 11. Moreover, the average, maximum and minimum binding free energy are also extracted which is given in Table 3. Lower the binding free energy of the molecule explains stronger attraction towards the HIV-1 integrase. On close inspection, it can be seen that H1 shows the highest affinity with the average binding energy value of -175.122 Kcal/mol followed by H2 with -62.979 Kcal/mol, H3 with -55.883 Kcal/mol, H4 with -38.758 Kcal/mol, and, BI-224436 with -26.927 Kcal/mol. It is worth to be noted that BI-224436 shows the lowest affinity in comparison to the proposed HIV-1 integrase inhibitors. From Figure 11, it can be seen that except H2, all molecules show consistency in binding energy for each frame during the simulation. Although estimation of the binding energy of H2 quite fluctuates for the different frame but not a single frame is found to have binding free energy 0 or higher than 0 Kcal/mol. Moreover, it has been observed that the Coulomb or electrostatic interaction and van der Waals interaction energy mainly contribute to accomplish higher binding free energy. Therefore, all proposed HIV-1 integrase inhibitors are found to show strong affection towards the protein and possess enough strength for the inhibition.

Table 3. Maximum, minimum and average Van der Waals, electrostatic and binding free energies of H1, H2, H3, H4 and BI-224436

| Energy | | H1 | H2 | H3 | H4 | BI-224436 |
|---|---|---|---|---|---|---|
| Van der Waals (Kcal/mol) | Minimum | -38.009 | -38.264 | -49.121 | -36.234 | -39.001 |
| | Maximum | -6.963 | 11.543 | 0.000 | 4.3234 | -2.312 |
| | Average | -25.139 | -14.660 | -27.292 | -12.871 | -9.692 |
| Electrostatic (Kcal/mol) | Minimum | -242.057 | -182.664 | -66.733 | -69.509 | -54.327 |
| | Maximum | -52.455 | 21.556 | 8.409 | 8.342 | 10.347 |
| | Average | -149.983 | -48.319 | -28.591 | -25.887 | -17.235 |
| Binding free energy (Kcal/mol) | Minimum | -264.116 | -191.441 | -103.095 | -85.171 | -42.744 |
| | Maximum | -78.678 | -2.162 | -17.644 | -6.696 | -5.749 |
| | Average | -175.122 | -62.979 | -55.883 | -38.758 | -26.927 |

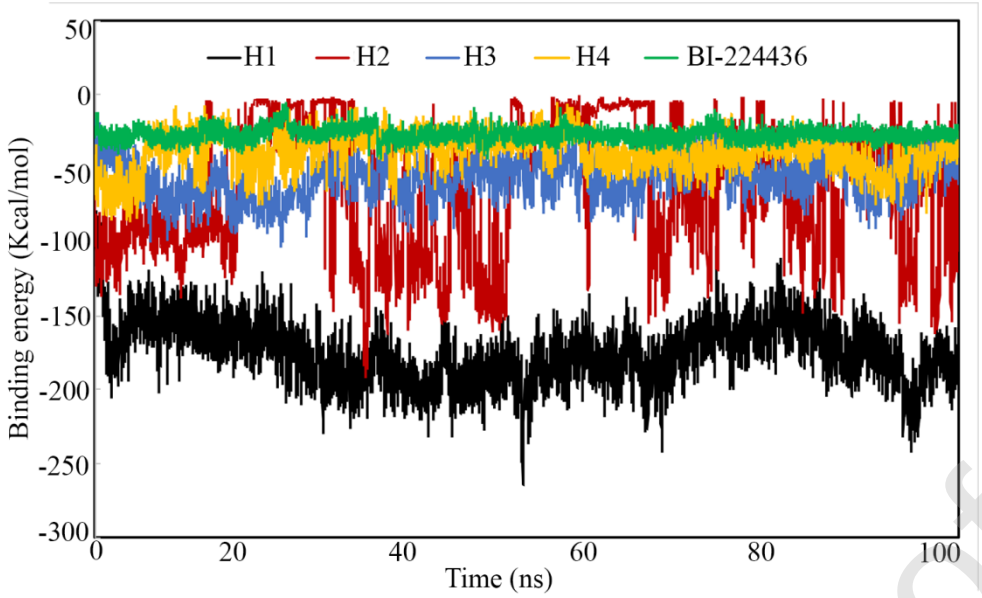



**Figure 11.** Binding free energy of H1, H2, H3, H4 and BI-224436 bound with HIV-1 integrase

## 4. Future prospects

Application of the pharmacoinformatics tools and algorithms in the drug discovery research brought in new benchmarks to identify promising chemical compounds for a specific target. There is no denial that pharmacoinformatics approaches are the boost to enhance the drug discovery pipeline. Proposed four molecules show strong binding affection towards the HIV-1 integrase through molecular docking, *in-silico* pharmacokinetics and MD simulation studies. Further, to check the potentiality of proposed molecules against HIV-1 integrase a number of experimental validations are needed. The thermal shift assay can be used to verify the binding affinity of the molecules. The kinetic study is the perfect approach to check the binding and unbinding mechanism of the molecules. The molecules may need further optimization on the basis of experimental assessment to improve the efficacy and therapeutic potentiality.

## 5. Conclusion

The structure-based designing such as *de novo* approach in LigBuilder is used to identify novel and potential chemical components for successful inhibition of HIV-1 integrase. For this purpose, a known inhibitor bound in the x-ray crystal structure of the HIV-1 integrase has been considered. A total of 80000 new HIV-1 integrase inhibitors are generated through the

*de novo* strategy with the help of LigBuilder, installed in a Linux-based server. Followed by the removal of duplicate molecules, a number of criteria are imposed to narrow down the chemical space. Initially, after verifying the LoF and Veber's rule all molecules are docked and the binding energy is compared with BI-225536. Molecules having better binding energy than BI-225536 are considered for further assessment. The toxic, mutagenic, irritant and carcinogenic molecules are identified through the TOPKAT and removed. Further, the drug-likeness and pharmacokinetics characteristics are assessed. Finally, the binding interactions of remaining molecules explored and four molecules are found to be promising HIV-1 integrase inhibitors. A number of catalytic amino residues are found to form strong binding interactions with the proposed molecules. Finally, the characteristics in the dynamic environment of each molecule are explored through a 100ns MD simulation study. The number of parameters included RMSD, RMSF and Rg which are calculated from the entire trajectory on successful completion of the MD simulation. All the above parameters undoubtedly explain that proposed molecules are capable enough to retain inside the receptor cavity of HIV-I integrase. Finally, the binding free energy of each molecule including BI-224436 is calculated using the MM-PBSA approach. Interestingly it is found that all proposed molecules show higher binding affection towards the HIV-1 integrase in comparison to the BI-224436. Hence, it can be concluded that newly designed and proposed molecules might be important to inhibit the HIV-1 integrase subjected to experimental validation.

**CRediT authorship contribution statement**

**Pooja Balasaheb Shinde:** Data curation, Formal analysis, Investigation, Methodology, **Shovonlal Bhowmick:** Data curation, Formal analysis, Investigation, Methodology, Writing - original draft, Writing – review & editing. **Etidal Alfantoukh:** Conceptualization, Writing – review & editing. **Preeti Chunarkar Patil:** Conceptualization, Writing – review & editing. **Saikh Mohammad Wabaidur**: Conceptualization, Writing – review & editing. **Rupesh V Chikhale:** Conceptualization, Writing – review & editing. **Md Ataul Islam:** Conceptualization, Investigation, Super-vision, Data curation, Writing - original draft, Writing – review & editing.

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Computational resource**

The CHPC (www.chpc.ac.za) is thankfully acknowledged for computational resources and tools.

**Conflict of interest**

Authors declare that there is no competing interest

**Acknowledgment**

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia through the Fast-track Research Funding Program.

**References**

[1] H.F. Gunthard, M.S. Saag, C.A. Benson, C. del Rio, J.J. Eron, J.E. Gallant, J.F. Hoy, M.J. Mugavero, P.E. Sax, M.A. Thompson, R.T. Gandhi, R.J. Landovitz, D.M. Smith, D.M. Jacobsen, P.A. Volberding, Jama, 316 (2016) 191-210.
[2]
[3] M.M. Lederman, P.M. Cannon, J.S. Currier, C.H. June, H.P. Kiem, D.R. Kuritzkes, S.R. Lewin, D.M. Margolis, J.M. McCune, J.W. Mellors, T.W. Schacker, R.P. Sekaly, P. Tebas, B.D. Walker, D.C. Douek, Pathogens & immunity, 1 (2016) 154-164.
[4] H.F. Zulfiqar, A. Javed, Sumbal, B. Afroze, Q. Ali, K. Akbar, T. Nadeem, M.A. Rana, Z.A. Nazar, I.A. Nasir, T. Husnain, Frontiers in public health, 5 (2017) 32.
[5] O. Delelis, K. Carayon, A. Saib, E. Deprez, J.F. Mouscadet, Retrovirology, 5 (2008) 114.
[6] D.J. Lee, W.E. Robinson, Jr., Journal of virology, 78 (2004) 5835-5847.
[7] Y. Luo, M.A. Muesing, Future medicinal chemistry, 2 (2010) 1055-1060.
[8] J. Demeulemeester, S. Vets, R. Schrijvers, P. Madlala, M. De Maeyer, J. De Rijck, T. Ndung'u, Z. Debyser, R. Gijsbers, Cell Host & Microbe, 16 (2014) 651-662.
[9] S.J. Smith, X.Z. Zhao, T.R. Burke, Jr., S.H. Hughes, Antimicrob Agents Chemother, 62 (2018).
[10] C. Liao, C. Marchand, T.R. Burke, Jr., Y. Pommier, M.C. Nicklaus, Future medicinal chemistry, 2 (2010) 1107-1122.
[11] T.K. Chiu, D.R. Davies, Current topics in medicinal chemistry, 4 (2004) 965-977.
[12] K. Sayasith, G. Sauve, J. Yelle, Expert opinion on therapeutic targets, 5 (2001) 443-464.
[13] R. Atanassova, G. Spassov, R. Balansky, K. Boev, The Journal of pharmacy and pharmacology, 44 (1992) 663-666.
[14] L.d.A. Machado, M.F.d.C. Gomes, A.C.R. Guimarães, Frontiers in Microbiology, 10 (2019).
[15] P.K. Quashie, Y.S. Han, S. Hassounah, T. Mesplede, M.A. Wainberg, PLoS One, 10 (2015) e0128310.

[16] Y. Pommier, A.A. Johnson, C. Marchand, Nature Reviews Drug Discovery, 4 (2005) 236-248.
[17] A. Savarino, Expert Opinion on Investigational Drugs, 15 (2006) 1507-1522.
[18] O. Delelis, K. Carayon, A. Saïb, E. Deprez, J.-F. Mouscadet, Retrovirology, 5 (2008) 114.
[19] J.C.-H. Chen, J. Krucinski, L.J.W. Miercke, J.S. Finer-Moore, A.H. Tang, A.D. Leavitt, R.M. Stroud, Proceedings of the National Academy of Sciences, 97 (2000) 8233-8238.
[20] D. Patel, J. Antwi, P.C. Koneru, E. Serrao, S. Forli, J.J. Kessl, L. Feng, N. Deng, R.M. Levy, J.R. Fuchs, A.J. Olson, A.N. Engelman, J.D. Bauman, M. Kvaratskhelia, E. Arnold, The Journal of biological chemistry, 291 (2016) 23569-23577.
[21] A. Engelman, J.J. Kessl, M. Kvaratskhelia, Current opinion in chemical biology, 17 (2013) 339-345.
[22] K. Shimura, E.N. Kodama, Antiviral chemistry & chemotherapy, 20 (2009) 79-85.
[23] J.L. Blanco, V. Varghese, S.Y. Rhee, J.M. Gatell, R.W. Shafer, The Journal of infectious diseases, 203 (2011) 1204-1214.
[24] B. Akil, G. Blick, D.P. Hagins, M.N. Ramgopal, G.J. Richmond, R.M. Samuel, N. Givens, C. Vavro, I.H. Song, B. Wynne, M. Ait-Khaled, Antiviral therapy, 20 (2015) 343-348.
[25] A. Pendri, N.A. Meanwell, K.M. Peese, M.A. Walker, Expert opinion on therapeutic patents, 21 (2011) 1173-1189.
[26] L. Elzi, S. Erb, H. Furrer, M. Cavassini, A. Calmy, P. Vernazza, H. Gunthard, E. Bernasconi, M. Battegay, H.I.V.C.S.G. Swiss, Aids, 31 (2017) 1853-1858.
[27] J. Peñafiel, E. de Lazzari, M. Padilla, J. Rojas, A. Gonzalez-Cordon, J.L. Blanco, J. Blanch, M.A. Marcos, M. Lonca, M. Martinez-Rebollar, M. Laguno, A. Tricas, A. Rodriguez, J. Mallolas, J.M. Gatell, E. Martinez, Journal of Antimicrobial Chemotherapy, 72 (2017) 1752-1759.
[28] A. Orta-Resendiz, R.A. Rodriguez-Diaz, L.A. Angulo-Medina, M. Hernandez-Flores, L.E. Soto-Ramirez, AIDS Research and Therapy, 17 (2020) 6.
[29] K. Shimura, E. Kodama, Y. Sakagami, Y. Matsuzaki, W. Watanabe, K. Yamataka, Y. Watanabe, Y. Ohata, S. Doi, M. Sato, M. Kano, S. Ikeda, M. Matsuoka, Journal of Virology, 82 (2008) 764-774.
[30] O. Goethals, R. Clayton, M. Van Ginderen, I. Vereycken, E. Wagemans, P. Geluykens, K. Dockx, R. Strijbos, V. Smits, A. Vos, G. Meersseman, D. Jochmans, K. Vermeire, D. Schols, S. Hallenberger, K. Hertogs, Journal of Virology, 82 (2008) 10366-10374.
[31] I. Malet, O. Delelis, M.-A. Valantin, B. Montes, C. Soulie, M. Wirden, L. Tchertanov, G. Peytavin, J. Reynes, J.-F. Mouscadet, C. Katlama, V. Calvez, A.-G. Marcelin, Antimicrobial Agents and Chemotherapy, 52 (2008) 1351-1358.
[32] F. Yang, G. Zheng, T. Fu, X. Li, G. Tu, Y.H. Li, X. Yao, W. Xue, F. Zhu, Physical chemistry chemical physics : PCCP, 20 (2018) 23873-23884.
[33] R. Badia, M. Pujantell, J. Torres-Torronteras, L. Menendez-Arias, R. Marti, A. Ruzo, E. Pauls, B. Clotet, E. Ballana, J.A. Este, E. Riveira-Munoz, Antiviral research, 142 (2017) 123-135.
[34] G.M. Ko, R. Garg, B.A. Bailey, S. Kumar, Current computer-aided drug design, 12 (2016) 135-153.
[35] L. Guasch, A.V. Zakharov, O.A. Tarasova, V.V. Poroikov, C. Liao, M.C. Nicklaus, Current topics in medicinal chemistry, 16 (2016) 441-448.
[36] F.E. Agharbaoui, A.C. Hoyte, S. Ferro, R. Gitto, M.R. Buemi, J.R. Fuchs, M. Kvaratskhelia, L. De Luca, European journal of medicinal chemistry, 123 (2016) 673-683.
[37] W.G. Gu, B.N. Liu, J.F. Yuan, Journal of drug targeting, 23 (2015) 134-139.
[38] M. Masso, G. Chuang, K. Hao, S. Jain, Vaisman, II, Antiviral research, 106 (2014) 5-12.

[39] A. Yan, S. Xuan, X. Hu, Combinatorial chemistry & high throughput screening, 15 (2012) 792-805.
[40] Z. Hu, D. Chen, L. Dong, W.M. Southerland, Ethnicity & disease, 20 (2010) S1-45-49.
[41] V. Hadi, Y.H. Koh, T.W. Sanchez, D. Barrios, N. Neamati, K.W. Jung, Bioorganic & medicinal chemistry letters, 20 (2010) 6854-6857.
[42] C.H. da Silva, V.B. da Silva, C.A. Taft, Drug metabolism letters, 2 (2008) 256-260.
[43] A.G. Cox, V. Nair, Antiviral chemistry & chemotherapy, 17 (2006) 343-353.
[44] E. Deprez, S. Barbe, M. Kolaski, H. Leh, F. Zouhiri, C. Auclair, J.C. Brochon, M. Le Bret, J.F. Mouscadet, Molecular pharmacology, 65 (2004) 85-98.
[45] G.S. Rao, S. Bhatnagar, V. Ahuja, Journal of biomolecular structure & dynamics, 20 (2002) 31-38.
[46] I.J. Chen, N. Neamati, A.D. MacKerell, Jr., Current drug targets. Infectious disorders, 2 (2002) 217-234.
[47] G. Sliwoski, S. Kothiwale, J. Meiler, E.W. Lowe, Jr., Pharmacological reviews, 66 (2014) 334-395.
[48] E. Lionta, G. Spyrou, D.K. Vassilatis, Z. Cournia, Current topics in medicinal chemistry, 14 (2014) 1923-1938.
[49] M. Batool, B. Ahmad, S. Choi, International journal of molecular sciences, 20 (2019).
[50] T. Fischer, S. Gazzola, R. Riedl, Expert Opinion on Drug Discovery, 14 (2019) 791-803.
[51] G. Schneider, D.E. Clark, Angewandte Chemie International Edition, 58 (2019) 10792-10803.
[52] A.S. Rifaioglu, H. Atas, M.J. Martin, R. Cetin-Atalay, V. Atalay, T. Doğan, Briefings in Bioinformatics, 20 (2018) 1878-1912.
[53] P.C. Koneru, A.C. Francis, N. Deng, S.V. Rebensburg, A.C. Hoyte, J. Lindenberger, D. Adu-Ampratwum, R.C. Larue, M.F. Wempe, A.N. Engelman, D. Lyumkis, J.R. Fuchs, R.M. Levy, G.B. Melikyan, M. Kvaratskhelia, Elife, 8 (2019).
[54] C. Fenwick, M. Amad, M.D. Bailey, R. Bethell, M. Bos, P. Bonneau, M. Cordingley, R. Coulombe, J. Duan, P. Edwards, L.D. Fader, A.M. Faucher, M. Garneau, A. Jakalian, S. Kawai, L. Lamorte, S. LaPlante, L. Luo, S. Mason, M.A. Poupart, N. Rioux, P. Schroeder, B. Simoneau, S. Tremblay, Y. Tsantrizos, M. Witvrouw, C. Yoakim, Antimicrob Agents Chemother, 58 (2014) 3233-3244.
[55] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, J Comput Chem, 30 (2009) 2785-2791.
[56] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, J Cheminform, 3 (2011) 33.
[57] O. Trott, A.J. Olson, J Comput Chem, 31 (2010) 455-461.
[58] M.O. Taha, M. Habash, Z. Al-Hadidi, A. Al-Bakri, K. Younis, S. Sisan, J Chem Inf Model, 51 (2011) 647-669.
[59] Y. Yuan, J. Pei, L. Lai, Journal of Chemical Information and Modeling, 51 (2011) 1083-1091.
[60] Y. Yuan, J. Pei, L. Lai, J Chem Inf Model, 51 (2011) 1083-1091.
[61] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Adv Drug Deliv Rev, 46 (2001) 3-26.
[62] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, J Med Chem, 45 (2002) 2615-2623.
[63] A.L. Hopkins, C.R. Groom, A. Alex, Drug Discov Today, 9 (2004) 430-431.
[64] C.H. Reynolds, S.D. Bembenek, B.A. Tounge, Bioorg Med Chem Lett, 17 (2007) 4258-4261.
[65] G.M. Keseru, G.M. Makara, Nat Rev Drug Discov, 8 (2009) 203-212.
[66] A. Daina, O. Michielin, V. Zoete, Sci Rep, 7 (2017) 42717.

[67] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmuller, A.D. MacKerell, Jr., Nat Methods, 14 (2017) 71-73.
[68] Y. Wardi, Journal of Optimization Theory and Applications 59 (1988) 307–323.
[69] R. Kumari, R. Kumar, A. Lynn, Journal of Chemical Information and Modeling, 54 (2014) 1951-1962.
[70] M.A. Islam, T.S. Pillay, J Biomol Struct Dyn, 37 (2019) 503-522.
[71] A.S. Rose, A.R. Bradley, Y. Valasatava, J.M. Duarte, A. Prlic, P.W. Rose, Bioinformatics, 34 (2018) 3755-3758.
[72] A.N. Engelman, The Journal of biological chemistry, 294 (2019) 15137-15157.
[73] J.J. Kessl, J.O. Eidahl, N. Shkriabai, Z. Zhao, C.J. McKee, S. Hess, T.R. Burke, Jr., M. Kvaratskhelia, Molecular pharmacology, 76 (2009) 824-832.
[74] V.K. Vyas, S. Shah, M. Ghate, Medicinal Chemistry Research, 26 (2017) 532-550.
[75] R.S. Bojja, M.D. Andrake, G. Merkel, S. Weigand, R.L. Dunbrack, Jr., A.M. Skalka, The Journal of biological chemistry, 288 (2013) 7373-7386.
[76] L.Q. Al-Mawsawi, A. Hombrouck, R. Dayam, Z. Debyser, N. Neamati, Virology, 377 (2008) 355-363.
[77] E.A. Semenova, N.M. Gashnikova, T.V. Il'ina, T.R. Pronyaeva, A.G. Pokrovsky, Biochemistry. Biokhimiia, 68 (2003) 988-993.
[78] A.M. Dirac, J. Kjems, European journal of biochemistry, 268 (2001) 743-751.
[79] J. Balzarini, A. Karlsson, V.V. Sardana, E.A. Emini, M.J. Camarasa, E. De Clercq, Proceedings of the National Academy of Sciences, 91 (1994) 6599-6603.
[80] A. Zamborlini, A. Coiffic, G. Beauclair, O. Delelis, J. Paris, Y. Koh, F. Magne, M.L. Giron, J. Tobaly-Tapiero, E. Deprez, S. Emiliani, A. Engelman, H. de The, A. Saib, The Journal of biological chemistry, 286 (2011) 21013-21022.
[81] A. Petrou, P. Eleftheriou, A. Geronikaki, M.G. Akrivou, I. Vizirianakis, Molecules, 24 (2019).