Computational Biology and Chemistry 79 (2019) 119–126

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Gp41 inhibitory activity prediction of theaflavin derivatives using ligand/structure-based virtual screening approaches

Tahereh Mostashari-Rad[a], Lotfollah Saghaei[a], Afshin Fassihi[a,b,*]

[a] Department of Medicinal Chemistry, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, 81746-73461, Isfahan, Iran
[b] Bioinformatics Research Center, School of Pharmacy and Pharmaceutical sciences, Isfahan University of Medical Sciences, Isfahan, Iran

ARTICLE INFO

Keywords:
Anti-HIV agents
Virtual screening
Gp41
Lipinski's Rule of Five
Molecular docking
Molecular dynamics simulation

ABSTRACT

Gp41 and its conserved hydrophobic groove on the NHR region is one of the attractive targets in the design of HIV-1 entry inhibitory agents. This hydrophobic pocket is very critical for the progression of HIV and host cell fusion. In this study different ligand-based (structure similarity search) and structure-based (molecular docking and molecular dynamic simulation) methods were performed in a virtual screening procedure to select the best compounds with the most probable HIV-1 gp41 inhibitory activities. In silico pharmacokinetics and ADMET (absorption, distribution, metabolism, excretion and toxicity) properties filtration also was considered to choose the compounds with best drug-like properties. The results of molecular docking and molecular dynamic simulations of the final selected compounds showed suitable stabilities of their complexes with gp41. The final selected hits could have better pharmacokinetics properties than the template compound, theaflavin digallate (**TF_3**), a naturally-originated potent gp41 inhibitor.

1. Introduction

Fusion of human immunodeficiency virus type 1 (HIV-1) with the immune system host cell membranes is mediated by two viral glycoproteins, gp41 and gp120, and some receptors located on the host cells. In the first step of viral entrance into the host cell, gp120 attaches to the host cell receptors. Then gp41 changes its conformation from a non-fusogenic to a fusogenic form. The fusogenic conformation of gp41 consists of three N-terminal heptad repeat coiled coil and three C-terminal heptad repeats (NHR and CHR, respectively) creating a six-helix bundle (6-HB) structure. After the formation of 6-HB structure the viral and host cell membranes fuse together and the viral genome moves into the host cell. Any chemical compound that inhibits these viral entry-involved proteins can inhibit the fusion process (Lu et al., 1995).

Gp41 is one of the attractive targets in the design of HIV-1 entry inhibitory agents. Some conserved parts of this glycoprotein are essential for all universal HIV strains. Studies have shown that the coiled coil intermediate of gp41 is exposed and susceptible to inhibition for quite a long period of time (Dimitrov et al., 2005). Thus, the conserved hydrophobic grooves on the NHR regions are considered as suitable targets for HIV entry inhibition drug design. These hydrophobic pockets are very critical for the formation and stability of 6-HB structure. Chemical compounds weighting up to 600 Da can occupy these pockets. C-peptid inhibitors derived from CHR region of gp41 have shown potent anti-HIV-1 inhibitory activities (Naider and Anglister, 2009). Enfovirtide, the first FDA-approved HIV-1 fusion inhibitor, is a peptide which consists of 36 amino acids derived from CHR region of gp41 (Matthews et al., 2004). Many other peptides targeting different parts of gp41 have been reported as potent HIV-1 entry inhibitors for HIV strains resistant to the common anti-HIV-1 therapies (Jiang et al., 1993). They can also be used as vaccine antigens but they commonly suffer from weak pharmacokinetic properties and high costs. Therefore, compounds having low molecular weights are regarded more valuable gp41 inhibitors. A number of small-molecules with different chemical structures such as *N*-substituted pyrrole, *N*-carboxyphenylpyrrole, furan and indole derivatives (Jiang et al., 2004; Liu et al., 2008; Zhou et al., 2011, 2014) have been reported as gp41 inhibitors.

Virtual screening is a computer-assisted protocol comprised of one or more computational methods applied to select the best compounds with the desired biological activities among the compounds in a large molecular database. Both ligand-based methods, such as similarity searching, QSAR modeling and pharmacophore modeling, and structure-based approaches for instance, molecular docking simulations are usually exploited in virtual screening projects. For ligand-based methods no structural information of the protein is needed and these

* Corresponding author at: Department of Medicinal Chemistry, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, 81746-73461, Isfahan, Iran.
E-mail address: fassihi@pharm.mui.ac.ir (A. Fassihi).

https://doi.org/10.1016/j.compbiolchem.2019.02.001
Received 4 November 2018; Received in revised form 28 January 2019; Accepted 2 February 2019
Available online 05 February 2019

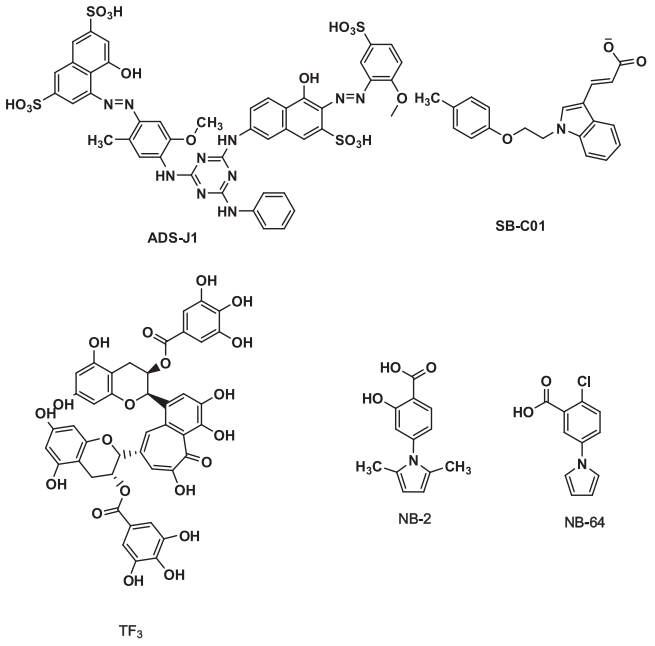

Fig. 1. Chemical structures of ADS-J, SB-C01, TF_3, NB-2 and NB-64.

methods are usually simple and fast. In structure-based approaches the three-dimensional structures of biological targets are needed. Sometimes a combination of both ligand-based and structure-based methods is used for a more successful virtual screening protocol which may start with a similarity search or pharmacophore screening based on a biologically active compound. Then more expensive computational structure-based methods such as molecular docking are used to narrow down the collection of hits.

There are a few reports of exploiting virtual screening methods to find small-molecule gp-41 inhibitors. Debnath, Jiang and coworkers were the first group to perform such methods targeting the conserved deep pocket on gp41 (Debnath et al., 1999; Jiang and Debnath, 2000). Using a structure-based protocol, they reached **ADS-J1** (Fig. 1) as a potent gp41 inhibitor. Due to its high molecular weight (1177 Da), **ADS-J1** was against the rules of drug-likeness properties of a suitable lead. In another research by Holden et al. (Holden et al., 2012) the interactions of 500 000 compounds from the ZINC database with gp41 lipophilic pocket residues were considered as the criteria for structure-based screening. Anti-HIV-1 evaluations of the best selected compounds resulted in a potent HIV-1 fusion inhibitor, **SB-C01** (Fig. 1).

Some small-molecule natural compounds have also shown anti-HIV-1 activities via inhibiting gp41 receptor. Catechins from green tea, theaflavins isolated from black tea and tannins extracted from Chinese medicinal herbs, all having polyphenolic structures, have been introduced as gp41 inhibitors. Theaflavins were more potent anti-HIV-1 compounds compared to catechins and tannins. Theaflavin digallate (**TF_3**, Fig. 1) had the best activity for inhibition of p24 production and low micromolar IC_{50}s for cell-cell and virus-cell fusion inhibition (Liu et al., 2005; Lü et al., 2004). **TF_3** can be regarded as a good lead for designing a new group of HIV-1 entry inhibitors. Jiang et al. have proven that these compounds can inhibit the HIV-1 fusion process by blocking gp41 and 6-HB formation. Polyphenols from black tea have been proven to be more potent gp41 inhibitors than the compounds from green tea (Liu et al., 2005).

One of the most important structural properties of gp41 inhibitors is the presence of one or more negatively-charged groups to interact with the positively-charged residues such as Lys574 or Arg579 in the gp41active site (Jiang et al., 2004). The hydroxyl groups on the polyphenolic natural compounds induce negative charges that can interact with the positively-charged amino-acids in gp41 lipophilic pocket. Docking studies have shown good placement and proper interactions of these compounds in this lipophilic pocket (Liu et al., 2005). Jiang's research group has introduced the polyphenolic compounds from black tea as lead compounds for HIV-1 entry inhibition drug design.

In the present study we selected Theaflavin digallate to search structurally related compounds in available databases in order to find some novel small molecular HIV-1 entry inhibitors. A similarity search from online PubChem database (Kim et al., 2016) was carried out and more than 70 000 compounds were extracted. These compounds were subjected to some filters to select the most desired compounds as gp41 inhibitors. These filters were 1) Lipinski's Rule of Five for drug-likeness properties (Lipinski et al., 2001), 2) docking of the compounds using PyRx (PyRx - Python Prescription Virtual Screening Tool 0.8, 2008-2010Python Prescription Virtual Screening Tool, 2008PyRx - Python Prescription Virtual Screening Tool 0.8, 2008-2010) to select the ones with the highest estimated binding energies, 3) evaluating the ADMET properties of the selected compounds using admetSAR online server (admetSAR@LMMD, http://lmmd.ecust.edu.cn/admetsar1/predict), 4) selecting the compounds with the lowest binding free energies and the best interactions with the gp41 lipophilic pocket's residues using autodock software, and 5) molecular dynamics simulations of the best compounds for further investigations of the interactions with gp41 amino acids. The exploited filters let us to select compounds with the best pharmacokinetics and pharmacodynamics features which render them drug-like properties. The overall virtual screening process applied in this research is illustrated in Fig. 2.

2. Methods

2.1. Similarity search

In the first step of this virtual screening experiment theaflavin isolated from black tea was used as a template for a similarity search in PubChem online database (Kim et al., 2016). This database contains chemical structures, bioactivity, health & safety, etc. that provides a good access of useful data for scientists.

2.2. Drug-likeness properties

All molecules with desired biological activities are not necessarily

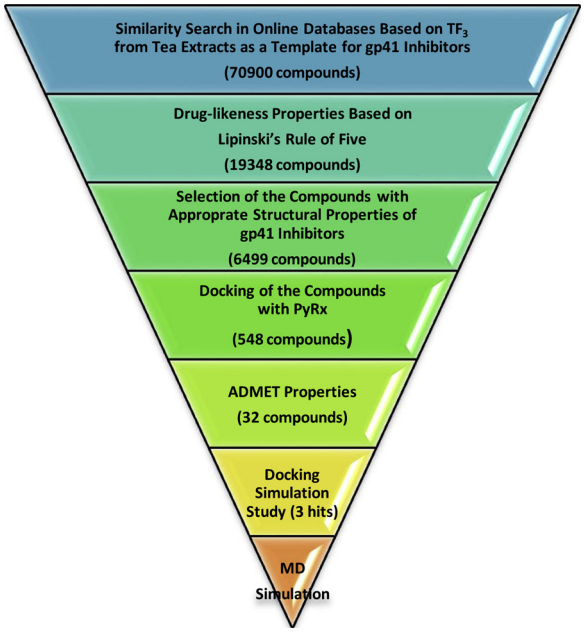

Fig. 2. The overall virtual screening process.

drug molecules. A drug molecule should have drug-like properties. Drug-like molecules should satisfy the Lipinski's Rule of Five and have a balance between lipophilicity and hydrophilicity. Compounds with more than one violation of the following properties do not have the chance to be an orally active drug: 1) not more than five hydrogen bond donors (the sum of OH and NH groups) 2) not more than ten hydrogen bond acceptors (all of N and O atoms), 3) a molecular weight under 500 g/mol, 4) octanol-water partition coefficient in terms of log P less than five. PubChem online database also possesses a filter for the Lipinski's Rule of Five. The obtained compounds from the similarity search step were passed through Lipinski's Rule of Five filter for drug-likeness properties available on this site. 19 348 compounds were in the Lipinski's Rule of Five criteria. 6 499 compounds with appropriate structural features for an orally active gp41 inhibitor were selected. The three-dimensional (3D) structure of these compounds in SDF format were downloaded and saved for docking studies.

2.3. Structural requirements for interacting gp41 binding site

Compounds with drug-like features selected from the previous step were saught for having the structural necessities for interacting gp41 binding site. These necessities were the presence of negatively charged groups such as hydroxyl, carboxyl, tetrazol and nitro which can interact with the positively-charged residues of the receptor and some lipophilic groups to interact with the lipophilic amino acids in the gp41 binding site.

2.4. Docking of the compounds with PyRx

The pdb structure of HIV-1 gp41 core structure (PDB ID; 1AIK) was downloaded from the Protein Data Bank (Berman et al., 2000; Chan et al., 1997; Protein Data Bank, www.rcsb.org) 6 499 molecules having proper drug-like characters as well as structural features for binding to gp41 binding site were docked into the lipophilic pocket of gp41 using autodock vina in PyRx 0.8, freely accessible on http://PyRx.sourceforge.net/downloads (PyRx - Python Prescription Virtual Screening Tool 0.8, 2008-2010Python Prescription Virtual Screening Tool, 2008PyRx - Python Prescription Virtual Screening Tool 0.8, 2008-2010). The number of runs for docking of each compound was 100. Lamarckian genetic algorithm (LGA) was selected as the search algorithm. The parameters for LGA were as follows: mutation rate of 0.02, 2.5×10^6 energy evaluations, initial population of 150 randomly placed individuals, a crossover rate of 0.80 and a maximum number of 27 000 generations. 548 compounds with the best docking scores were selected for in silico ADMET properties analysis.

2.5. In silico ADMET evaluation

Prediction and determination of the absorption, distribution, metabolism, elimination and toxicity (ADMET) properties of a group of selected compounds can speed-up the lead identification. The prediction of ADMET properties is possible according to a series of QSAR/QSPR models and the structural properties of molecules (Hassan Khan, 2010). AdmetSAR (admetSAR@LMMD, 2019 http://lmmd.ecust.edu.cn/admetsar1/predict/; Cheng et al., 2012), a chemoinformatics-based web server that can predict the most important molecular properties was exploited to acquire the ADMET properties of the 548 selected compounds to yield 32 ones with the best properties for further in silico investigations.

2.6. Docking simulation study

The 32 compounds obtained from the previous step were docked into the lipophilic pocket of gp41 using AutoDock 4.2 (Morris et al., 2009). The crystal structure of gp41 (PDB ID; 1AIK) was downloaded from RCSB Protein Data Bank (Berman et al., 2000; Protein Data Bank, www.rcsb.org). All the crystalized water molecules were deleted from the pdb file of the protein by Accelrys Discovery Studio Visualizer 4.0 program (Discovery Studio 4.0 Visualizer, Accelrys Software Inc. San Diego CA USA, 2014). The other steps of the preparation of the protein were carried out in the AutoDockTools package. At first, hydrogens were added to the protein then, after calculating the Kolman united atom charges, non-polar hydrogens were merged into their adjacent carbon atoms. This protein was saved as a pdbq and then pdbqt file formats. The ligands structures were drawn in ChemDraw program (ChemDraw Ultra12.0, Cambridge soft, USA, 2009). Geometry optimization of the ligands structures were carried out by HyperChem8 software (HyperChem 8.0.3, 2007) using MM^+ force field and PM3 semi-empirical calculations. Then, hydrogens were added to the ligands structure in AutoDockTools environment. The charges were calculated with Gasteiger-Marsili procedure and the non-polar hydrogens were merged with their corresponding carbon atoms. After determination of the torsion root and rotatable bonds, ligands were saved as pdbqt files. For grid parameter file preparation there should be one map for each atom of the ligand, an electrostatics and a desolvation map. After determination of the atom types, the size and the place of the grid box were determined. The center of the grid box was set to the coordinates of the α-carbon of the catalytic residue Lys574. The grid box with $60 \times 60 \times 60$ points in x, y and z directions and the grid point spacing of 0.375 Å was generated. The grid box in this size would let the ligands to rotate freely and contains all the important residues of the gp41 lipophilic pocket. Then the grid parameter file was generated using AutoGrid. The Lamarckian genetic algorithm (LGA) was selected for the global optimum binding position search. For all docking procedures, 100 runs of AutoDock search were performed with the step sizes of 0.2 Å for translations and 5° for orientations and torsions. The default settings that were used for LGA were as follows: a maximum number of 2.5×10^6 energy evaluations, a maximum number of 2.7×10^4 generations and mutation and crossover rates of 0.02 and 0.8, respectively. The results of the dockings were ranked based on the estimated binding free energies and clusters presented in docking log files (dlg). The visualization of the ligand-protein interactions based on the resulting dlg files were carried out by Autodock Tools.

2.7. Molecular dynamic simulation study

The three ligands selected in molecular docking studies as well as **TF_3** were subjected to molecular dynamic simulation study. For each compound a 25 ns MD simulation was performed using the GROMACS 5.1.2 software package. PRODRG online server (Schuttelkopf and van Aalten, 2004) was used to generate the necessary topology files and other force field parameters for the four compounds. In the first steps, the Gromos43a1 force field and the SPC water model was employed. Sufficient amount of sodium counter ions were used to neutralize the system. A dodecahedron box was chosen to wrap the whole system. At first the energy minimization of the system was carried out using the steepest descent method (Hess et al., 1997). Position restraint procedure was performed in association with NVT and NPT ensembles after energy minimization. An NVT ensemble was set at constant temperature of 300 K with time duration of 50 ps. Then, an isothermal–isobaric ensemble (NPT) with constant pressure of 1.0 bar and time duration of 50 ps was performed. The cut-off and the Particle-Mesh Ewald (PME) methods were used to treat the van der Waals and the long-range electrostatic interactions (Darden et al., 1993; Essmann et al., 1995). Berendsen thermostat was used to keep the temperature at 300 K (Berendsen et al., 1984).

3. Results and discussion

3.1. Similarity search

Theaflavin digallate (**TF_3**) was selected as a template for a similarity search in a virtual screening experiment. The online database of Pubchem was searched for compounds with 80% similarity to theaflavin digallate. The PubChem System produces a binary substructure fingerprint for a special given query. The format of this fingerprint is binary data with a four-byte integer prefix, indicating the length of the bit list. This similarity search system can find compounds with different degrees of "similarity" by considering structural connectivity and stereochemical data (Bolton et al., 2008). The Tanimoto equation and a dictionary-based fingerprint, analogous to the Molecular ACCess System (MACSS) structure-based keys (Durant et al., 2002) are used in this search process. The result of this similarity search was a library of 70 900 ligands which were subjected to the following pharmacokinetics and pharmacodynamics filters.

3.2. Drug-likeness properties

The "drug-like" phrase commonly refers to compounds that have chemical or physical properties in accordance with the known drug molecules. One of the most popular and useful rules for this purpose is the "Lipinski's Rule of Five". This rule states the physicochemical properties of small-molecule drugs. The Lipinski's Rule includes molecular weight, number of H-bond donors, number of H-bond acceptors and partition coefficient (Chan et al., 1997). The exceptions of Lipinski's Rule of Five are drugs that use the active transport systems, oligo-nucleotides, proteins and natural compounds (Berman et al., 2000; Protein Data Bank, 2019, www.rcsb.org). This rule helps to make an easier selection of molecules with better pharmacokinetic properties in human body for oral formulations. Among the 70 900 compounds in the similarity search, 19 345 compounds passed this filter.

3.3. Structural requirements for interacting gp41 binding site

The main structural necessities for interacting gp41 binding site are: 1) some hydroxyl groups to induce negative charges or carboxyl, tetrazol and nitro groups being able to interact with the positively-charged residues of Lys574 or Arg579, 2) some lipophilic groups to interact with the lipophilic amino acids in the binding site of gp41 such as Leu568, Val570, Trp571 and Trp631 (Liu et al., 2005, 2008; Zhou et al., 2014). Based on these features in the structures of the known gp41 inhibitors, 6 499 molecules were selected among the 19 345 compounds and their three dimensional structures in sdf file format were downloaded to enter the next step of the screening.

3.4. Docking studies with PyRx

After docking all 6 499 compounds their resulting docking scores in terms of estimated free energies of binding were checked. The estimated binding free energies of some known gp41 inhibitors were lower than −7.00 kcal/mol; thus, molecules with ΔG_{bind} lower than −7.00 kcal/mol were selected which were 548 out of 6 499. The sdf file formats of these 548 molecules are provided as supplementary information.

3.5. In silico ADMET properties

The use of ADMET prediction of ligands is to eliminate the weak drug candidates and focus on the compounds with more likely successful drug properties. Many in silico methods have been introduced for prediction of ADMET properties. A large number of them are based on structural and chemical properties of the known compounds. They use QSAR or QSPR models to predict ADMET properties of new ligands. Admetsar webserver was used to predict the ADMET properties of 548 compounds that passed the last step. Factors that were considered for ligand selection were as follows: Caco-2 permeability, aqueous solubility, blood-brain barrier permeability, human intestinal absorption, P-glycoprotein inhibitory activity, possibility of being a P-glycoprotein substrate, CYP450 (1A2, 2C19, 2C9, 2D6 and 3A4) inhibitory activity, possibility of being CYP450 2D6 and 3A4 substrate, carcinogenicity, fish toxicity, honey bee toxicity and rat acute toxicity. 32 compounds with the best ADMET properties were selected in this stage. The sdf file formats of these compounds are provided as supplementary information.

3.6. Docking simulation study

Docking simulation study was used to investigate the binding modes of the 32 selected compounds which had proper drug-likeness and ADMET properties. At the first step, the binding modes of **NB-2** and **NB-64** (Fig. 1), two well-known gp41 inhibitors (Jiang et al., 2004), within the entire protein were explored via a blind-docking procedure to validate the applied docking protocol.

Most of molecular docking study programs including Autodock, present docking scores in terms of binding free energies. But due to the limitations of these programs these quantities are not reliable estimators of the true binding free energy. The docking scores (estimated ΔG_{bind}) and interactions with the key amino acids of each ligand were explored to find the best conformation and orientation of the ligand within the binding site of gp41. The final docking results were sorted according to the docking scores. The runs with the best docking scores were regarded as the most stable conformations and orientations. Visualization of the protein-ligand interactions were carried out by the discovery studio software package.

All the final docking information of the 32 compounds including their estimated ΔG_{bind} and the intermolecular interactions are presented in Tables S2 and S3 in supporting information. As it is shown in these tables, all the 32 compounds had shown almost similar interactions with the binding site amino acids. Ligands with the most negative free energies of binding were considered as the most stable ones in the gp41 binding site.

Three compounds, **1**, **2** and **3** (Fig. 3), with the best estimated free energies of binding and the most important interactions with the key amino acids of the gp41 lipophilic pocket were chosen as the most stable ligands in the protein binding site, among the 32 docked compounds. They showed higher binding affinities than **TF_3** and **NB-64**. Compound **1** with the best docking score was the only compound that had electrostatic, π–π and cation-π interactions all together. It also showed lipophilic interactions with all the most critical lipophilic residues in the gp41 binding site such as Leu568, Val570, Trp571 and Trp631, and Hydrogen bonds with Trp571, Gln575 and Arg579.

All the docked molecules except **4**, **7**, **8**, **13**, **18**, **32**, **31** and **24** had lipophilic interactions with Val570. Compounds **1**, **14**, **18** and **34** had π-π interactions with Trp571. Only eight ligands (**1**, **4**, **8**, **12**, **13**, **22**, **25** and **33**) were able to form cation-π interactions with Lys574. Most of the 32 compounds except **18**, **22**, **23**, **24**, **26**, **32** and **27** had appropriate lipophilic interactions in the protein binding site. The structures of the best poses of **1**, **2** and **3** are shown in Fig. 3.

The lipophilic pocket of gp41 has enough space to accommodate the molecules with 500–600 Da molecular mass. The **NB-2** and **NB-64** (molecular weight = 231 and 222, respectively) that have shown good gp41 inhibitory activities do not fill this pocket completely. It has been proven that the ligands with longer molecular backbones and additional hydrophobicity and hydrogen bonding functionalities that can occupy more space inside the lipophilic pocket are stronger gp41 inhibitors (Jiang et al., 2011). The docking results show that the three new hits with longer molecular backbones than **NB-2** and **NB-64** can occupy more space inside this pocket. They also have more lipophilic and hydrogen bonding functionalities that provides better docking scores and

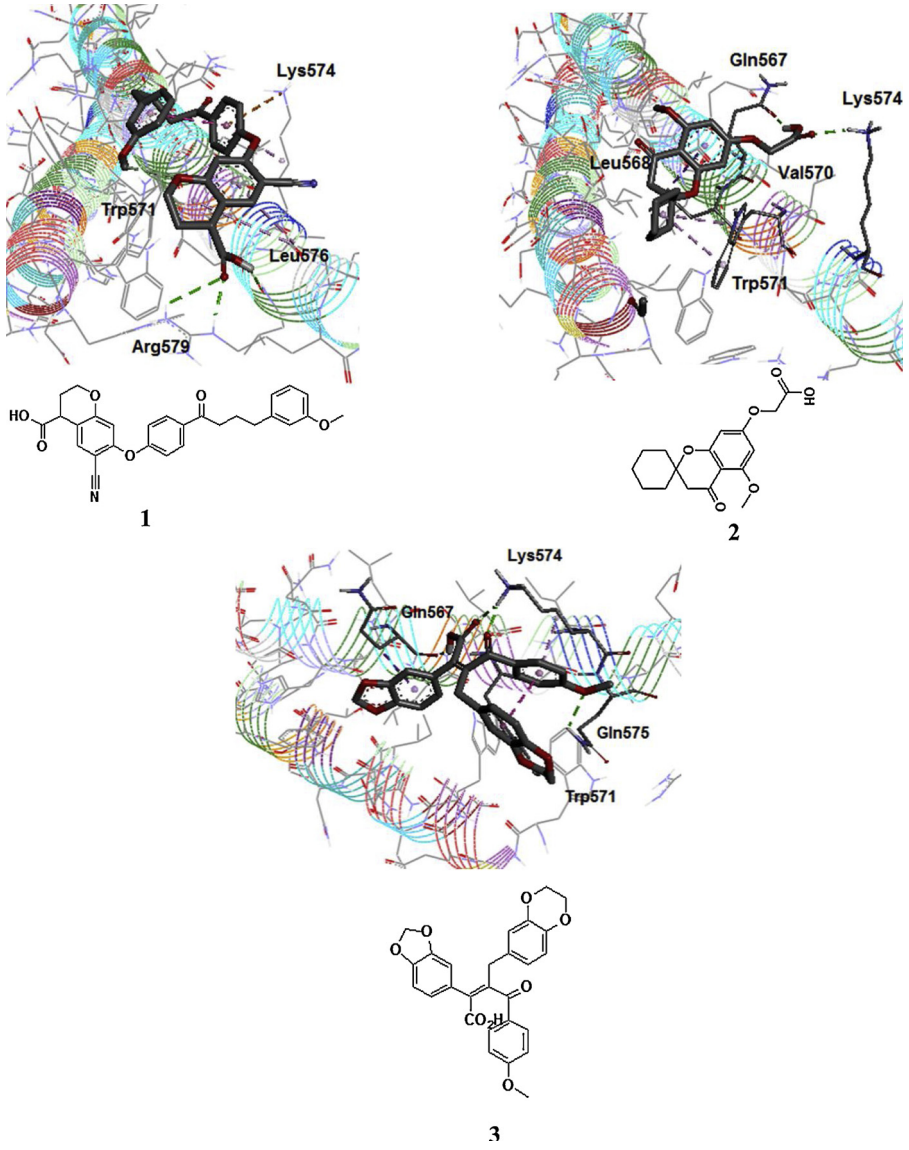

Fig. 3. The chemical structures and intractions of the three selected hits, **1**, **2** and **3** inside the lipophilic pocket of gp41.

more binding affinities than **NB-2** and **NB-64**.

Most of the introduced gp41 inhibitors in the literature have negatively-charged groups such as carboxyl and tetrazole to interact with the positively-charged residues (Lys574 or Arg579) in the gp41 binding pocket via electrostatic interactions. Experimental data have shown that the elimination of these negatively-charged groups abolishes the gp41 inhibitory activity (Liu et al., 2008). **1**, **2** and **3** also have carboxyl groups. The results of docking studies for compound **1** determined electrostatic interaction between the carboxyl group and Arg579. Compounds **2** and **3** showed electrostatic interactions with Lys574.

In the best molecular docking poses **1** showed π-π stacking interactions between chromene moiety and indole ring of Trp571 and cation-π interactions between phenyl ring and NH_3^+ of Lys574. But two other ligands (**2** and **3**) had no π-π and cation-π interactions with the residues in the binding site of gp41. All the three hit compounds had hydrophobic interactions with Gln567, Val570 and Trp571. There were some other hydrophobic interactions with Leu568 Lys574, Gln575, Ala578 and Trp631. The overall docking results show that these three compounds can be proposed as suitable inhibitors of gp41 with anti-HIV-1 activities.

3.7. Molecular dynamic simulation study

Molecular dynamics (MD) simulation is an in silico method useful in identification of protein behavior, drug–receptor interactions, solvation of molecules and conformational changes that macromolecules undergo in biological environments. Unlike molecular docking simulations, MD simulations consider macromolecules to be flexible body structures and more similar to their natural structures in biological environments. Since the study of macromolecule structures in the lab is very difficult and time-consuming, MD simulation studies will be very helpful alternatives. To investigate the detailed positions and interactions of the selected ligands in the molecular docking step an MD simulation study was performed. The best docked conformations of the three selected compounds (**1**, **2** and **3**) and **TF_3** as the reference with the best docking scores and orientations in the active site were selected for MD simulations.

After 25 ns molecular dynamic simulations of the three selected compounds and **TF_3** in complex with gp41, the resulting trajectories were completely analyzed to evaluate the stability, structural properties and convergence of the system. Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were evaluated to assess the stability and fluctuations of protein backbone and alpha carbon atoms.

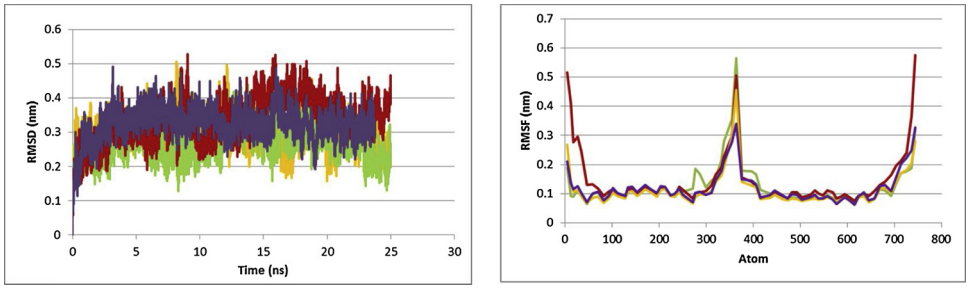

Fig. 4. Superimposed RMSD of the backbone atoms and RMSF plots of gp41 in complex with **1** (green), **2** (red), **3** (purple) and **TF$_3$** (yellow).

The average RMSD values for each complex were between 0.15 and 0.50 nm confirming the system stability.

As can be seen in Fig. 4, compounds **3** and **TF$_3$** had the least fluctuations in RMSD, thus the most stability during the MD simulations compared to the other two compounds under the same MD simulation conditions. The RMSD values for **3** had a rising in the first 3 ns and then stayed stable in the rest of the simulation time (Fig. 4). The superimposed RMSD plot, Fig. 4, shows almost the same values for all the compounds during 3–5 ns and 21–23 ns of the MD simulations.

RMSF measurement is a useful way to recognize the residues that have the strongest interactions with a specific ligand. The higher value of RMSF shows the higher amount of atomic mobility of the Cα atoms of the protein in the MD simulations. Fig. 4 shows the residues in the N-terminal and C-terminal of gp41 that are far from the binding site have the highest RMSF values (0.2–5.2 nm). The key residues in the active site (Trp571, Lys574 and Arg579) have the lowest amounts of RMSF in the four MD simulations. The small peak in RMSF of compound **1** for atoms 270–310 is related to Cα atoms of Gln563 and His564. The more flexibility of these carbon atoms is due to the fewer interactions of these amino acids with the ligand compared to other residues in the gp41binding site. The RMSF values for the three hits and also **TF$_3$** show suitable and stable placement of the ligands in the lipophilic pocket of gp41 during the MD simulations.

Hydrogen bonds are one of the most important interactions in a stable gp41-ligand complex. Fig. 5 shows the numbers of hydrogen bonds formed in the complex of the three hits and **TF$_3$** with gp41 during 25 ns MD simulations. Compound **1** showed more hydrogen bonds with gp41. In all the other three simulations almost the same amounts of hydrogen bonds were observed.

The radius of gyration (Rg) in MD simulation is an indicator of the compactness of the protein. Time-dependency plots of the radius of gyration for the simulated complexes are shown in Fig. **S2**. For the complexes of **2** and **3** with gp41 the Rg reduced slightly showing the increase in the compactness of the complexes. But, for the complexes of **1** and **TF$_3$** the Rg values were almost constant during the simulations.

At the end of MD simulations some changes were observed in the positions and orientations of the ligands compared to the most stable conformers of the molecular dockings. The chemical interactions after MD simulations of **1** and **3** are shown in Fig. 6. These observations show some of the useful applications of molecular dynamic simulations after molecular docking of ligands. In compound **1** very small changes in orientation caused different chemical interactions compared to molecular docking results. For example the nitrogen atom of cyano- group had a hydrogen bond with Lys574, but in molecular docking studies the carbon chain of this amino acid showed π-alkyl interactions with the central phenyl ring. This ring had another π-alkyl interaction with Ala578. As it was observed in the molecular docking studies, the lipophilic residues such as Leu568, Trp571 and Leu581 are positioned in proximity of compound **1** in MD results. The terminal phenyl ring of this compound had a π-alkyl interaction with Val570. For compound **3** a conformational change in MD simulations resulted in the formation of π–π stacking interactions between 2,3-dihydrobenzo[*b*][1,4]dioxine groups and Trp571 while this amino acid formed a hydrogen bond with one of the 2,3-dihydrobenzo[*b*][1,4]dioxine groups in molecular docking studies. Other residues in the lipophilic pocket such as Gln567, Leu568, Val570, Trp631 and Glu634 are positioned around this ligand with small changes in the spatial coordinates compared to the docking study results.

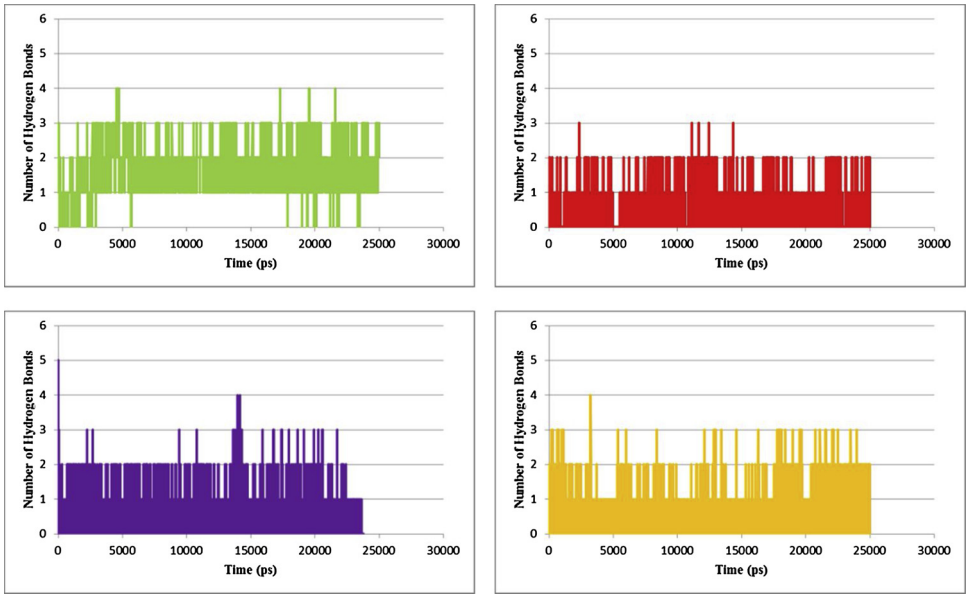

Fig. 5. Numbers of hydrogen bonds formed between **1** (green), **2** (red), **3** (purple) and **TF$_3$** (yellow) and gp41 during MD simulation.

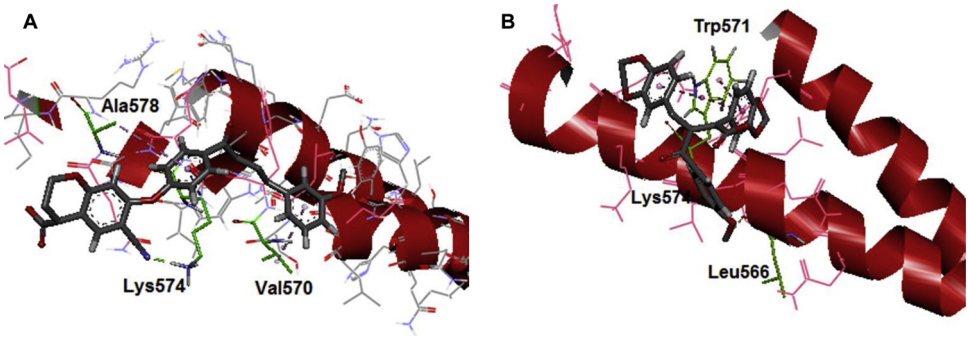

Fig. 6. Chemical interactions of **1** (A) and **3** (B) in the lipophilic pocket of gp41 after MD simulations.

4. Conclusions

In this study three small molecules were proposed as possible gp41 inhibitors based on a combination of ligand-based and structure-based methodologies in the format of a virtual screening protocol. In the first step, a structure similarity search using the theaflavin digallate (**TF_3**), a potent gp41 inhibitor, as the template compound was performed. Lipinski's Rule of Five was considered to select compounds with better pharmacokinetics properties that could be more suitable for oral absorption in human body. Molecular docking investigations with PyRex helped to estimate the affinities of the selected compounds to gp41 lipophilic pocket. The compounds with the higher affinities to gp41 were subjected to in silico ADMET properties prediction. Molecular docking with Autodock was carried out to select the ligands with the best docking scores among the compounds with appropriate ADMET properties. Comparisons of the chemical interactions and docking scores of the docked compounds with **TF_3** and **NB-64**, two known gp41 inhibitors, were also considered. Compounds **1**, **2** and **3** with higher docking scores, binding affinities and more electrostatic interactions with Lys574 and Arg579 than **TF_3** and **NB-64** were selected as the best potential gp41 inhibitors. Although **TF_3** is a potent gp41 inhibitor, it has some violations of the Lipinski's rule of 5 such as having more than five hydrogen bond donors (13 hydroxyl groups) and a molecular weight higher than 500 g/mol (868.7 g/mol). The final three selected hits do not have these violations. Hydroxyl groups in **TF_3** structure induce negative charges for electrostatic interactions with the gp41 binding site but they make it too hydrophilic to pass through the blood brain barrier and it has low gastrointestinal absorption. Compounds **1**, **2** and **3** indicate a perfect balance between hydrophilicity and lipophilicity and could have better pharmacokinetics properties than **TF_3**. Docking results showed that the carboxylic acid moieties in the structure of these compounds serve for electrostatic interactions the same as the **TF_3** hydroxyl groups. Finally molecular dynamics simulations of the three best potential gp41 inhibitors and **TF_3** were executed. The resulting RMSD, RMSF and Rg changes during the time of simulations confirmed the stability of the complexes of the three proposed hits and **TF_3** with gp41. Future experimental investigations of these compounds will verify the predicted computational activities.

Author contribution statement

This article was extracted from the PhD research project of Tahereh Mostashari-Rad. She performed all this research and wrote the manuscript by herself. Lotfollah Saghaei and Afshin Fassihi were the supervisors of this project.

Conflicts of interest

There are no conflicts of interest in this report.

Acknowledgment

The authors would like to thank the School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran for financing this project.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compbiolchem.2019.02.001.

References

admetSAR@LMMD, http://lmmd.ecust.edu.cn/admetsar1/predict/. (Accessed 1 August 2018).
Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684–3690. https://doi.org/10.1063/1.448118.
Berman, H.M., Feng, J.Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242 PMID: 10592235.
Chan, D.C., Fass, D., Berger, J.M., Kim, P.S., 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell. 89, 263–273 PMID: 9108481.
ChemDraw Ultra12.0, 2009. Cambridge soft, USA.
Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P., Tang, Y., 2012. admetSAR: a comprehensive source and free tool for evaluating chemical ADMET properties. J. Chem. Inf. Model. 52, 3099–3105. https://doi.org/10.1021/ci300367a.
Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. https://doi.org/10.1063/1.464397.
Debnath, A.K., Radigan, L., Jiang, S., 1999. Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. J. Med. Chem. 42, 3203–3209. https://doi.org/10.1021/jm990154t.
Dimitrov, A.S., Louis, J.M., Bewley, C.A., Clore, G.M., Blumenthal, R., 2005. Conformational changes in HIV-1 gp41 in the course of HIV-1 envelope glycoprotein-mediated fusion and inactivation. Biochemistry (Moscow) 44, 12471–12479. https://doi.org/10.1021/bi051092d.
Discovery Studio 4.0 Visualizer, 2014. Discovery Studio 4.0 Visualizer. Accelrys Software Inc., San Diego CA USA.
Durant, J., Leland, B., Henry, D., Nourse, J., 2002. Reoptimization of MDL keys for use in drug discovery. J. Chem. Inf. Comput. Sci. 42, 1273–1280. https://doi.org/10.1021/ci010132r.
Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577–8593. https://doi.org/10.1063/1.470117.
Hassan Khan, M.T., 2010. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modeling approaches. Curr. Drug Metab. 11, 285–295 PMID: 20450477.
Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J., 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
Holden, P.M., Kaur, H., Goyal, R., Gochin, M., Rizzo, R.C., 2012. Footprint-based identification of viral entry inhibitors targeting HIVgp41. Bioorg. Med. Chem. Lett. 22, 3011–3016. https://doi.org/10.1016/j.bmcl.2012.02.017.
HyperChem 8.0.3, 2007.
Jiang, S., Debnath, A.K., 2000. Development of HIV entry inhibitors targeted to the coiled-coil regions of gp41. Biochem. Biophys. Res. Commun. 269, 641–646. https://doi.org/10.1006/bbrc.1999.1972.
Jiang, S., Lin, K., Strick, N., Neurath, A.R., 1993. HIV-1 inhibition by a peptide. Nature 365, 113. https://doi.org/10.1038/365113a0.
Jiang, S., Lu, H., Liu, S., Zhao, Q., He, Y., Debnath, A.K., 2004. N-Substituted pyrrole

derivatives as novel human immune deficiency virus type 1 entry inhibitors. Antimicrob. Agents Chemother. 48, 4349–4359. https://doi.org/10.1128/AAC.48.11.4349-4359.2004.

Jiang, S., Tala, S.R., Lu, H., Zou, P., Avan, I., Ibrahim, T.S., Abo-Dya, N.E., Abdelmajeid, A., Debnath, A.K., Katritzky, A.R., 2011. Design, synthesis, and biological activity of a novel series of 2,5-disubstituted furans/pyrroles as HIV-1 fusion inhibitors targeting gp41. Bioorg. Med. Chem. Lett. 21, 6895–6898. https://doi.org/10.1016/j.bmcl.2011.08.081.

Kim, S., Thiessen, P., Bolton, E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B., Wang, J., Yu, B., Zhang, J., Bryant, S., 2016. PubChem substance and compound databases. Nucleic Acids Res. 44, 202–213. https://doi.org/10.1093/nar/gkv951.

Lipinski, C., Lombardo, F., Dominy, B., Feeney, P., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 46, 3–26 PMID: 11259830.

Liu, S., Lu, H., Zhao, Q., He, Y., Niu, J., Debnath, A.K., Wu, S., Jiang, S., 2005. Theaflavin derivatives in black tea and catechin derivatives in green tea inhibit HIV-1 entry by targeting gp41. Biochim. Biophys. Acta 1723, 270–281. https://doi.org/10.1016/j.bbagen.2005.02.012.

Liu, K., Lu, H., Hou, L., Qi, Teixeira, Z.C., Barbault, F., Fan, B.T., Liu, S., Jiang, S., Xie, L., 2008. Design, synthesis, and biological evaluation of N-carboxyphenylpyrrole derivatives as potent HIV fusion inhibitors targeting gp41. J. Med. Chem. 51, 7843–7854. https://doi.org/10.1021/jm800869t.

Lu, M., Blacklow, S., Kim, P., 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nat. Struct. Biol. 2, 1075–1082 PMID: 8846219.

Lü, L., Liu, S., Jiang, S., Wu, S., 2004. Tannin inhibits HIV-1 entry by targeting gp41. Acta Pharmacol. Sin. 25, 213–218 PMID:14769212.

Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., Bolognesi, D., 2004. Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. Nat. Rev. Drug Discov. 3, 215–225. https://doi.org/10.1038/nrd1331.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J.J., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. Comput. Chem. 30, 2785–2791. https://doi.org/10.1002/jcc.21256.

Naider, F., Anglister, J., 2009. Peptides in the treatment of AIDS. Curr. Opin. Struct. Biol. 19, 473–482. https://doi.org/10.1016/j.sbi.2009.07.003.

Protein Data Bank, www.rcsb.org. (Accessed 15 April 2018).

Python Prescription Virtual Screening Tool, http://PyRx.sourceforge.net/downloads. (Accessed 15 July 2018).

Schuttelkopf, A.W., Van Aalten, D.M.F., 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D Biol. Crystallogr. 60, 1355–1363. https://doi.org/10.1107/S0907444904011679.

Zhou, G., Wu, D., Snyder, B., Ptak, R.G., Kaur, H., Gochin, M., 2011. Development of indole compounds as small molecule fusion inhibitors targeting HIV-1 glycoprotein-41. J. Med. Chem. 54, 7220–7231. https://doi.org/10.1021/jm200791z.

Zhou, G., Sofiyev, V., Kaur, H., Snyder, B., Mankowski, M.K., Hogan, P., Ptak, R.G., Gochin, M., 2014. Structure-activity relationship studies of indole-based compounds as small molecule HIV-1 fusion inhibitors targeting glycoprotein-41. J. Med. Chem. 57, 5270–5281. https://doi.org/10.1021/jm500344y.