An insight into the binding mechanism of Viprinin and its morpholine and piperidine derivatives with HIV-1 Vpr: molecular dynamics simulation, principal component analysis and binding free energy calculation study

Satyabrata Majumder and Kalyan Giri

Department of Life Sciences, Presidency University, Kolkata

Communicated by Ramaswamy H. Sarma

ABSTRACT

HIV-1 Vpr is an accessory protein responsible for a plethora of functions inside the host cell to promote viral pathogenesis. One of the major functions of Vpr is the G2 cell cycle arrest. Among several small molecule inhibitors, Viprinin, a coumarin derivative, has been shown to specifically inhibit the G2 cell cycle arrest activity of Vpr thus making it an excellent choice for a lead molecule to design antiretroviral drug. But the exact mechanism of binding of the Viprinin and its two potent derivatives with Vpr is still not understood. In this study with combined molecular docking, molecular dynamics simulation, Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method, Principal component analysis and Umbrella sampling simulation, we have explored the binding mechanism of Viprinin and its two derivatives with Vpr. MM-PBSA and Umbrella sampling calculations suggest that Viprinin and ViprininD1 have higher binding energy than ViprininD2. Molecular dynamics simulation shows that the ligands are not very stable inside the initial binding pocket and various hydrophobic interactions are responsible to hold the ligands with Vpr. Vpr backbone Principle Component Analysis (PCA) shows various unique essential motions of Vpr bound with Viprinin and its two derivatives. This study may give detailed insight of the mode of binding of the specified compounds at atomic scale and provide valuable information about the possibility of using these compounds as a potent Vpr inhibitor.

Abbreviations: HIV-1: Human Immunodeficiency Virus type 1; Vpr: Viral protein r; MM-PBSA: Molecular Mechanics Poisson-Boltzmann Surface Area; RMSD: Root mean square deviation; RMSF: Root mean square fluctuation; PCA: Principal Component analysis; PMF: Potential of Mean Force; AIDS: Acquired immune deficiency syndrome; GA: Genetic Algorithm

1. Introduction

Human immunodeficiency virus 1 is the causative agent of AIDS and it belongs to lentiviral family that encodes six accessory proteins, Tat, Rev, Vpu, Vif, Nef, and Vpr along with retroviral Gag, Pol and Env proteins. Viral protein R (Vpr) is a 96 amino acid, 14 kDa protein (Yuan et al., 1990) and it plays crucial roles in virus spread, pathogenesis and virus replication efficiency (Levy et al., 1994; Ogawa et al., 1989). It performs a myriad of functions inside the host cell to promote disease progression and hence contributes to the development of AIDS. Vpr is characterized by the presence of three well-defined α-helices spanning residues 17-33, 38-50 and 56-77 and random coils at the N- and C-termini (Morellet et al., 2003). Amino acid residues residing in these α-helices are important for its multifunctional role. For example, residues W18, Q3, L22, K27, L23, and F34 of Helix 1 are associated with the cytopathic functions of Vpr (Barnitz et al., 2011; Chen et al., 1999; Stromájer-Rácz et al., 2010); Residues 52-96 (Helix 3 and C-terminal random coil) are shown to bind with ANT and results in the induction of apoptosis (Jacotot et al., 2001); residues in the region 71-82 are involved in mitochondrial membrane permeabilization-inducing (MMP) activity (Jacotot et al., 2001) etc. Among all the functions, two most important functions of Vpr are viral Pre-Integration Complex(PIC) transportation from cytoplasm to the host cell nucleus via interaction with importin-α and G2 cell cycle arrest (González, 2017).

Nuclear import of Viral PIC in non-dividing cells, such as Macrophages, is one of the major function of Vpr (Popov et al., 1998). Previously it was demonstrated that Vpr transports PIC through Importin-α mediated pathway and the region 17-74 is responsible for the interaction (Kamata et al., 2005). Recent crystal structure of Vpr C-terminal 85-96 in complex with Importin-α2 reveals the actual mode of binding between Vpr and Importin-α and the process of Nuclear transportation (Miyatake et al., 2016). In the case of Vpr mediated G2 cell cycle arrest, different mechanisms have been proposed, but the exact mechanism is still poorly understood. A study demonstrated that Vpr forms a chromatin-associated nuclear focus by recruiting VprBP-DDB1 complex and targets a chromatin-bound substrate whose Ubiquitination and proteolysis initiates the activation of ataxia
telangiectasia-mutated and Rad3-related-mediated G2/M checkpoint and induce G2 arrest (Belzile et al., 2010; 2010). Recent crystal structure of DDB1–DCAF1–Vpr–UNG2 complex reveals the mechanism of how Vpr guides UNG2, Uracil DNA glycosylase, towards destruction and induces G2 cell cycle arrest (Wu et al., 2016). Vpr is an excellent choice for therapeutic interventions due to its important role in AIDS pathogenesis.

Several small molecule inhibitors against Vpr have been tested to date but still no drug candidate has been developed to use in Anti-Retroviral Therapy (ART). Fumagillin, a fungal metabolite, was shown to inhibit Vpr mediated cell cycle arrest activity in both yeast and mammalian cells (Popov et al., 1998). Damnacanthal, another compound extracted from Noni (a small evergreen tree), was shown to prevent Vpr-associated cell death but does not inhibit cell cycle arrest (Kamata et al., 2006). A stable derivative of Hematoxylin has been synthesized which specifically inhibits HIV-1 replication in macrophages but not G2 cell cycle arrest (Hagiwara et al., 2015). Recently several Vpr inhibitors have been extracted from the medicinal plants of Myanmar (Win et al., 2017) but further work in optimising these lead compounds into anti-HIV drugs is still insufficient. Through a high-throughput screening (HTS) system, using a library of compounds in the RIKEN NPDepo, it was found that a 3-phenyl coumarin compound NPD4456 (known as Viprinin) was able to show anti Vpr activity (the cell cycle arrest activity) (Ong et al., 2011). It was also found that this compound binds in the hydrophobic core region of Vpr. Further in the same study, several structural derivatives of Viprinin were produced maintaining the 3-phenyl coumarin scaffold and through Structure-Activity Relationship (SAR) study it was found that the Piperidine and Morpholine derivative of 3-phenyl coumarin (ViprininD1 and ViprininD2 respectively) shows high potency against Vpr (Ong et al., 2011) and these compounds also binds in the same hydrophobic region. Structural level details of mode of binding of these three potent compounds will guide development of more potent drug like molecule which will be more selective towards Vpr.

In this study we have explored three aspects: i) the mechanism of how and two of its highly potent derivatives binds with Vpr (Ong et al., 2011) at atomistic level through molecular dynamics simulation and Umbrella sampling simulations, ii) Principal component analysis to investigate the major conformational changes in Vpr backbone caused due to VPR-ligands complex formation during simulation and iii) MM-PBSA method to quantify the binding strengths of the Viprinin and its derivatives with Vpr. Investigation of the mechanism and strength of binding of Viprinin and its derivatives with Vpr may create a direction for AIDS (acquired immune deficiency syndrome) therapy in near future.

2. Materials and methods

2.1. Protein and ligand preparation

HIV-1 Vpr sequence was retrieved from Uniprot (Uniprot ID: P69726, isolate HXB2) and homology model was built using SWISS-MODEL (Schwede et al., 2003) taking the NMR structure of Vpr (PDB ID: 1M8L) (Morellet et al., 2003) as template. Prepare Protein module of BIOVIA Discovery Studio 2018 was used to prepare the protein. This module cleans the protein, optimizes side chain conformation of the residues, protonates the structure at pH 7.4 by predicting titration site pKas and minimizes the protein using CHARMM36 force field (Huang & Mackerrill, 2013) with 1000 steps of Steepest Descent followed by 1000 steps of Conjugate gradient. The side chains of the protein were ionized consistent with a pH of 7.4. Viprinin (NPD4456) molecule was downloaded from Pubchem (PubChem CID: 4837694) and the two derivatives of this compound were drawn in ACD/Chemsketch according to the research article by Ong et al. (Ong et al., 2011). Structural details and Ames mutagenicity details about Viprinin and its derivatives are given in Supplementaries 1 and 2. These three compounds were then prepared using Prepare ligands module of Discovery Studio. pH based ionization method (pH was set to 7.4) was used and all possible Tautomers and Isomers were allowed to generate but no Tautomers and Isomers were generated in this process. The compounds were energy minimized with the same protocol as stated earlier.

2.2. Molecular docking

Two derivatives of Viprinin were named as ViprininD1 and ViprininD2. Viprinin and its two derivatives were docked with Vpr using Autodock 4.2.6 (Morris et al., 2009). The docking simulation box was fixed at the site encompassing Glu25 and Gln65 as specified by the research article by Ong et al. (Ong et al., 2011) with box size 60 Å at all three directions (Xcenter, Ycenter and Zcenter coordinates are −2.909, −5.097 and −0.331 respectively) and the grid spacing was set to 0.375 Å. Initial coordinates of the minimized ligands were used during initiation of docking. Three independent docking runs were performed with varying GA runs (100, 200 and 300), Population size (250, 350 and 400) and RMS cluster tolerance (1 Å, 2 Å and 2 Å). Rest of the parameters were set to its default values. Lamarckian Genetic algorithm search space was used for docking simulation. Docked poses with the highest negative binding energy (i.e. complex with lowest energy/score) from the three docking runs were subjected to Molecular dynamics simulation run for each of the Vpr-ligand system.

2.3. Molecular dynamics simulation

GROMACS 2018.3 software (Abraham et al., 2015) was used to simulate Viprinin-Vpr, ViprininD1-Vpr, ViprininD2-Vpr complexes. GROMOS96 54a7 force field (Schmid et al., 2011) with SPC water model and Dodecahedron simulation box was used in the simulation. ligand parameters were constructed using PRODRG (van Aalten et al., 1996) server (parameter files of the ligands are given in Supplementary 8). The solvated Protein-ligand complexes were neutralized by adding two Na⁺ ions. Steepest Descent method (maximum no. of steps 50000) was employed to energy-minimize the neutralized system. Then the energy minimized systems were subjected to 1 ns NVT equilibration simulation step to raise the temperature of the systems to 300K and successively the NVT equilibrated systems were subjected to 1 ns NPT
equilibration at 1 atm pressure with same temperature. During NVT and NPT equilibration period, the protein-ligand complexes were fully position-restrained with a force constant of 1000 kJ/mol/nm² and 2 fs time step was used. NPT equilibrated systems were then subjected to 50 ns production run with 2 fs time step with no restraint. Smooth Particle-Mesh Ewald method (Essmann et al., 1995) was used to calculate long-range electrostatic interactions and a 12 Å cutoff was used for both PME and van der Waals interactions. For each of the protein-ligand complexes, seven independent 50 ns MD simulations were conducted starting from different random number seeds to assign different sets of initial velocities in each of these simulations. Vpr without any bound ligands was also subjected to seven independent MD simulation run with different initial random number seeds. A total of 1.4 µs of simulation has been performed.

2.4. Simulation trajectory analysis

MD simulation trajectory analysis was done using GROMACS command line tools. Root mean square deviation (RMSD) of protein backbone and ligand heavy atom was calculated using gmx rms, Root mean square fluctuation calculation of protein residues was done using gmx rmsf commands. Plotting was done using Python matplotlib module.

2.5. Principal component analysis (PCA)

PCA is an unsupervised machine learning algorithm; in the context of atomistic simulation it is better described in terms of relation to normal modes or quasi-harmonic analysis that breaks down the MD trajectories into the principal modes of motion sampled. As these motions are often important for protein function (Hayward & Groot, 2008), the dynamics in this low-dimensional subspace spanned by the first few principal components was termed “essential dynamics” and analysis of these PCs performed on Cartesian coordinates has proven to be a valuable tool for studying conformational changes. In a summary, to obtain PC modes first correlation matrix is computed using the following formula:

$$C_{ij} = \frac{1}{N} \sum_{k=1}^{N} (x^k_i - \bar{x}_i)(x^k_j - \bar{x}_j)$$

Here $x^k_i$, $x^k_j$ are a pair of elements of vector $x^k$ (where $x$ is a vector containing the Cartesian coordinates of the $C_a$ atoms of the protein), which describes the configuration of the system at time step $k$, while $\bar{x}_i$, $\bar{x}_j$ are their average values calculated from the $N$ structures sampled in the MD simulation. Then this correlation matrix is diagonalized using the formula $\Lambda = T^T CT$ where $T$ is transformation matrix whose columns are the eigenvectors of the motions and the diagonal elements of $\Lambda$ are the associated eigenvalues. When performed in protein dynamics the eigenvectors shows the direction and magnitude of motion of the backbone and the associated eigenvalues are the frequency or amplitude of the motion. PCA was carried out using ProDy (Bakan et al., 2011) python module where the last 20 ns trajectory was used for the analysis.

2.6. MM-PBSA binding free energy calculation

Binding free energies of protein-ligand complexes were calculated using Molecular Mechanics/Poisson–Boltzmann surface area (MM-PBSA) (Vorontsov & Miyashita, 2011) (Bakan et al., 2011) method. g_mmpbsa tool was used for this calculation (Kumari, Kumar, Lynn, & Lynn, 2014). The default parameters of the mentioned tool were used for the calculation (Supplementary 8). 200 snapshots were taken (from every 50th ps trajectory) from last 10 ns of the trajectory (40 ns-50 ns) of each run and these 1400 snapshots were used for the calculation.

2.7. Umbrella sampling simulation

Umbrella sampling simulation is not only used to calculate the binding energy ($\Delta G_{bind}$) of protein-ligand system (derived from the potential of mean force), but it can also give valuable insights into the mechanism of binding. Three independent pulling simulations were performed starting from energy minimized solvated complex structure but with different random number seed to generate different initial velocities (different initial velocities were assigned in NPT equilibration step). Caver program (Chovancova et al., 2012) was used to predict the unbinding pathway and the complexes were rotated along the Z-axis to fit the unbinding pathway. Gromacs 2018.3 software was used for the Umbrella sampling simulation. In each of the three pulling simulations, Vpr was set as the immobile reference and viprinin, viprininD1 and viprininD2 were pulled along the z-axis (with the box size along x, y and z axis are 6.56 Å, 4.362 Å and 14 Å respectively) over the course of 500 ps by using a 600 kJ/mol/nm² force constant. Pull rate was set to 0.005 nm per ps according to the previous studies (Ngo et al., 2016; Tam et al., 2018). During the unbinding process, conformations of the solvated complexes were recorded at every step of ~ 0.2 nm and then used for the US simulations. Number of snapshots varied between different complexes (between 11 to 16 snapshots). However, as the ligand narrowly diffuses in the bound state, an additional conformation was taken so that it results in 0.1 nm spacing between the first three windows (Ngo et al., 2019). After that, the conformations were subjected to 100 ps NPT equilibration prior to 5 ns MD run at 300K temperature. Using the outputs from the Umbrella sampling simulation, the potential of mean force (PMF) was extracted using weighted histogram analysis (WHAM) utilizing gmx wham module. $\Delta G_{bind}$ value was then calculated by subtracting the highest and lowest values of PMF curve. Error of computations was estimated over 100 rounds of Bayesian bootstrapping analysis (Efron, 1979).

3. Results and discussion

3.1. Molecular docking between Vpr and Viprinin, ViprininD1 and ViprininD2

Molecular docking is a reliable method to investigate binding orientation of a ligand molecule on a target protein. In this study Viprinin and its two derivatives were docked three
times at the specified site, each time the GA runs, Population size and RMS cluster tolerance were varied in order to find best possible pose. Pose with highest docking energy from three independent docking simulations were taken for further analysis. Highest Docking energy of Viprinin, ViprininD1 and ViprininD2 are $-6.31\text{ Kcal/mol}$, $-7.7\text{ Kcal/mol}$ and $-6.99\text{ Kcal/mol}$ respectively (Figure 1). Viprinin forms sole hydrogen bond (H-bond) with Ala30 and 12 hydrophobic interactions (Figure 1(a,d)); ViprininD1 do not form any H-bond but it is compensated by the formation of 14 hydrophobic interactions (Figure 1(b,e)); ViprininD2 forms 2 H-bonds with Arg73 and Cys76 and 8 hydrophobic interactions (Figure 1(c,f)). An unfavourable acceptor-acceptor interaction between O26 atom of ViprininD2 and amide bond O atom of Phe69, which can be considered as steric clash but it does not affect the binding orientation. Though in this docking procedure, Vpr is held rigid during docking simulation, it may not be the case in actual scenario, but these docking energies are in well agreement with the experimental result (Ong et al., 2011) where it is shown that ViprininD1 and ViprininD2 are more effective in functional inhibition of Vpr than Viprinin. From visual inspection of protein-ligand complex in Discovery Studio, it is found that ViprininD1 invades the hydrophobic space (formed by the three $\alpha$ helices) of Vpr using its Piperidine (a cyclic Carbamate) ring. This piperidine ring is alone responsible for five pi-Alkyl interactions (mean distance $4.482 \pm 0.283 \text{ Å}$) with Ile60, Ile61, Leu64, Lys27 and Leu39. ViprininD2 also contains a Morpholine group but it is less penetrated in the hydrophobic core of Vpr. It forms three pi-Alkyl interactions (mean distance $4.683 \pm 0.615 \text{ Å}$) with Leu26, Ala30 and Ile61. XLogP3 is the partition coefficient value which measures how hydrophilic and hydrophobic a chemical substance is (Leo et al., 1971). Higher invasiveness of the Piperidine ring is obvious as it has positive XLogP3 value ($+0.8$) (Lipophilic) whereas Morpholine ring has negative XLogP3 value ($-0.9$) (Hydrophilic).

3.2. MD Simulation of Vpr-ligand complexes

Vpr segment 17-77 was used (regular secondary structure elements of Vpr is shown in Supplementary 9) in MD simulation study for all three Vpr-ligand complexes. N-terminal 1-16 and C-terminal 78-96 were discarded for three reasons: (i) Residues lining 17-77 is mainly responsible for the formation of DCAF1-Vpr-UNG2 complex which is the functional construct of G2 cell cycle arrest (Wu et al., 2016) (ii) these two parts are random coil which will increase the computation time during simulation; (iii) the C-terminal 85-96 is responsible for Importin-$\alpha$ mediated transport of viral Pre-integration complex (PIC) (Miyatake et al., 2016) and Viprinin do not inhibit this function. Docking study assumes the receptor as a rigid body; therefore it is necessary to investigate the role of the Vpr backbone flexibility on the binding of these three ligands. In order to enhance conformational sampling, simulations were repeated from same initial docking pose but assigning different initial velocities by changing the random number seed. Conclusions drawn from single simulation may lead to false positive data and the conclusion drawn from multiple shorter runs are more reliable. Average value drawn from five to ten runs tends to give more reproducible data (Ngo et al., 2019). All the MD simulation results presented in this study are drawn from seven independent runs with error.
values. Also multiple-run MD simulation enhances the sampling of the conformational space near the native state of a protein. The average value (calculated from seven independent runs for each frame) of RMSD with standard deviation simply indicates spreading of conformational variability achieved during the simulation of each system starting from same initial structure (Figure 2). For our convenience, Vpr protein bound with Viprinin, ViprininD1 and ViprininD2 are designated as Vpr_V, Vpr_V1 and Vpr_V2 (where these are the abbreviations for backbone atoms only) respectively and these abbreviations are used in the paper for analysis purposes. As can be seen from Figure 2, RMSD of Vpr backbone somewhat stabilizes around 20-30 ns in all the simulations but at a higher value (> 0.3 nm). High RMSD value indicates large structural deviations from initial starting structure. Average RMSD value of Vpr_V1 at plateau phase (20 ns-50 ns) is much higher (0.5 ± 0.09 nm) than Vpr, Vpr_V and Vpr_V2 indicating marked structural deviations resulting from the said ligand. RMSD of Vpr, Vpr_V and Vpr_V2 shows similar value throughout the plateau phase (20 ns-50 ns) indicating absence of marked structural deviations caused from the ligands (Figure 2(d)). Three dimensional NMR structure of Vpr was solved in 10-30% CD$_3$CN solution which being a hydrophobic solution stabilizes the alpha helices. In aqueous solution and physiological pH, Vpr does not stabilize and tends to unfold to form aggregates (Morellet et al., 2003). As the simulation of the free Vpr and complexes were carried out in aqueous environment, large backbone deviations are inevitable but, in this study results are drawn from seven independent runs and comparative analysis shows Vpr_V1 exhibits large deviations from Vpr suggesting the effect of ViprininD1 in the protein.

Two out of seven runs of Vpr shows partial unfolding of helix 2 as the simulation progresses and the secondary structure analysis also suggest this fact (Supplementary 3). Another striking feature is the bending of α-helix 3 at the junction of Leu68 in these two runs (run 1 and run 5) which is introduced very early in the simulation and continued till end (Supplementary 4). This bend may have arisen due to the destabilizing effect of abrupt termination of C-terminal helix region (Padhi et al., 2013). This random coil may be responsible for the stabilization of α-helix 3 as evident from the ensemble of 10 NMR structures of Vpr52-96 where we can see that due to the presence of terminal random coil region (78-96), no bend is seen (Bourbigot et al., 2005) (Supplementary 12). Another ensemble of 20 NMR structures of Vpr59-86 shows similar kind of bending at the C-terminal Arg77-Gln86 region, Arg77 being the point of bending (Yao et al., 1998) but due to the presence of some random coil residues (78-86) no huge bend is seen in core helix region (Supplementary 13). This indicates the importance of terminal amino acids for stabilization of α-helices. Vpr, Vpr_V, Vpr_V1 and Vpr_V2 show a common characteristic: transformation of Cys76 and Arg77 into coil from α-helix in all seven runs. This is again due to absence stabilization effects of C-terminal 78-96 segment.

From Figure 2, it can be seen that RMSD values of Vpr backbones in all complexes reaches a plateau after 10 ns except Vpr_V1 which takes near 20 ns to reach plateau phase. Vpr_V1 shows high RMSD value of 0.5 ± 0.09 nm at plateau phase whereas Vpr, Vpr_V and Vpr_V2 shows RMSD values of 0.36 ± 0.07 nm, 0.37 ± 0.09 nm and 0.36 ± 0.04 nm respectively indicating similar convergence of these three trajectories. Detailed analysis of root-mean-square-fluctuation

Figure 2. (a, b and c) are the RMSD data with error values of the backbone Vpr bound with Viprinin, Vpr bound with ViprininD1 and Vpr bound with ViprininD2 from seven independent simulations. In figure d the three mentioned RMSD values are plotted with free Vpr backbone RMSD values for comparative analysis purposes. X axis is the number of frames in all the plots.
RMSF of all the Vpr backbones indicates that N-terminal residues (Glu17-Glu21) (> 0.25 nm) and C-terminal Ile74-Arg77 shows very high fluctuation (> 0.3 nm), (Figure 3) which is partially responsible for higher value of RMSDs. This high fluctuation of terminal residues is obvious due to the absence of stabilizing forces from other residues. The marked structural change in Vpr_V1 is due to full penetration of ViprininD1 into the hydrophobic core during the simulation (Figure 4) which is seen in six out of seven runs. As ViprininD1 has more hydrophobic Piperidine ring than the Morpholine ring of ViprininD2, this penetrative characteristic is favoured during simulation. Asp52-Trp54 of the loop2 region in Vpr_V and Vpr_V2 shows mean RMSF value around 0.20 nm whereas in case of Vpr_V1 these values is greater than 0.25 nm indicating large structural rearrangement in the core hydrophobic region. Vpr_V also shows high flexibility in loop1 region (34-37). From Figure 3d it is evident that except Vpr_V1, Vpr_V and Vpr_V2 displays overall similar flexibility as ligand free Vpr backbone.

From the above analysis it can be inferred that Viprinin, ViprininD1 and ViprininD2 do not produce similar effects in Vpr and they are not very position-restrained in their primary binding groove. To quantify how much ligand position has varied relative to the protein, the overall rotation and translation of the protein is removed via backbone least-squares fitting and the resulting ligand RMSD indicates how well the binding pose has been preserved during the course of the simulation. High RMSD values of the ligands in all the simulations indicate that they are not fixed at the initial starting position (Figure 5(a)–(c)). As can be seen from Figure 5a,b,
after 20 ns of initial equilibration period, Viprinin and ViprininD1 is somewhat stabilized with very high RMSD values (average values of 0.8 ± 0.4 nm and 0.9 ± 0.3 nm respectively); all three trajectories of three ligands are not converged to a similar path thus indicating the absence of any stable specific pocket in the Vpr. From the initial binding pocket it roams around the hydrophobic pocket (in case of Viprinin and ViprininD2) or penetrates the hydrophobic pocket distorting the triple helix structure of the protein. ViprininD2 shows two plateau phase: 13 ns-30 ns (RMSD value of 0.6 ± 0.2), then a slight shift around 30 ns-32 ns and then second plateau phase at 32 ns-50 ns (RMSD value of 0.7 ± 0.2) (Figure 5(c)). From visual inspection of all seven runs of all protein-ligand system, it is seen that during simulation most of the time the ligands are bound with the protein through multiple pi-pi stacked, pi-sigma and pi-Alkyl interactions. We have computed the distribution of the total number of hydrogen bonds over all 7 runs against number of frames (Figure 6) and it is seen that in all three complexes, the ligands are bound by single hydrogen bond in majority of the total time frame. So we can conclude that various hydrophobic interactions have dominated to keep the ligands bind with Vpr. Frames where H-bond number is highest, hydrophobic interactions still high in number (Supplementary 7). This can be explained by the absence of any H-bond donor group in Viprinin and its derivatives. H-bond counts between Vpr and ligand do not exceed three and four in case of Viprinin, ViprininD1 and ViprininD2 which is very low. As we have computed the total number of hydrogen bonds against the number of frames, variations among the runs are not cleared. For an example, in run 4 of Vpr-ViprininD1 simulation, total 2926 frames of two H-bonds were captured whereas all other six runs have below 800 frames of two H-bonds (in a single run there are total 5000 frames). In run 4 and run 7 there are total 118 and 153 frames of three H-bonds and other five runs have below 50 frames.
frames of two H-bonds. Similar scenario can be seen in other protein-ligand complexes where number of frames of H-bonds varies. To address the interplay relationship between dynamics and the binding affinity we have quantified the nonbonded interaction energy between (specifically the short-range Coulombic interaction energy and Lennard-Jones energy) protein and ligand in all seven runs. Although decomposition of interaction energies further into these components is not necessarily real but total interaction energy (Coulomb + Lennard-Jones) can give intuition about the strength of affinity during simulation. Computed interaction energy between Vpr-Viprinin is \[130.2 \pm 16.3 \text{ kJ/mol,} \]
Vpr-ViprininD1 is \[218.5 \pm 14.6 \text{ kJ/mol,} \]
Vpr-ViprininD2 is \[159.6 \pm 14.8 \text{ kJ/mol} \] which resembles the pattern of docking simulation.

3.3. Principal component analysis

To identify overall patterns of motion in Vpr models, we have projected the eigenvectors (on to the C-alpha backbone atoms of representative Vpr protein) of the first three eigenvalues which captures the maximum variances (Figure 7). Multiple-run MD simulation samples the conformational space near the native state of a protein. Therefore, PCA was performed by concatenating the trajectories of seven independent runs (trajectories from 30 ns to 50 ns). This time range was chosen because in all simulation systems and in all runs, the trajectories were somewhat converged to equilibrium at the last 20 ns. We have also shown possible representatives from the seven runs of each simulation system by computing the norm of the deviation of RMSD value (from last 20 ns) from the average value (Supplementary 5) and the projected structural
ensemble of first three PCs of these representatives are shown in Supplementary 11. From Figure 7 we can clearly see that the modes of motion differ from one complex to other. To investigate whether the ligands produced any effect on Vpr backbone, we projected the structural ensemble onto the eigenvectors of first three principal components (Figure 7(c)). Visually, it is very much cumbersome to describe the relative directionality of the vectors between the proteins. To describe vector directionality quantitatively between the proteins, we have computed cosine correlation or overlap value (Supplementary 6) which measures the similarity between two vectors of an inner product space. It is calculated by the cosine of the angle between the two vectors and determines whether two vectors are pointing roughly in same direction (Majumder et al., 2021). Obviously negative sign means opposite vector directionality and vice versa. For an example, from the overlap table (Supplementary 6 A) we can see strong opposite collective motion of three helices between PC1 of Vpr and PC3 of Vpr_V (-0.44), PC2 of Vpr and Vpr_V (-0.55) whereas PC3 of Vpr and PC1 of Vpr_V shows moderate strong same directionality of motion (+0.48). This scenario can also be deducted from Figure 7 qualitatively. To compare overall global collective motion of the three principal components between the proteins, we have calculated root mean squared inner product (RMSIP) which computes quantitative comparison value between the sets of principal components (Majumder et al., 2021). RMSIP values between free Vpr and Vpr with ligands (Vpr-Vpr_V = 0.62, Vpr-Vpr_V1 = 0.60, Vpr-Vpr_V2 = 0.60) indicates that ligands produced substantial effects in the backbone which results in variation in the collective motions.

Figure 9. Average per residue binding energy of (a) Vpr complexed with Viprinin, (b) Vpr complexed with ViprininD1 and (c) Vpr complexed with ViprininD2. Important residues are highlighted.
Global motion between the complexes also varies with following RMSIP values: \( \text{Vpr}_V-\text{Vpr}_V1 = 0.62; \text{Vpr}_V-\text{Vpr}_V2 = 0.60; \text{Vpr}_V1-\text{Vpr}_V2 = 0.60 \). A common feature of all the motions is that the N-terminal (Glu17-Glu21) and C-terminal residues (His71-Arg77) are responsible for the overall high flexibility of Vpr backbone. From Figure 7, it is evident that ViprininD1 and ViprininD2 produces substantial effect on the Vpr backbone during dynamics run which distorts the overall three helix architecture (Figure 7(c,d)). In case of Vpr\_V1, this distortion increases the binding affinity for ViprininD1 by creating an extra space in the Vpr hydrophobic cavity and binding free energy calculation also prove this scenario. In case of Vpr\_V2, this distortion reduces the binding affinity.

### 3.4. MM-PBSA Calculation

To quantify the strength of interaction between Viprinin, ViprininD1 and ViprininD2, MM-PBSA methods were applied by taking 200 snapshots from 40 to 50 ns from all seven runs (thus total \( 200 \times 7 = 1400 \) snapshots for each system) and average value was calculated (Figure 8). We have taken last 10 ns of the trajectory from each seven run because in last 10 ns the trajectories reach a plateau indicating the convergence of the simulation.

Interestingly from MM-PBSA calculation it is found that Viprinin has higher binding energy (-97.2 ± 12.09 kJ/mol) than ViprininD2 (-96.4 ± 14.4 kJ/mol) but the deviation is small and not significant. ViprininD1 has highest binding energy of -140 ± 9.6 kJ/mol and significantly higher than other two. Vpr does not possess any type of stable ligand binding site in the dynamic mode. Average per residue binding energy contribution plot has been computed from the runs (Figure 9). Figure 9c indicates that binding preferences of Viprinin, ViprininD1 and ViprininD2 with Vpr are different. Viprinin and ViprininD2 bind with \( \alpha \)-Helix 3 (59-77) during simulation because amino acid residues in this region have higher binding energy contributions than \( \alpha \)-Helix 1 and \( \alpha \)-Helix 2 (Figure 9(a,c)). As ViprininD1 enters into the hydrophobic pocket during simulation, it make contact with the residues of all three \( \alpha \)-Helices (Figure 9(b)) resulting in higher binding energy.

Apart from DCAF1-Vpr-UNG2 complex formation, it was demonstrated that the Leucine rich Vpr\_61-68 region binds with VprBP and this Vpr-VprBP complex serves as a precursor of G\_2 cell cycle arrest activity (DeHart et al., 2007). As these three compounds make constant binding contacts with the amino acid residues in the region 61-68, VprBP may not be able to bind with Vpr and G\_2 cell cycle arrest activity is lost. Leu26, Le68, Phe69 and Phe72 residues of Vpr have made the major contributions in binding of Viprinin and ViprininD2 but in case of ViprininD1, Leu64 and Leu68 have the highest energy contribution of \(-5.68 ± 2.4\) kJ/mol and \(-5.02 ± 1.7\) kJ/mol respectively (Figure 9(b)). Although the binding site consists of Glu25
and Gln65 and previously it was found that mutation in the Glu25 residue inhibits fumagillin binding (Watanabe et al., 2006), these two residues contribute very less in Viprinin, ViprininD1 and ViprininD2 binding. In particular, Leu and Ile residues in vicinity of above mentioned two residues play an important role in ligand binding.

3.5. Umbrella sampling simulation

To investigate the binding mechanism of Viprinin, ViprininD1 and ViprininD2, three independent Umbrella pulling simulation has been run starting from same NPT equilibrated complex structure for each protein-ligand system. The potential of mean force (PMF) was extracted from US simulations and binding free energy ($\Delta G$) was computed by taking the difference between the highest and lowest values from PMF graph (Figure 10). Computed average $\Delta G$ values of Viprinin, ViprininD1 and ViprininD2 are $-10.5 \pm 3.7$ kcal/mol, $-13.0 \pm 2.3$ kcal/mol and $-7.8 \pm 1.07$ kcal/mol respectively, indicating similar MM-PBSA binding energy and significance pattern. In all three representative PMF graphs, the free energy starts at zero value, then drops to a prominent minimum and finally increases to a stable value when $\zeta$ reaches 2.0-2.3 nm. This range indicates to the state where the various noncovalent bonds between protein and ligand are broken. Overlaps of each umbrella pulling simulation of the representatives are given in Supplementary 10.

The unbinding process of all the ligands involves multiple hydrophobic interaction breakage. From the MM-PBSA calculation it is found that Leu26, Leu68, Phe69 and Phe72 contribute heavily in the binding of Viprinin and similar scenario is also found during the pulling process. At starting conformation (frame 0), a pi-Anion bond with Glu29, Alkyl interaction with Cys76, hydrogen bond with Arg73 and multiple Pi-Pi and Pi-Alkyl interaction with Phe72 hold Viprinin. This pi-Alkyl bond is formed between the pi electron cloud of benzene ring of Phe72 side chain and C15 alkyl group of Viprinin. Cys76 also make pi-sulfur interaction by its sulphur group with the aromatic benzene ring of Viprinin. Then all these various interactions break and in the next conformation (frame 77) a single Pi-Alkyl bond holds the ligand. In the next and final stage of unbinding (frame 140), Viprinin breaks a sole Pi-sulfur interaction with Cys76 and then completely freed from protein (Figure 11(a)). PMF curve of Viprinin (Figure 10(a)) shows two prominent and two very small energy minima and the snapshots used in Figure 11a are taken from these three energy minima. The two minima indicates that after unbinding from initial primary pocket, Viprinin transiently binds with residues from another secondary pocket along the path of reaction coordinate.

A dip in the PMF profile of ViprininD1 (Figure 10(b)) around 1 nm reflects a lack in sampling but it does not affect the $\Delta G$ value. At the starting conformation, ViprininD1 is confined in the hydrophobic group of Vpr through multiple Pi-Alkyl, Pi-Anion, Amide-Pi stacked interactions from Leu26,
Lys27, Ile60, Leu68, Gln65. Then in the next step of dissociation process (frame 240), Leu26 holds ViprininD1 through a single Alkyl interaction. Before complete dissociation, two Alkyl interactions (with Cys76 and Arg73) are broken.

ViprininD2 shows similar unbinding mode as Viprinin and ViprininD1. During the unbinding process Arg73 makes the sole H-bond with the ligand Phe72 and Cys76 hold ViprininD2 through three pi-Alkyl bonds before complete detachment (Figure 11(c)).

4. Conclusion

In this study the binding mechanism of Viprinin and its two highly potent derivatives with Vpr is thoroughly explored through molecular docking, molecular dynamics simulation and Umbrella sampling simulation. Viprinin and its derivatives are not very much fixed at the initial binding pocket of Vpr but they are bound to protein through various hydrophobic interactions. Principal component analysis reveals major dynamic modes of Vpr backbone complexed with Viprinin and its two derivatives. MM-PBSA and Umbrella sampling calculation indicates that ViprininD1 have significantly higher binding energy than Viprinin and ViprininD2. PCA analysis indicates that Helix 3 of Vpr show major motions resulting from the ligand dynamics. This study may create a future direction for possible usage of Viprinin and its derivative as drug molecule against AIDS pathogenesis. As this study is conducted using single Vpr sequence, it is still not understood if the high sequence variation in different HIV-1 subtypes will have any effect in the binding mechanism of Viprinin and its derivatives.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by FRPDF grant of Presidency University and Project grant from West Bengal Department of Higher Education, Science & Technology and Biotechnology (243(Sanc.)/ST/P/S&T/9G-60/2017). The authors acknowledge Joyeeta Datta and Dwaiypayan Chaudhuri for their help to prepare the manuscript.

ORCID

Kalyan Giri http://orcid.org/0000-0001-5681-2612

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-levelparallelism from laptops to supercomputers. SoftwareX, 1-2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001
Bakan, A., Meireles, L. M., & Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics (Oxford, England), 27(11), 1575–1577. https://doi.org/10.1093/bioinformatics/btr168
Barnitz, R. A., Chaing-Delalande, B., Bolton, D. L., & Lenardo, M. J. (2011). Exposed Hydrophobic Residues in Human Immunodeficiency Virus Type 1 Vpr Helix-1 Are Important for Cell Cycle Arrest and Cell Death. PLoS One, 6(9), e24924. https://doi.org/10.1371/journal.pone.0024924
Belzile, J.-P., Abrahamyan, L. G., Gérard, F. C. A., Rougeau, N., & Cohen, É. A. (2010). Formation of mobile chromatin-associated nuclear focus containing HIV-1 Vpr and VPRBP is critical for the induction of G2 cell cycle arrest. PLoS Pathog, 6(9), e1001080. https://doi.org/10.1371/journal.ppat.1001080
Belzile, J.-P., Richard, J., Rougeau, N., Xiao, Y., & Cohen, E. A. (2010). HIV-1 Vpr induces the K48-linked polyubiquitination and pro tease degradation of target cellular proteins to activate ATR and promote G2 arrest. J Virol, 84(7), 3230–3233. https://doi.org/10.1128/JVI.02590-09
Bourbigot, S., Beltz, H., Denis, J., Morellet, N., Roques, B. P., Mély, Y., & Bouaziz, S. (2005). The C-terminal domain of the HIV-1 regulatory protein Vpr adopts an antiparallel dimeric structure in solution via its leucine-zipper-like domain. The Biochemical Journal, 387(Pt 2), 333–341. https://doi.org/10.1042/BJ20041759
Chen, M., Elder, R. T., Yu, M., O’Gorman, M. G., Selig, L., Benarous, R., Yamamoto, A., & Zhao, Y. (1999). Mutational analysis of Vpr-induced G2 arrest, nuclear localization, and cell death in fission yeast. Journal of Virology, 73(4), 3236–3245. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10074177 https://doi.org/10.1128/JVI.73.4.3236-3245.1999
Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., & Damborsky, J. (2012). CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. PLoS Computational Biology, 8(10), e1002708. https://doi.org/10.1371/journal.pcbi.1002708
DeHart, J. L., Zimmerman, E. S., Ardon, O., Monteiro-Filho, C. M. R., Arganaraz, E. R., & Planelles, V. (2007). HIV-1 Vpr activates the G2 checkpoint through manipulation of the ubiquitin proteasome system. Virology Journal, 4(1), 57. https://doi.org/10.1186/1743-422X-4-57
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1–26. https://doi.org/10.1214/aos/1176345522
Esmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. The Journal of Chemical Physics, 103(19), 8577–8593. https://doi.org/10.1063/1.470117
González, M. E. (2017). The HIV-1 Vpr Protein: A multifaceted target for therapeutic intervention. International Journal of Molecular Sciences, 18(1), 126. https://doi.org/10.3390/ijms18010126
Hagiwara, K., Ishii, H., Murakami, T., Takeshima, S.-N., Chutwiitoonchai, N., Kodama, E. N., Kawaji, K., Kondoh, Y., Honda, K., Osaka, H., Tsunetsugu-Yokota, Y., Suzuki, M., & Aida, Y. (2015). Synthesis of a Vpr-binding derivative for use as a novel HIV-1 Inhibitor. PLoS One, 10(12), e0145573. https://doi.org/10.1371/journal.pone.0145573
Hayward, S., & Groot, B. L. (2008). Normal modes and essential dynamics. Methods in Molecular Biology (Clifton, N.J.), 443, 89–106. https://doi.org/10.1385/1-59745-177-2:5
Huang, J., & Mackereil, A. D. (2013). CHARMm36 all-atom additive protein force field: Validation based on comparison to NMR data. J Comput Chem, 34(25), 2135–2145. https://doi.org/10.1002/jcc.23354
Jacotot, E., Ferri, K. F., El Hamel, C., Brenner, C., Druillennec, S., Hoebeke, J., Rustin, P., Metvier, D., Lenoir, C., Geuskens, M., Vieira, H. L., Loeffler, M., Belzacq, A. S., Briand, J. P., Zamzami, N., Edelman, L., Xie, Z. H., Reed, J. C., Roques, B. P., & Kroemer, G. (2001). Control of Mitochondrial Membrane Permeabilization by Adenine Nucleotide Translocator Interacting with HIV-1 Viral Protein r and Bcl-2. The Journal of Experimental Medicine, 193(4), 509–520. https://doi.org/10.1084/jem.193.4.509
Kamata, M., Nitahara-Kashahara, Y., Miyamoto, Y., Yoneda, Y., & Aida, Y. (2005). Importin-α alpha promotes passage through the nuclear pore complex of human immunodeficiency virus type 1 Vpr. Journal of Virology, 79(6), 3557–3564. https://doi.org/10.1128/JVI.79.6.3557-3564.2005
Kamata, M., Wu, R. P., An, D. S., Saxe, J. P., Damoiseaux, R., Phelps, M. E., Huang, J., & Chen, I. S. Y. (2006). Cell-based chemical genetic screen identifies damnacanthal as an inhibitor of HIV-1 Vpr induced cell
