Comparative study of the unbinding process of some HTLV-1 protease inhibitors using Unbiased Molecular Dynamics simulation

Fereshteh Noroozi Tiyoula[1], Hassan Aryapour*[1], Mostafa Javaheri Moghadam[2]

[1]Department of Biology, Faculty of Science, Golestan University, Gorgan, Iran

[2]Department of Chemistry, University of New Brunswick, 30 Dineen Drive, Fredericton, New Brunswick E3B 5A3, Canada

These authors contributed equally to this work.

*Corresponding author. Department of Biology, Faculty of Science, Golestan University, Gorgan, Iran. Tel No: +98-17-32254161; Fax No. +98-17-32245964; E-mail address: hassan.aryapour@gmail.com

Abstract

The HTLV-1 protease is one of the major antiviral targets to overwhelm this virus. Several research groups have been developing protease inhibitors over the years, but none has been successful. In this regard, the development of new HTLV-1 protease inhibitors based on fixing the defects of previous inhibitors will overcome the absence of curative treatment for this oncovirus. Thus, we decided to study the unbinding pathways of the most potent (compound 10, Ki = 15 nM) and one of the weakest (compound 9, Ki = 7900 nM) protease inhibitors, which are very structurally similar, with the PDB IDs: 4YDG, 4YDF, using the Supervised Molecular Dynamics (SuMD) method. In this project, we had various short and long-time-scale simulations, that in total, we could have 12 successful unbindings (a total of 14.8 μs) for the two compounds in both mp forms. This comparative study measured all the essential factors simultaneously in two different inhibitors, which improved our results. This study revealed that Asp32 or Asp32′ in the two forms of mp state similarly exert super power effects on maintaining both potent and weak inhibitors in the binding pocket of HTLV-1 protease . In parallel with the important impact of these two

residues, in the potent inhibitor's unbinding process, His66′ was a great supporter, that was absent in the weak inhibitor's unbinding pathway. In contrast, in the weak inhibitor's unbinding process, Trp98/Trp98′ by pi-pi stacking interactions were unfavorable for the stability of the inhibitor in the binding site. In our opinion, these results will assist in designing more potent and effective inhibitors for the HTLV-1 protease.

Keywords: HTLV-1; ATLL; HAM/TSP; Supervised Molecular Dynamics method (SuMD); Unbinding pathway

Introduction

Human T-cell leukemia virus type 1 (HTLV-1) was discovered in 1980 as the first oncogenic retrovirus in the project "War on Cancer" in the United States [1]. According to the latest information, 5–10 million people are infected with this virus worldwide, and only 0.25–5% of them are affected by Adult T-cell Leukemia/Lymphoma (ATLL) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) [2]. and also HTLV-1-associated ocular diseases. These diseases are known as HTLV-1 uveitis (HU) and ATL-related ocular [3]. Certainly, the reported numbers are not terrible, but there is no standard treatment for all types of diseases [4]. In addition, only a few regions were evaluated, and many unknown infected people could transmit the virus [5]. So, even low-risk areas are in danger because of Global Village. After HTLV-1 discovery, all its components were identified gradually, and its protease was discovered in 1989 [6]. HTLV-1 protease is a homodimer protein containing 125 residues in each subunit, which is one of the A2 family of aspartic proteases, with two key aspartates in the catalytic dyad. This enzyme is essential for viral growth because it cleaves the Gag-Pro-Pol-Env polyprotein, a necessary viral replication component [7]. Since this part is vital for the viral life cycle, it is an interesting target for HTLV-1 demise.

Toward this end, many research groups in different countries succeeded in designing and synthesizing various compounds with inhibitory effects in the micromolar

to nanomolar ranges [8, 9]. Finally, some German scientists considered the structural similarities between HTLV-1 and HTLV-3 (HIV) and determined the X-ray structure of Indinavir complexed with HTLV-1 protease, which is the only AIDS protease drug that has an inhibitory effect on HTLV-1 protease in low micromolar concentration. Unfortunately, this drug failed to be used to eradicate HTLV-1 [10]. After being frustrated with AIDS drugs, this team, in 2015, succeeded in synthesizing ten inhibitors that contain the most potent nonpeptidic inhibitor of HTLV-1 protease up to now [11].

All reported HTLV-1 protease compounds only remain as inhibitors, and we do not have any specialized FDA-approved drug for this virus. It is evident that experimental researches alone are not sufficient, and *in silico* methods, like unbiased molecular dynamics (UMD), are needed to provide valuable information for rational drug design, which is the primary goal of all researchers in this field. MD simulation offers information about the reaction pathways of the ligand-protein complexes, and it has been considered by many research groups over these years and led to effective drug design [12, 13]. Therefore, besides the importance of one particular drug's binding affinity to a target protein in traditional drug design, the binding and unbinding processes and the residence time the compound interacts with the protein in each intermediate state are just as important. So by a complete understanding of the unbinding mechanism, we can uncover the key elements in the protein-ligand complex interactions, ligand flexibility, and solvation effects that are more critical in the rational drug design. The obtained vital information will ultimately appear in a scenario with fully atomistic details [14]. For investigating unbinding pathways of inhibitors, some advanced MD simulation approaches like metadynamics and supervised metadynamics (suMetaD) simulation have been used before [15, 16], and one of the newest MD approaches is the supervised molecular dynamics (SuMD) method. This unbiased method performs simulation in replicas with fixed parameters and prepares information regarding metastable intermediate ligand-bound states. In this regard, the SuMD has been utilized to discover the reaction pathways of various ligands in molecular targets [17]. Therefore, we decided to examine the unbinding

pathways of the most potent and one of the weakest HTLV-1 protease inhibitors retrieved from the last designed compounds using SuMD.

Methods

The X-ray crystallography structures of HTLV-1 protease-ligand complex (PDB IDs: 4YDG, 4YDF [11]) were obtained from the Protein Data Bank. At first, for protein preparation, all protein missing residues and atoms in 4YDG were remodeled and fixed using UCSF Chimera software [18]. Then, for the ligand preparation, according to the practical information [11], the nitrogen of the pyrrolidine ring was protonated in both compounds and parameterized by ACEPYPE using default settings (the GAFF atom type and BCC partial charges) [19]. After preparation of complexes, based on the monoprotonated (mp) form of the catalytic dyad Asp32-Asp32′ in the active site [20], each catalytic Asp was considered separately as an ionization state. Finally, we constructed our systems in GROMACS 2018 [21] using the OPLS all-atom force field [22] and with the TIP3P water model [23]. The considered holoproteins were located in the center of the triclinic box with a distance of 1 nm from each edge. The next step was to provide a 150 mM neutral physiological salt concentration, sodium, and chloride ions. Then all systems were relaxed in energy minimization using the steepest descent algorithm and reached Fmax of less than 1000 $kJ.mol^{-1}.nm^{-1}$. All covalent bonds by Linear Constraint Solver (LINCS) algorithm were constrained to maintain constant bond lengths [24]. The long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method [25], and the cut-off radii for Coulomb and Van der Waals (VdW) short-range interactions were set to 0.9 nm for all systems.

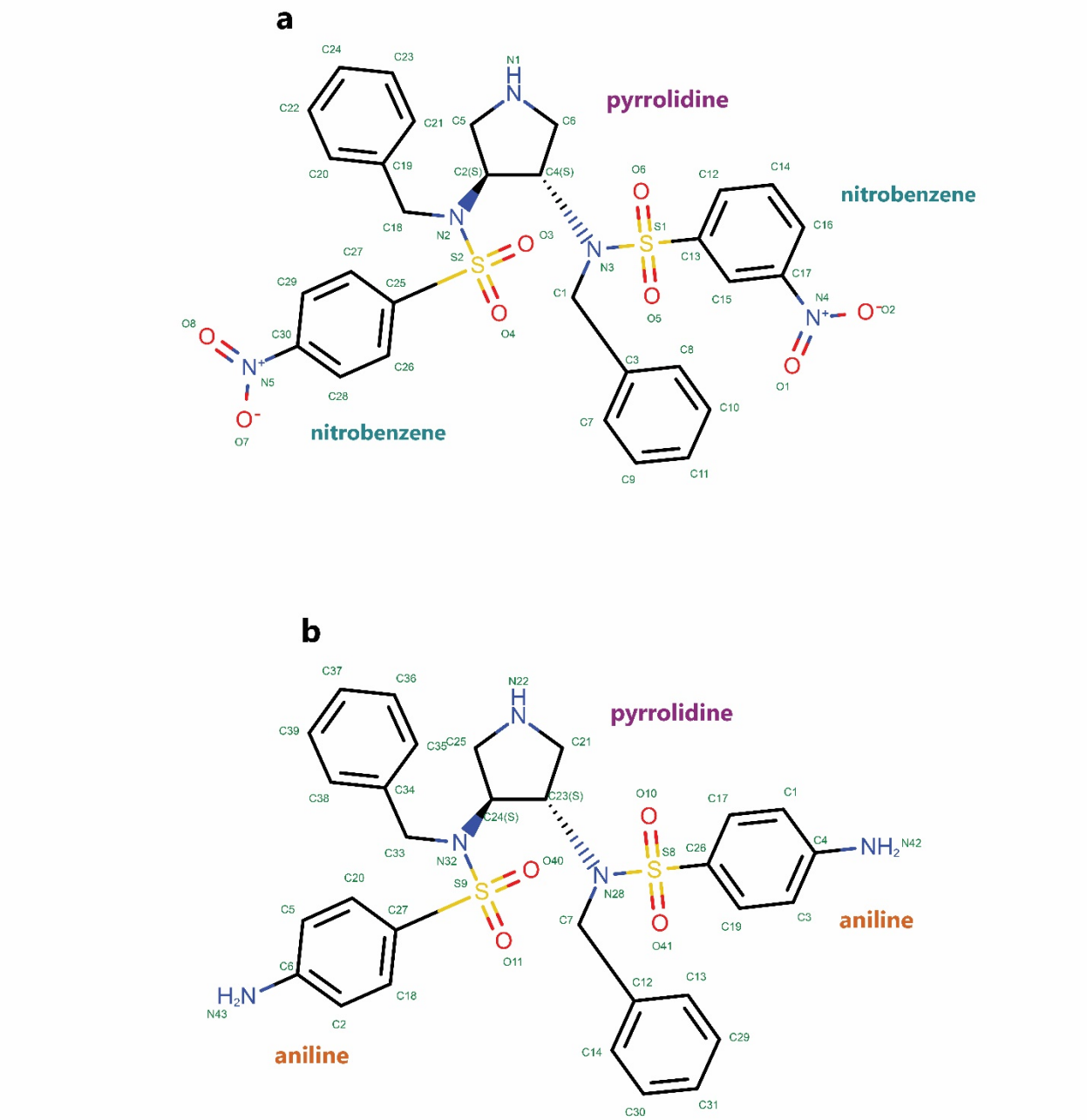
At last, the modified Berendsen (V-rescale) thermostat [26] and Parrinello-Rahman barostat [27] were applied for 100 and 300 ps, for the equilibrations and keep the system in stable environmental conditions (310 K, 1 Bar) and got ready to begin molecular dynamic simulations with a time step of 2 fs and without applying any human or non-

human biasing force or potential. In this regard, to reach complete unbinds, we performed 12 separte series of replicas (three replicas for each type of mp form) with fixed duration times, by the SuMD method with some modifications [28]. During the simulation, the distance of the center of masses of the ligand and selected residues in the binding site was monitored in a fixed time window until full unbind occurs. This method is base on tabu-like supervision algorithm without applying any biasing force or potential. Herein, we set the center of mass (COM) of ligands as a first spot and the COM of the catalytic aspartic acids (Asp32, Asp32′) as second spots and ran all simulations with a time window of 500 ps. After finishing each run, the frame with the longest distance between selected spots was selected automatically to extend the next 500 ps simulation. These processes were continued until complete unbind was obtained, which is equal to a distance of 50 Å between the mentioned spots. Finally, all events in every concatenated trajectory file were investigated carefully with GROMACS utilities for data analysis. To picture the important interactions , we used UCSF chimera and used Daniel's XL Toolbox (v7.3.4) to create plots [29] , and using Matplotlib to show free energy landscapes [30]. The free energy landscapes plots were made base on three variables time, ligand RMSD, and protein RMSD. The ligand and protein RMSD values were selected because they were meaning full and had sharp changes as a function of time during unbindings. Analyzing these plots can reveal the stable states of inhibitors, as well as the residence time of inhibitors in each state over unbinding. Areas that tend to turn blue color indicate that the inhibitor has been present in this area for a longer time.

Results and Discussion

Since the only structural difference between compounds 9 and 10 is in the amino and nitro groups on the benzene ring (Figure 1a, 1b), compound 10 (K_i = 15 nM) is approximately 526 times more potent in complex with HTLV-1 protease [11]. Therefore, a proper understanding of the unbinding pathways of these compounds is vital to unveiling secrets that a minor structural difference can have a dramatic effect on inhibitory effects.

137



138

139 **Figure 1.** The 2D structures of selected compounds were obtained from PDB. **a**, Compound 9, the inhibitor
140 in (PDB ID 4YDF). **b,** Compound 10, the inhibitor in (PDB ID 4YDG).

141

142 For a complete understanding of unbinding mechanisms of these two compounds,
143 it is better to get more familiar with this less-known virus′s protease structure and inhibitors

features at first. This homodimer protein has some particular regions with strategic effects in keeping ligands in the protein's binding pocket that are obtained from analysis of trajectories. The active site region (Leu31-Val39 and Leu31′-Val39′) contains catalytic dyad aspartate residues (Asp32, Asp32′) that are so important in protein-inhibitor interactions. The second essential region is the flaps (Val56-Thr63 and Val56′- Thr63′). The specific residues of Ala59-59′ consider as flap tips in the region of the flaps. Finally, Lateral Loops or 95S loops part of protease (Lys95-Gly102 and Lys95′- Gly102′) are other key regions in this aspartic protease (Figure 2a). For the inhibitors, both compounds have pi-pi self-interactions. With more details, in compound 9, the nitrobenzene ring can form face-to-face pi-pi interaction with the benzene ring (Figure 2b), and in compound 10, aniline ring can form T-shaped edge-to-face pi-pi interaction with the benzene ring (Figure 2c).

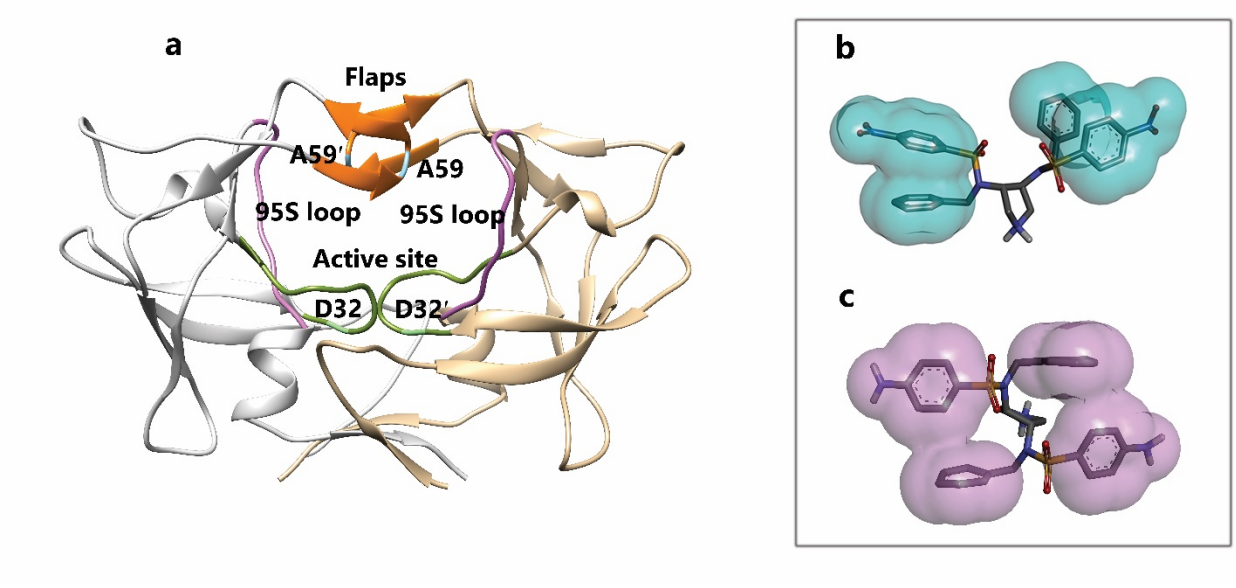


Figure 2. The 3D structure of HTLV-1 protease (PDB ID 4YDF) and self-interactions of inhibitors. **a**, All important domains of HTLV-1 protease: the green area is the active site region, the purple area is Lateral Loops or 95S loops part of protease, the orange area is the flaps region, and the blue area is the flap tips part. **b,** Compound 9. **c,** Compound 10.

As mentioned, one of the essential parts of this protein is the flaps region, which showed high flexibility during our simulations (Figure 3e, 3f, 3g & 3h). So, during our simulations, four modes were observed for the flaps. Herein, we considered two factors to show these modes: the distance between COMs of Ala59 and Ala59′ (d1), and the second one is the distance between COMs of Ala59′ and Asp32′ (d2). The second factor can be even between COMs of Ala59 and Asp 32 due to flaps′ handedness opening. In the close form, the maximum amount of d1 and d2 are 10 and 15 Å, respectively (Figure 3a). In the semi-open form, the maximum amount of d1 and d2 is 14 and 20 Å, respectively (Figure 3b). In the open state, the minimum amount of d1 is 14 Å, and the maximum amounts of d1 and d2 are 20 Å (Figure 3c). In the wide-open form, d1 and d2 must be more than 20 Å (Figure 3d).

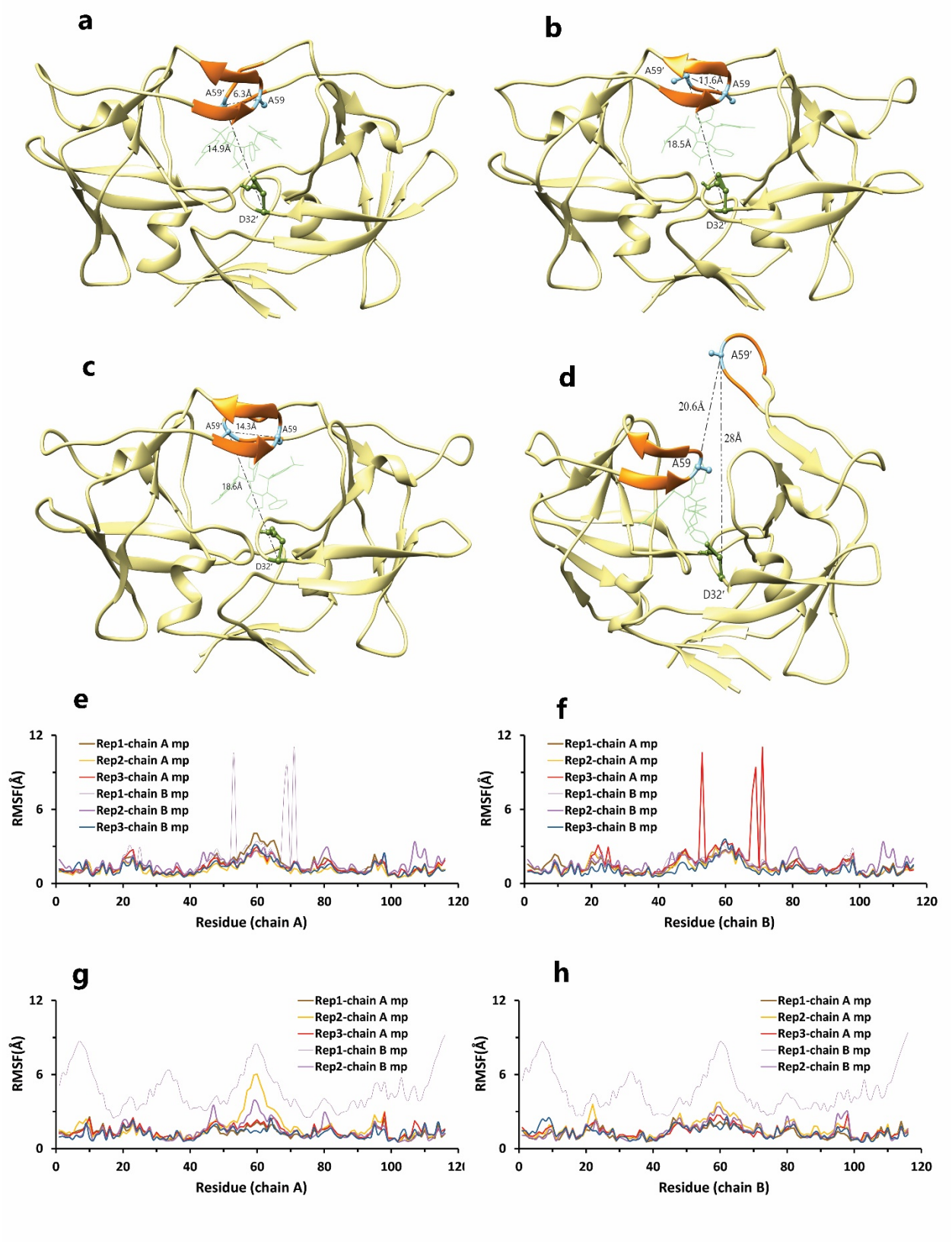


Figure 3. Different modes of the flap. **a**, Close form of the flap. **b,** Semi-open form of the flap. **c,** Open form of the flap. **d**, Wide-open form of the flap. **e, f,** RMSF values of HTLV-1 protease in the 4YDG PDB code, during our simulations. **g, h,** RMSF values of HTLV-1 protease in the 4YDF PDB code, during our simulations.

For our purpose, we had various short and long-time-scale simulations, that in total, we could have 12 successful unbindings (a total of 14.800 μs) for the two compounds from a minimum of 94 ns to a maximum of 4.4 μs in both mp forms. At last, for providing comprehensive information, all the events in each frame of trajectories were investigated carefully, and different analyses were performed on them.

We had two mp forms of the potent compound (AspH32 and AspH32′) like the weak compound. In this regard, in the duration times of 4.4us (Figure 4a) and 260 and 305 ns (Figure 4b), which were in the chain A, Asp32 protonated state, we saw a uniform mechanism to unbind with some important differences that caused a significant difference in one of simulation time. So, in the first state of rep1, 2, and 3 (Figure 4c, 4d, 4e), Asp32′, which had salt bridge interaction with the positive charge of the pyrrolidine ring, play a crucial role in preserving ligand in the binding pocket of protease. This acidic residue is essential because it is located almost at the bottom and center of the binding pocket. This residue considers as a strategic residue due to the positive charge of pyrrolidine. Parallel to that, His66′ by cation–pi interaction with an aniline ring was the second important preserving factor. In addition, Ala59′ in the flap tip by forming H-bond with the atom of O_{10} (Figure 4h) and also Ala35′ by VdW interactions with a benzene ring, Asp36′ by forming H-bond with an aniline fragment in the active site and finally Ile100′ in 95S loop (Figure 4j, 4k, 4l), with VdW interaction, blocked all the exit routes, like the fence (Figure 4f). As mentioned before, His66′ was the second important residue in this state, which was a supporter of Asp32′ to fix the inhibitor in the binding pocket. As time passing, Asp32′ loosed its superpower of preserving, and the inhibitor entered the second intermediate state. In this state, Lys95′ by forming a hydrogen bond with the atom of O_{10}, along with His66′

cation-pi interaction with the benzene ring, was a third essential residue. This residue increased protein-ligand interactions time in the rep1 and was absent in rep2 and rep3 (Figure 4g). According to significant differences in replicas simulation times, the effects of the Lys95′ hydrogen bond (Figure 4i) appear more pronounced. Finally, ligand pi-pi self-interactions, which were observed in the whole time of simulations (Figure 5a, 5b, 5c), weakened all important protein-inhibitor interactions slowly. The critical point was that, over the entire simulation time, flaps positioning impacted ligand's behaviors, so the exit process started when the flaps began to open, and Ala59′ loosed its effect (Figure 5d, 5e, 5f) in the second intermediate state gradually with the help of water mediation (Figure 6a, 6b, 6c).

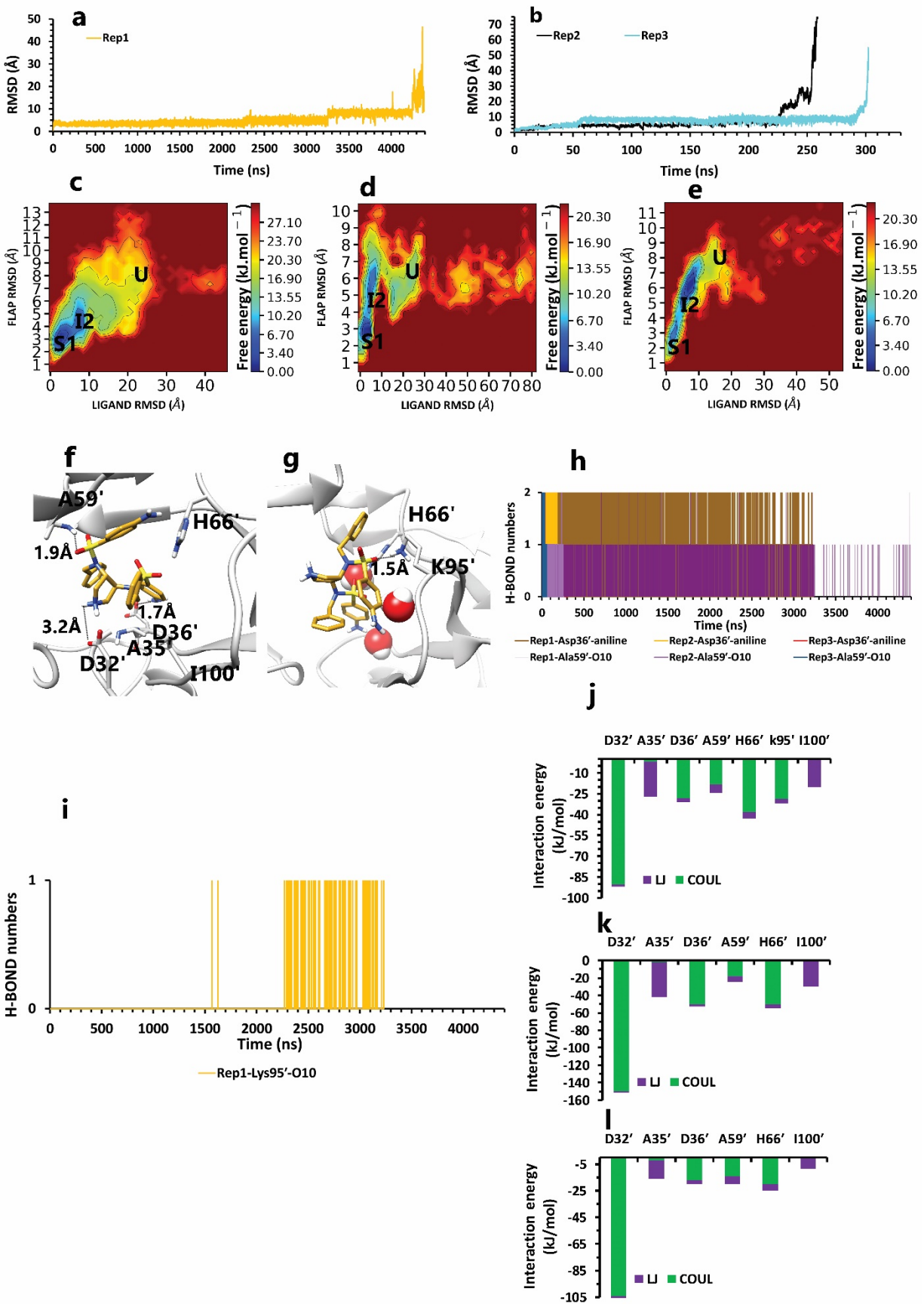


Figure 4. The details of compound 10 unbinding pathways in complex with HTLV-1 protease when Asp32 of chain A was protonated in three replicas. **a**, RMSD value of the ligand from binding pose to complete unbinding in the rep1. **b,** RMSD values of the ligand from binding pose to complete unbinding in the rep2 and rep3. **c, d,** and **e,** The free energy landscape of rep1, 2, and 3 during the unbinding process (state (S), intermediate state (I), unbound (U)), respectively, which was calculated by using "gmx sham". **f,** The interactions between the ligand and important residues in the binding pose of rep1, 2, and 3. **g,** The new interactions between the inhibitor and particular residues in the second intermediate state of rep1. **h,** Hydrogen bond numbers of Asp36′ and Ala59′ with the inhibitor in rep1, 2, and 3 **i,** Hydrogen bond numbers of Lys95′ with the inhibitor in rep1. **j, k,** and **l,** The average of most important interaction energies of the protein-ligand complex in rep1, 2, and 3, respectively.

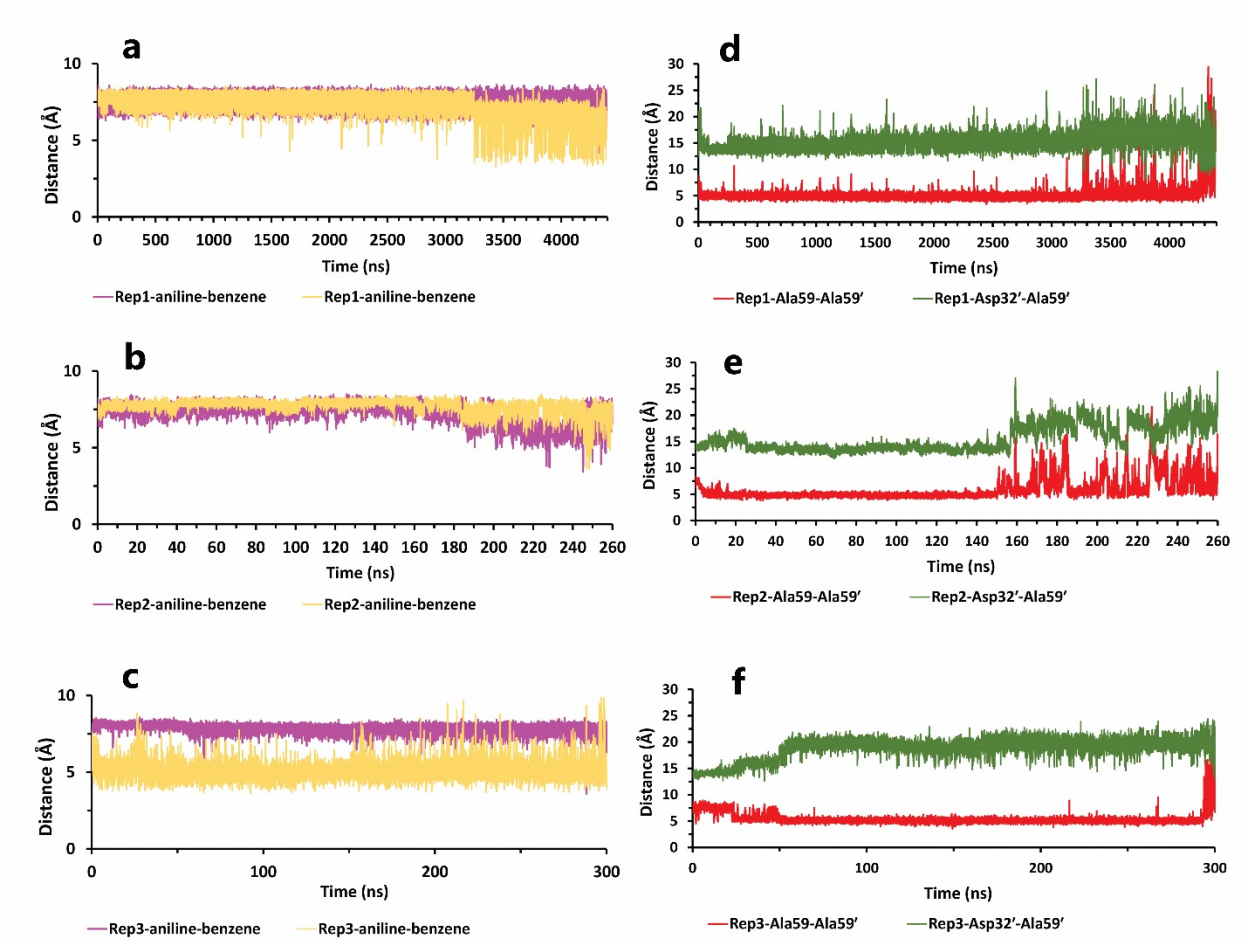
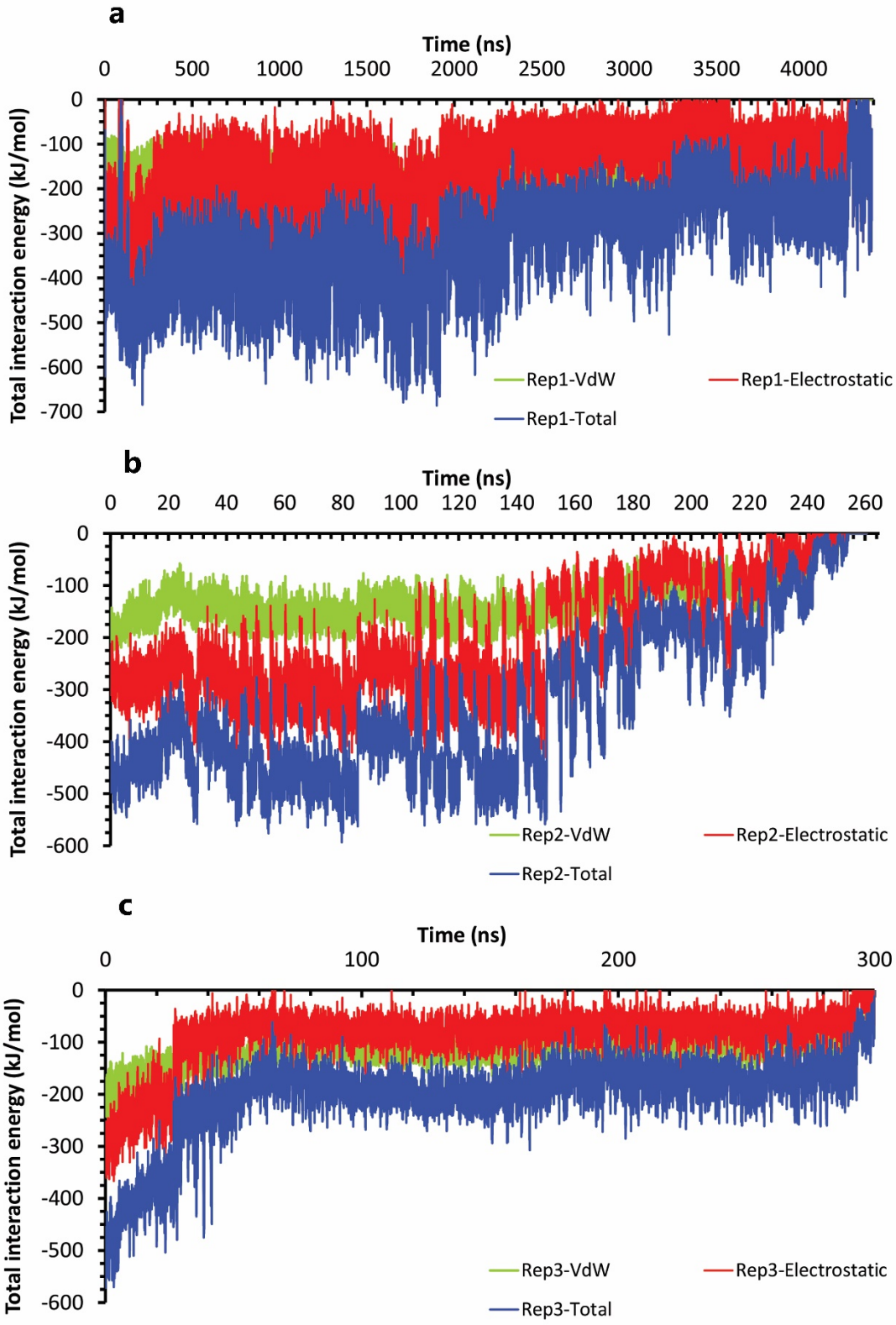


Figure 5. The details of distances between particular parts in compound 10 in complex with HTLV-1 protease, when Asp32 of chain A was protonated in three replicas. **a, b,** and **c,** The distance between COMs of both aniline rings and benzene rings, which were in a position that could form pi-pi self-interactions in

all replicas. These plots prove that during the whole simulation, these fragments were so close together. **d, e** and **f,** The distance between COMs of Ala59 and Ala59′, and also Asp32′ and Ala59′ in all replicas (these plots should be checked along with Figure 3).



232 **Figure 6.** The interaction energies plots of compound 10 in complex with HTLV-1 protease when Asp32
233 of chain A was protonated in three replicas. **a, b,** and **c,** The total VdW, and electrostatic interactions
234 energies of protein-inhibitor complexes in rep1, 2, and 3.

235 Conversely, in the other state of protonation (AspH32′), we saw a uniform pathway
236 that was dissimilar to the previous model with the different lengths of times involving: 94,
237 320, and 790 ns (Figure 7a, 7b). In the first state (S1) of these pathways (Figure 7c, 7d, 7e),
238 Asp32 was so important as expected. Asp36 and Asp36′, Leu57, and Ala59′ are the residues
239 that acted as auxiliary agents (Figure 7j, 7k, 7l) to the pivotal amino acid (Asp32). Details,
240 at first times of simulation along with the salt bridge of Asp32 and pyrrolidine fragment
241 (Figure 7f), both aniline rings had H-bonds with Asp36 and Asp36′ in the active site (Figure
242 7h). Along with these residues, Leu57 and Ala59′ formed a hydrogen bond with an aniline
243 fragment and O_{41} atom of inhibitor, respectively (Figure 7g, 7i). In the following, in the
244 lack of His66 and Lys95 effects, after time passing with the help of pi-pi ligand self-
245 interactions and water molecules effect (Figure 8a, 8b, 8c), active site and flaps' important
246 residues lost their effects, and full unbind was observed between the flaps (Figure 8d, 8e,
247 8f) (Figure 9a, 9b, 9c).

248

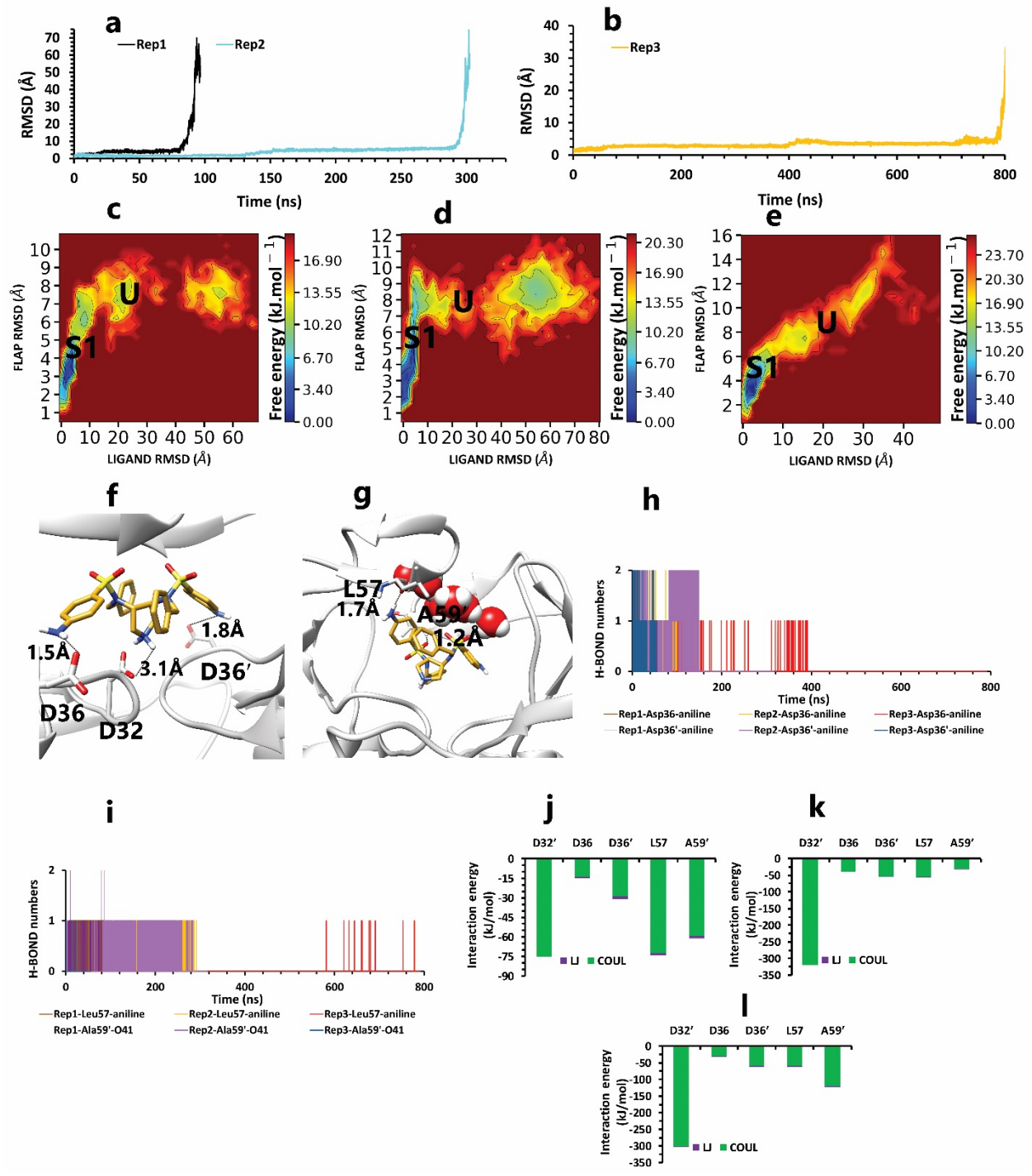


Figure 7. The details of compound 10 unbinding pathways in complex with HTLV-1 protease when Asp32′ of chain B was protonated in three replicas. **a**, RMSD values of the ligand from binding pose to complete unbinding in the rep1 and rep2. **b,** RMSD value of the ligand from binding pose to complete unbinding in the rep3. **c, d,** and **e,** The free energy landscape of rep1, 2, and 3 during the unbinding process (state (S), intermediate state (I), unbound (U)) respectively, which was calculated by using "gmx sham". **f,** The

interactions between the ligand and important residues of the active site in the binding pose of rep1, 2, and 3 **g,** The interactions between the ligand and important residues of flaps region in the binding pose of rep1, 2, and 3. **h,** Hydrogen bond numbers of Asp36 and Asp36′ with aniline fragment in rep1, 2, and 3. **i,** Hydrogen bond numbers of Leu57 and Ala59′ with the inhibitor in rep1, 2, and 3. **j, k** and **l,** The average of most important interaction energies of the protein-ligand complex in rep1, 2, and 3, respectively.

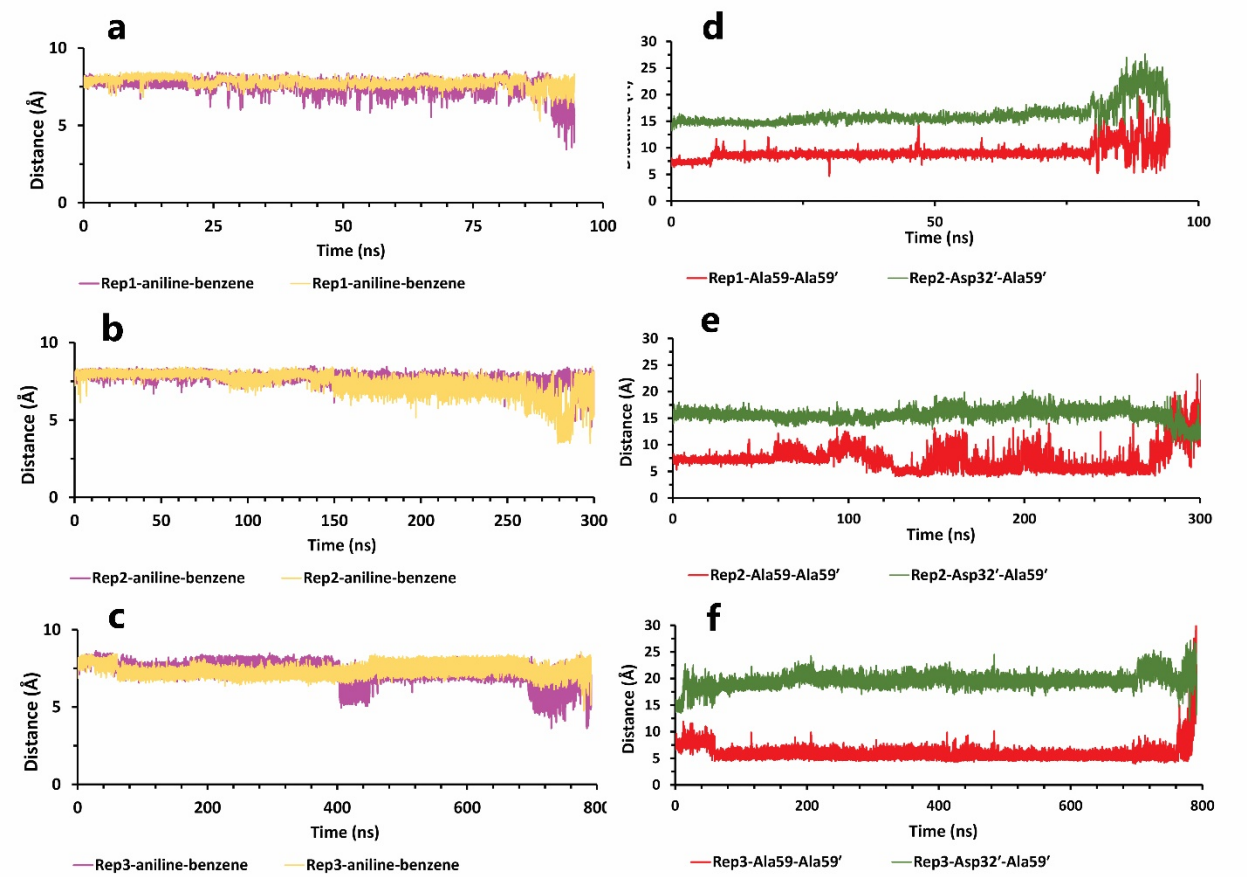


Figure 8. The details of distances between particular parts in compound 10 in complex with HTLV-1 protease, when Asp32 of chain B was protonated, in three replicas. **a, b,** and **c,** The distance between COMs of both aniline rings and benzene rings, which were in a position that could form pi-pi self-interactions in all replicas. These plots prove that during the whole simulation, these fragments were so close together. **d, e,** and **f,** The distance between COMs of Ala59 andAla59′ and also Asp32′ Ala59′ in all replicas (these plots should be checked with Figure 3).

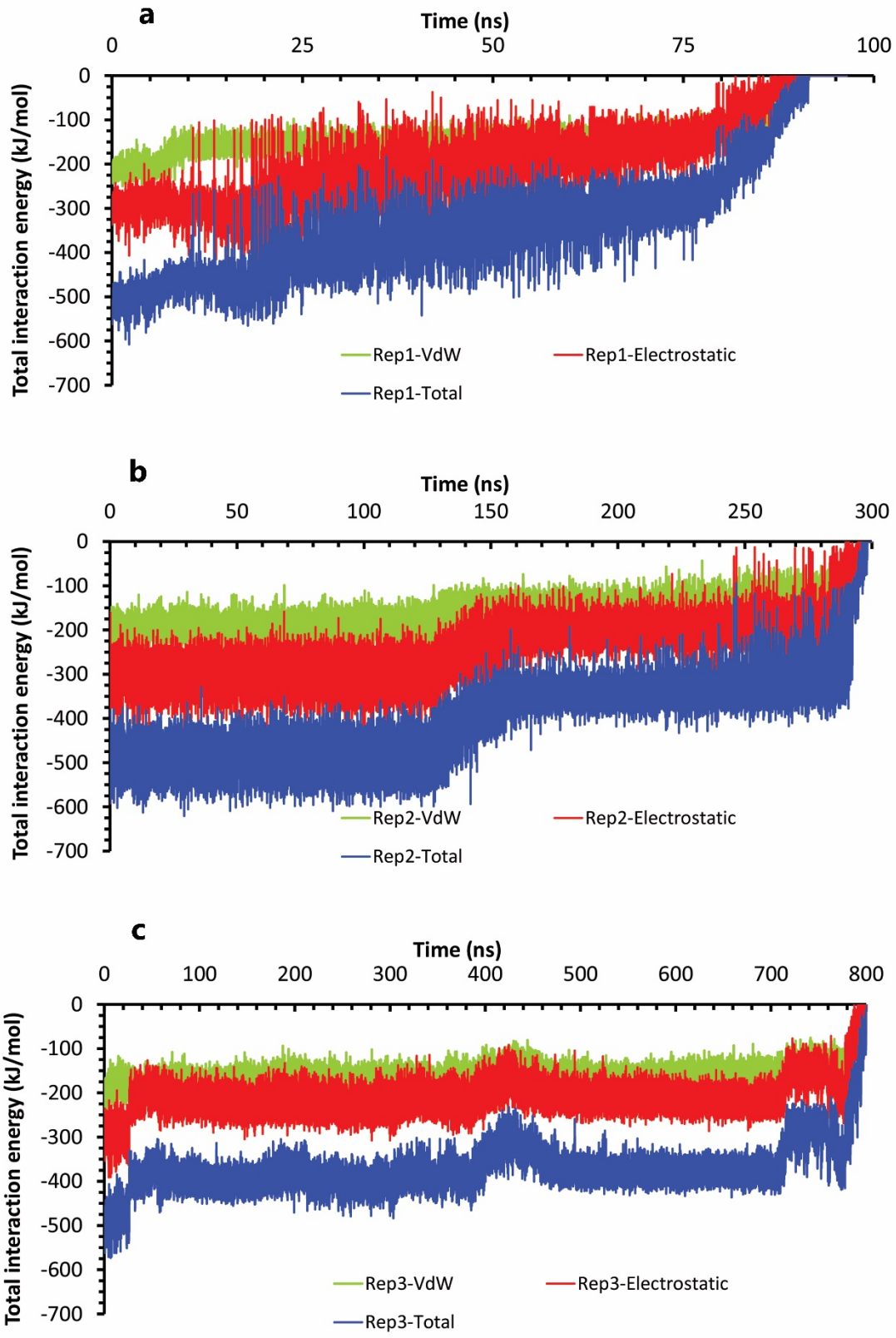


Figure 9. The interaction energies plots of compound 10 in complex with HTLV-1 protease when Asp32 of chain B was protonated in three replicas. **a, b,** and **c,** The total VdW and electrostatic interactions energies of protein-inhibitor complexes in rep1, 2, and 3

We had different unbinding mechanisms for the weaker inhibitor, depending on close and open flaps and mp states. Accordingly, compound 9 was unbound in the duration times of 148 ns (Figure 10a), 3.5 μs and 3 μs (Figure 10b) when Asp 32 of chain A was protonated. In the first state (S1) (Figure 10c) of the rapid unbinding pathway (rep1), a repulsive force occurred between the pyrrolidine ring of ligand and AspH32 of the binding pocket, and because of no attractive interactions in this area, AspH32 forced the ligand to push out. In addition to this interaction, VdW interaction between both nitrobenzene rings and one of the benzene ring of inhibitor and Leu57, Gly58 and Ala59 in the close flap region and pi-pi stacking interaction of Trp98′ and nitrobenzene fragment and also pi-alkyl interaction of Ile100′ with the benzene ring, were other protein-inhibitor important interactions, which were not potent enough to prevent from repulsive interaction effect (Figure 10f, 10k). In the two other long-time simulations, compound 9 was unbound in 3.5 μs in wide-open form of flaps (rep2) and 3 μs in close and semi-open forms of flaps (rep3).

Similarly, Asp32′ was the most important amino acid with its salt bridge and the only common point in both pathways. In the first state of rep2 (S1) (Figure 10d), due to handedness opening, only one of the flaps had forward and backward motions, so Leu57′, Gly58′, Ala59′ by VdW interactions, kept the ligand in exposing to Asp32′. Also, in this state, Trp98 in the lateral loop built up pi-pi stacking interaction with nitrobenzene ring of the ligand and Trp98′ built up pi-pi stacking interaction with benzene ring of another side of inhibitor (Figure 10g, 10l). So even with enough space for the exit, the inhibitor was still in blockage. These important protein-inhibitor interactions were maintained until the effect of the Asp32′ became faded, and other agents, one after another, lost their effect. Unexpectedly, the interesting point was that the complete unbinding process does not occur from the region of the flaps. In the rep3 pathway, that the flaps were close or semi-open

the whole time, from the first state (Figure 10e), not only Asp32′ was important, and Asp36 in a close position to Asp32′ was powerful too (Figure 10m).

On the other hand, during the first two states, Asp36 by forming pi-anion interaction with nitrobenzene fragment, was momentous as a second ligand preserving residue (Figure 10h), which was promoted to the first important factor in the next intermediate state by replacing pi-anion interaction with the salt bridge with pyrrolidine fragment (Figure 10i). From a holistic view, even though Asp32′ was more critical for protein, it was effective until the second intermediate state or until 2 μs, but Asp36 (Figure 10j) was effective until complete unbind. Actually, the ligand in all replicas showed face-to-face pi-pi self-interactions between mentioned fragments that caused weakened important protein-ligand interactions gradually with the help of water mediation effect (Figure 11a, 11b, 11c). Finally, for the flaps behaviors in all replicas, we saw a new opening form for the rep2 as it was opened from chain A (Figure 11e), and for rep1 and rep3 wide opening (Figure 11d, 11f) was not seen until complete unbound (Figure 12a, 12b, 12c).

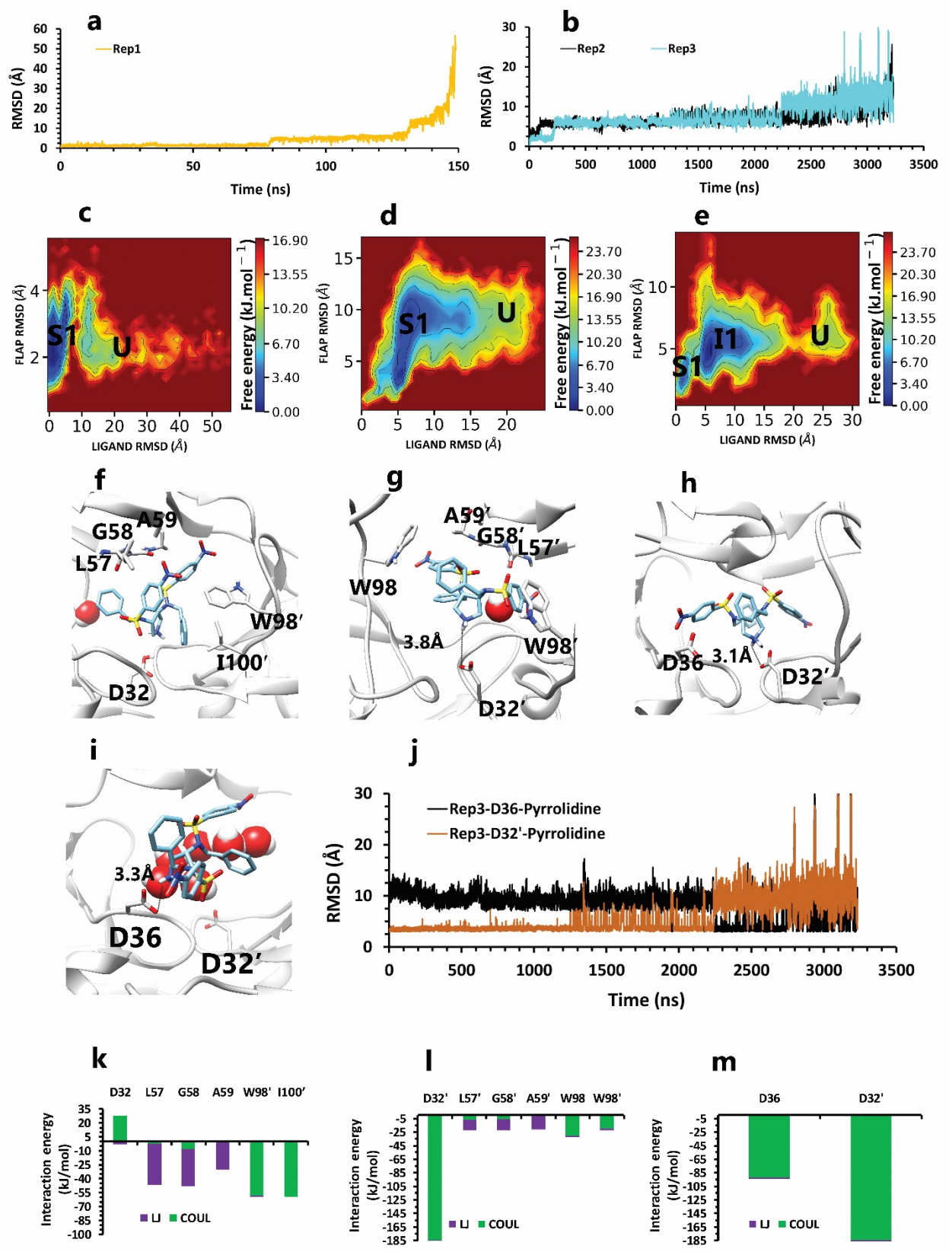


Figure 10. The details of compound 9 unbinding pathways in complex with HTLV-1 protease when Asp32 of chain A was protonated in three replicas. **a**, RMSD value of the ligand from binding pose to complete unbinding in the rep1. **b,** RMSD values of the ligand from binding pose to complete unbinding in the rep2 and rep3. **c, d** and **e,** The free energy landscape of rep1, 2 and 3 during the unbinding process (state (S), intermediate state (I), unbound (U)) respectively, which was calculated by using "gmx sham". **f,** The interactions between the ligand and important residues of the active site in the rep1. **g,** The interactions between the ligand and important residues of the active site in the rep2. **h,** The interactions between the ligand and important active site residues in the binding pose of rep3. **i,** The new interactions between the inhibitor and particular residues in the second intermediate state of rep3. **j,** The distance between COMs of pyrrolidine ring and Asp36 and Asp32′ in rep3, to show after 2us of simulation this fragment get closer to Asp36 and get farther from Asp32′. **k, l** and **3,** The average of most important interaction energies of the protein-ligand complex in rep1, 2 and 3, respectively.

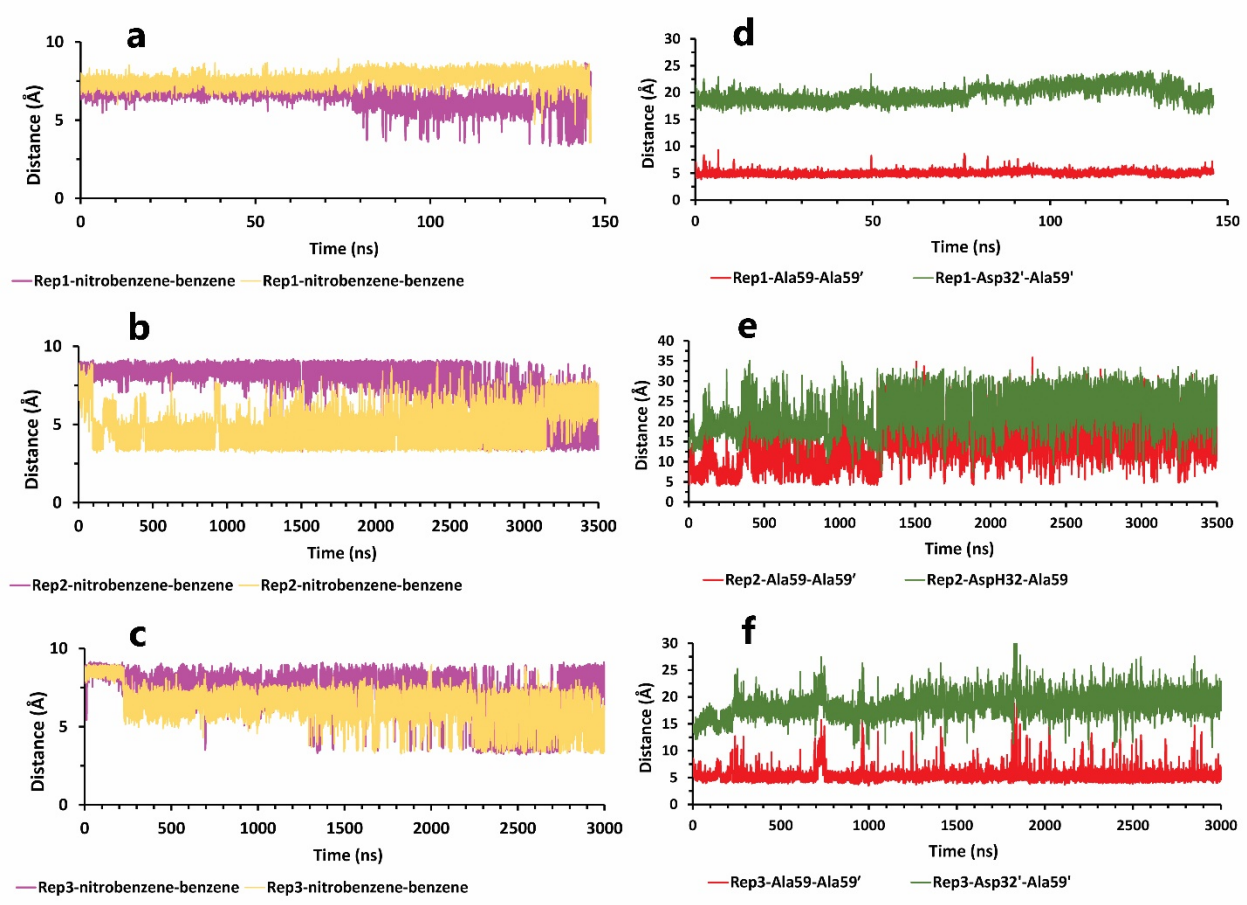


Figure 11. The details of distances between particular parts in compound 9 in complex with HTLV-1 protease, when Asp32 of chain B was protonated, in three replicas. **a, b** and **c,** The distance between COMs

of both aniline rings and benzene rings, which were in a position that could form pi-pi self-interactions in all replicas. These plots prove that during the whole simulation, these fragments were so close together. **d,** The distance between COMs of Ala59 andAla59′ and also Asp32′ Ala59′ in the rep1 (these plots should be checked with Figure 3). **e,** The distance between COMs of Ala59 andAla59′ and also AspH32 Ala59 in the rep2. **f,** The distance between COMs of Ala59 andAla59′ and also Asp32′ Ala59′ in the rep3. **g** and **h,** The total interactions energies of protein-inhibitor complexes in rep1, 2, and 3.

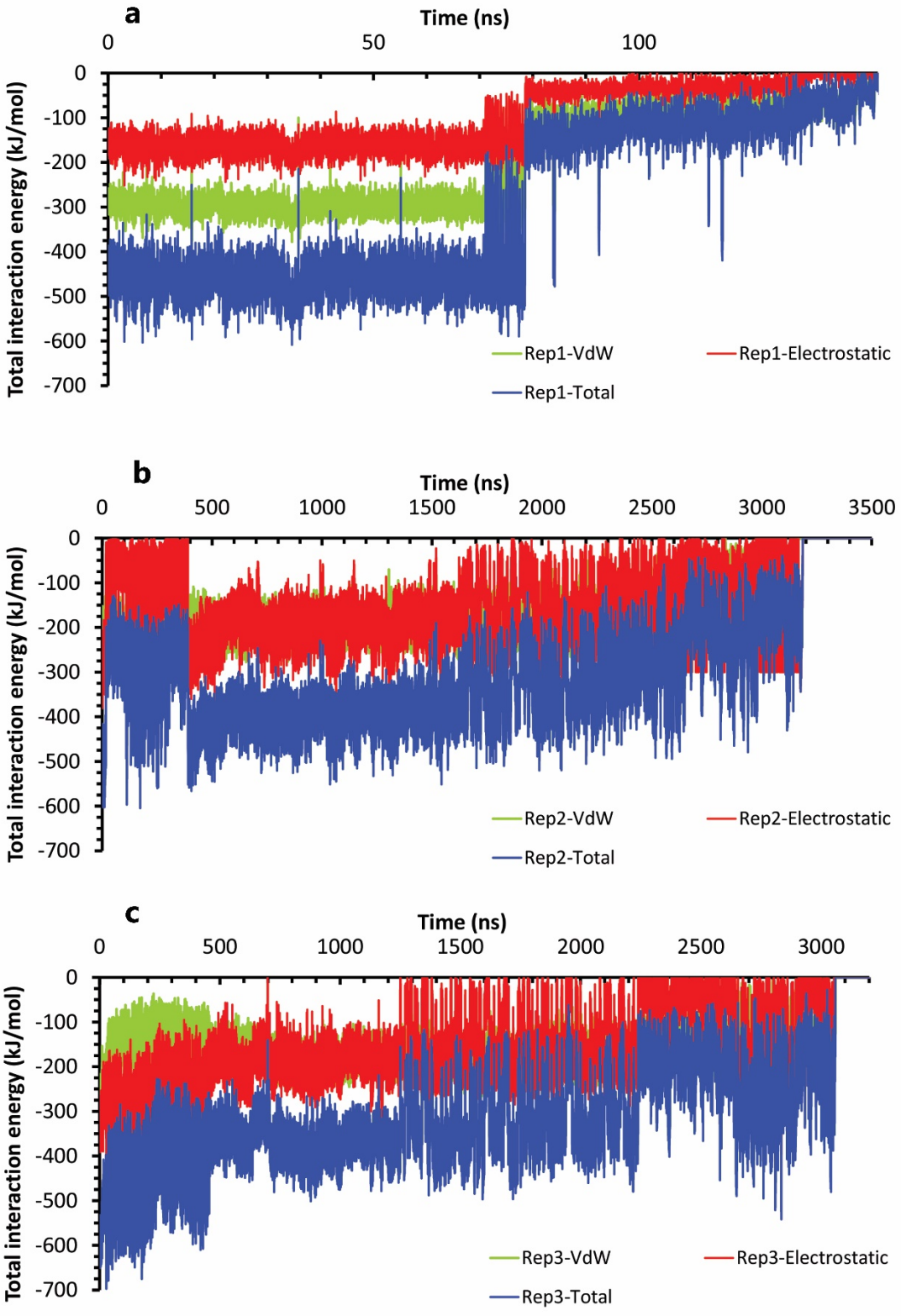


Figure 12. The interaction energies plots of compound 9 in complex with HTLV-1 protease when Asp32 of chain A was protonated in three replicas. **a, b,** and **c,** The total VdW and electrostatic interactions energies of protein-inhibitor complexes in rep1, 2, and 3

On the contrary, when the Asp32 of chain B was protonated, we saw the same mechanism during the 1.2 μs, 410, and 450 ns of simulations (Figure 13a, 13b). In the first state of these replicas (Figure 13c, 13d, 13e), the ligand was surrounded by interactions of some residues in both chains (Figure 13g, 13h, 13i). In more detail, Asp32 had salt bridge interaction with pyrrolidine fragment as a most important interaction. This fragment also had VdW interaction with Gly34. For the other fragments, one of the nitrobenzene rings was in VdW interactions with Leu57′ and Gly58′ in the flaps regions. The benzene rings were in important interactions involving: pi-pi interaction with Trp98 and pi-alkyl interaction with Ile100 in one side, and pi-pi interactions with Trp98′ and pi-alkyl interaction with Ile100′ on the other side (Figure 13f). It may be due to the high number of important factors; it seems that compound 9 is potent, but except Asp32 other agents did not have any significant effect. So, they could not keep the ligand after disappearing of Asp32 effect. Thus, as time passed, ligand self-interactions and water mediation contributed to full unbinding in these three replicas (Figure 14a, 14b, 14c). Ultimately for the flaps effects, in rep1, the flaps showed high motions, and even though the flaps were wide open (Figure 14d), the full unbind did not occur from this region. In rep2 and 3, the inhibitor unbounded between semi-open forms of flaps (Figure 14e, 14f) (Figure 15a, 15b, 15c).

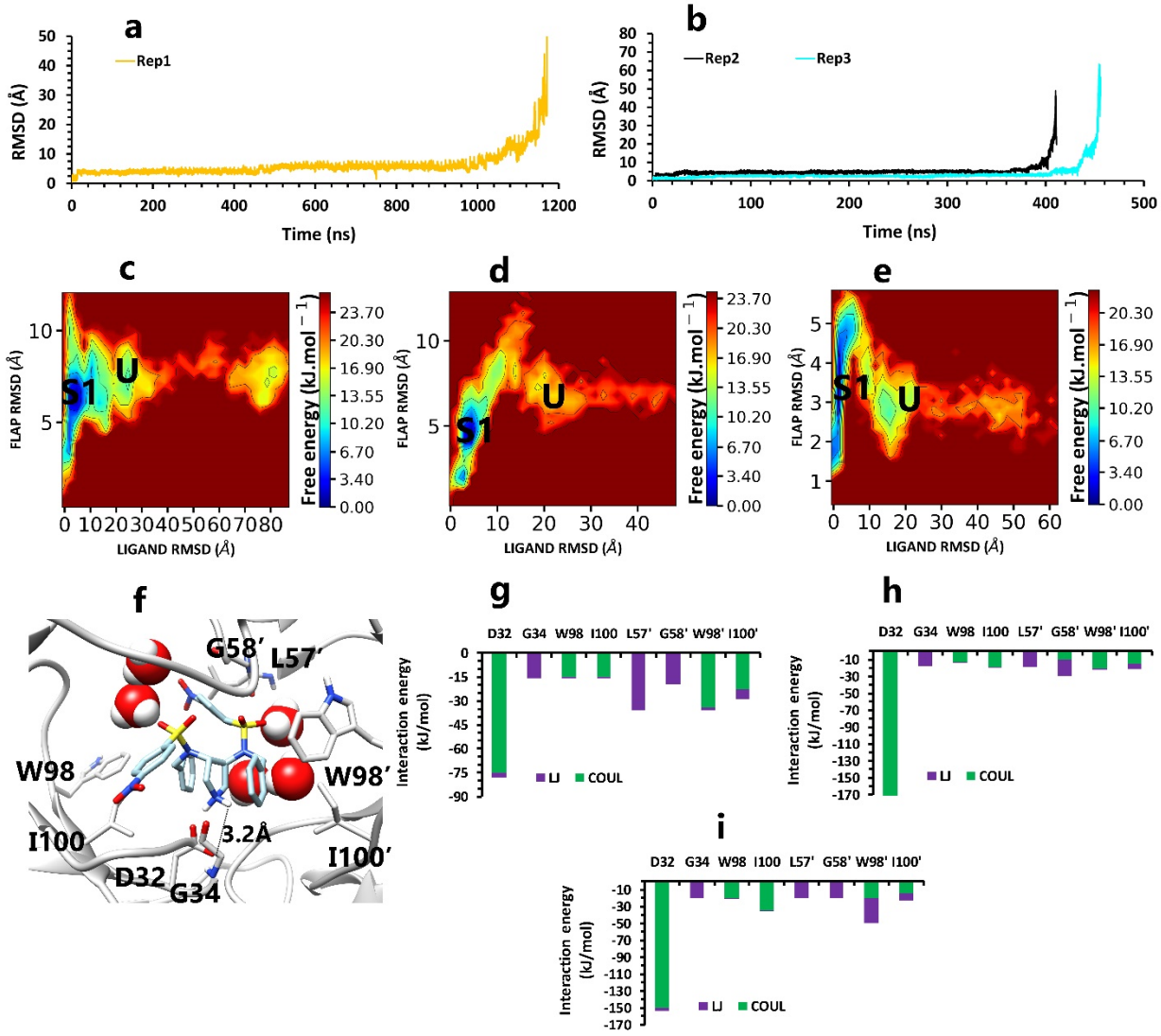


Figure 13. The details of compound 9 unbinding pathways in complex with HTLV-1 protease (state (S), intermediate state (I), unbound (U)) when Asp32 of chain B was protonated, in three replicas. **a**, RMSD value of the ligand from binding pose to complete unbinding in the rep1. **b,** RMSD values of the ligand from binding pose to complete unbinding in the rep2 and rep3. **c, d** and **e,** The free energy landscape of rep1, 2, and 3 during the unbinding process, respectively, was calculated using "gmx sham". **f,** The interactions between the ligand and important active site residues in the rep1, 2, and 3. **g, h,** and **i,** The average of most important interaction energies of the protein-ligand complex in rep1, 2, and 3, respectively.

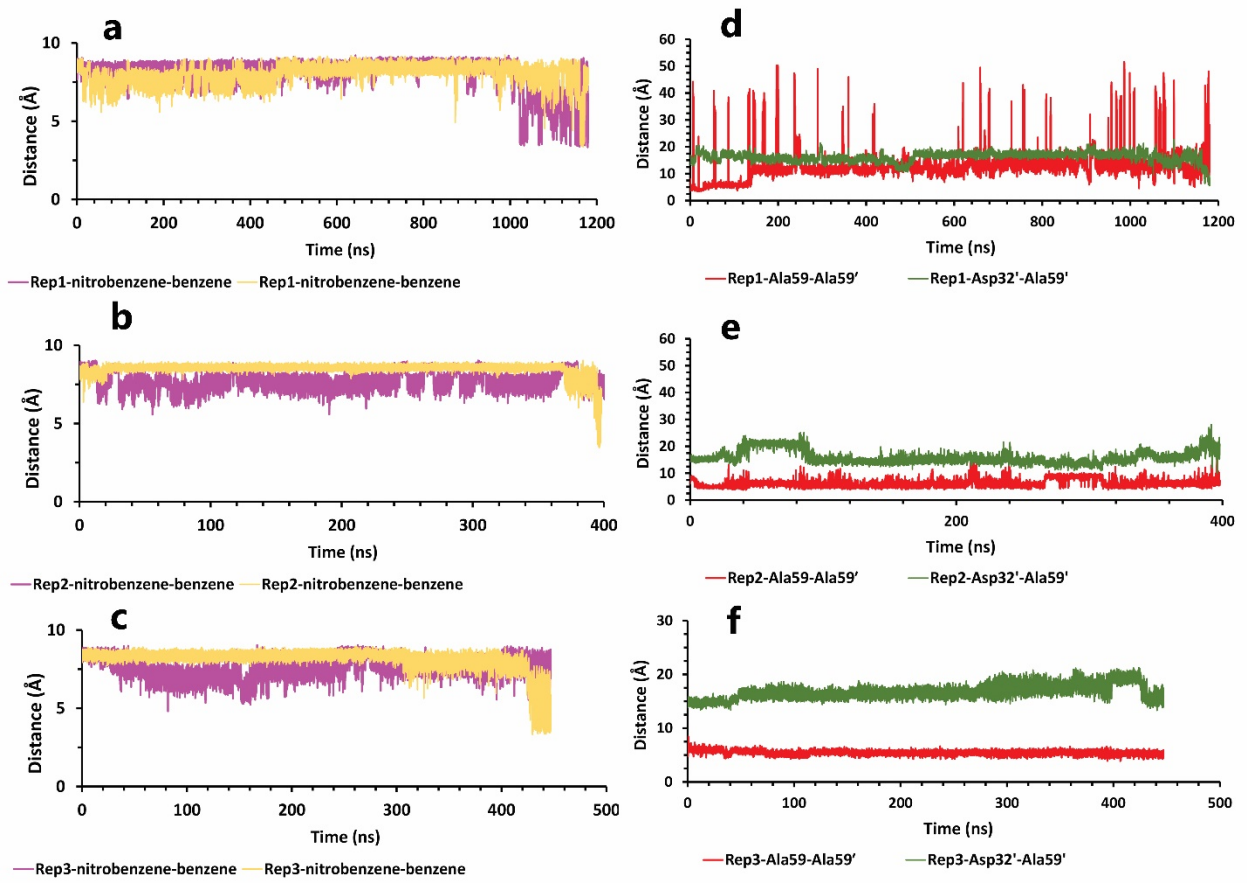
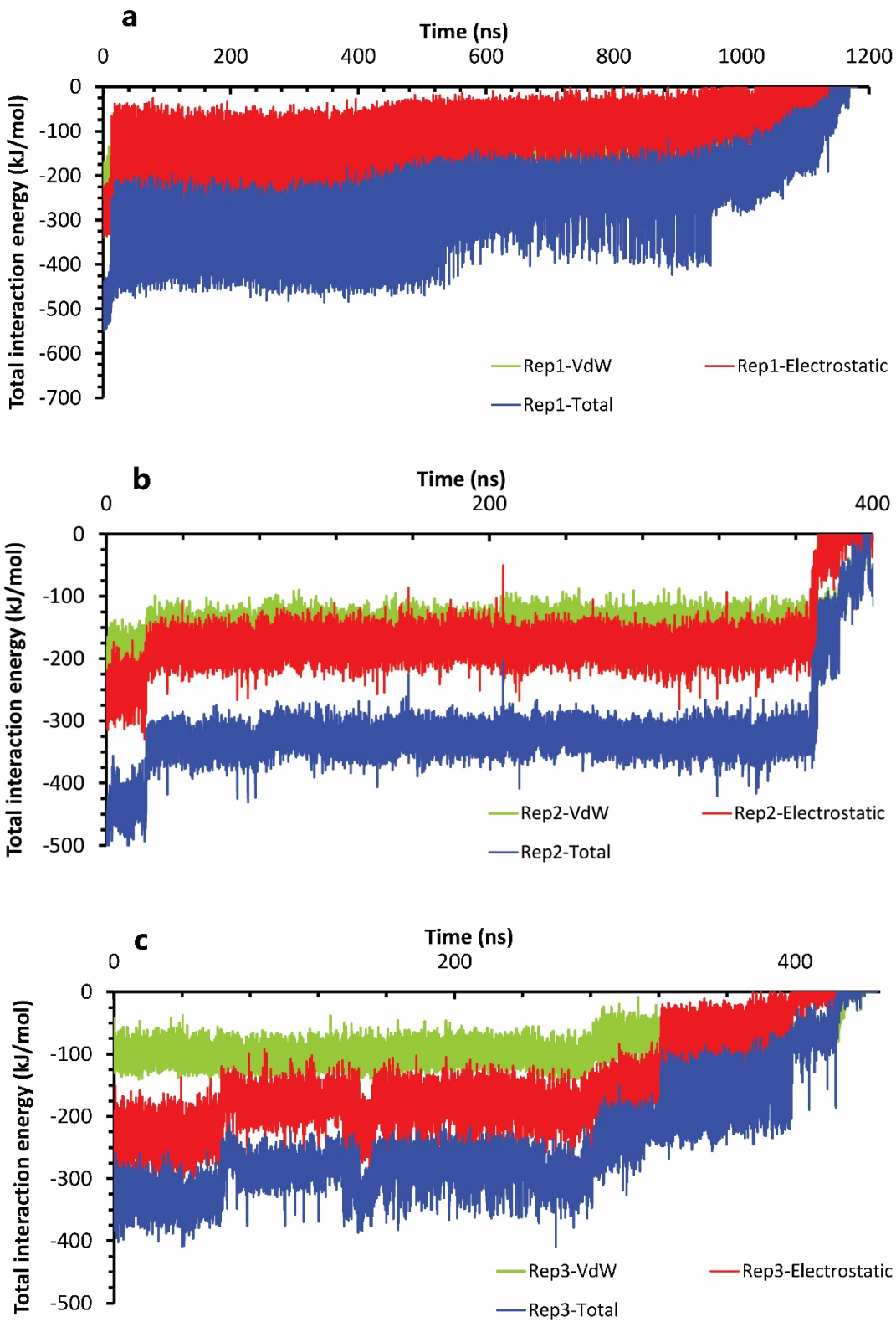


Figure 14. The details of distances between particular parts in compound 9 in complex with HTLV-1 protease when Asp32 of chain B was protonated in three replicas. **a, b** and **c,** The distance between COMs of both aniline rings and benzene rings, which were in a position that could form pi-pi self-interactions in all replicas. These plots prove that during the whole simulation, these fragments were so close together. **d, e** and **f,** The distance between COMs of Ala59 andAla59′ and also Asp32′ Ala59′ in all replicas (these plots should be checked with Figure 3).



372 **Figure 15.** The interaction energies plots of compound 9 in complex with HTLV-1 protease when Asp32
373 of chain B was protonated in three replicas. **a, b,** and **c,** The total VdW and electrostatic interactions energies
374 of protein-inhibitor complexes in rep1, 2, and 3

375

376 ## Conclusion

377 By putting together, the atomistic details of unbinding pathways of selected
378 inhibitors in all replicas, with various times, the importance of Asp32′ in chain A
379 protonation state and Asp32 in chain B protonation state are pretty straightforward. Due to
380 its strategic position, this effective residue could play a critical role in keeping the ligand
381 in the binding pocket for a long time, so the more exposed to Asp32 or Asp32′, the more
382 inhibitory effects. The pyrrolidine fragment was held well by Asp32 or Asp32′ from the
383 native binding pose of the two compounds. Thus, the interactions of other fragments with
384 other residues in different regions of protein caused significant differences.

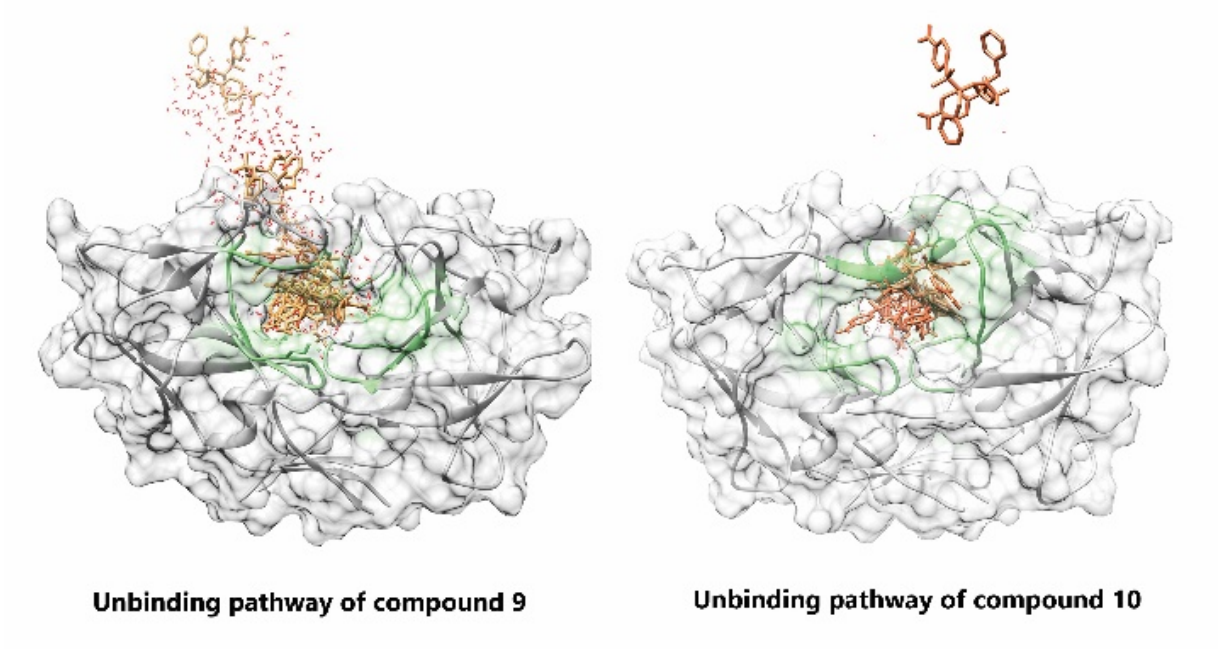
385 Herein, we cannot conclude certainly which state of protonation actually
386 occurs, so with our obtained information for the potent compound in chain A protonation
387 state, His66′ with its cation–pi interaction with an aniline ring of inhibitor was a perfect
388 supporter to Asp32′. This residue's effect was absent in the other form of protonation state
389 and caused a significant difference in simulation time. In the weak inhibitor unbinding
390 pathways, Trp98 and Trp98′ with pi-pi interactions, due to their close position to one of
391 the exit areas were not good supporters for Asp32 or Asp32′, like His66′, His66′, due to its
392 far position from the bottom and center of the binding pocket, could fixed aniline fragment
393 and decreased ligand fluctuations. The two mentioned tryptophan were closer to the
394 important aspartic acids, and there was enough space for ligand fluctuations. For this
395 reason, Asp36 in the active site that was close to the exit area could be a competitor with
396 Asp32′ and was not a good interaction for keeping the ligand in the binding pocket.
397 Similarly, attenuating effect of Trp98/Trp98′ residues in unbinding pathways of the weak
398 inhibitor has a correlation with another research result. These residues' interactions are
399 unfavorable to Indinivar's stability in complex with HTLV-1 protease and result in it being

a weak inhibitor. [31]As we said before, both compounds had self-interactions that caused weakening critical protein-ligand interactions by time passing. These two compounds did not have the same self-interaction type, so in the weak inhibitor, face-to-face pi-pi interactions caused to loosed important pi interactions with the protein, but the potent inhibitor could have formed more important pi interactions with the protein along with self-interactions. Overall, this obtained information is valuable for designing a new generation of inhibitors against this molecular target.

Data and Software Availability

All data and analysis files are uploaded to this address (https://zenodo.org/record/5633143#.YZXlftBBzIU) and are available with DIO:10.5281/zenodo.5633143. Also, in the method section, all used software is listed.

Table of Contents graphic



References

1. A. Vahlne, "A historical reflection on the discovery of human retroviruses," *Retrovirology*, vol. 6, pp. 1–9, 2009.

2. M. E. Haziot *et al.*, "Detection of clinical and neurological signs in apparently asymptomatic HTLV-1 infected carriers: Association with high proviral load," *PLoS Negl. Trop. Dis.*, vol. 13, no. 5, pp. 1–9, 2019.

3. K. Kamoi, "HTLV-1 in Ophthalmology," *Front. Microbiol.*, vol. 11, no. March, pp. 1–12, 2020.

4. Y. Tagaya, M. Matsuoka, and R. Gallo, "40 years of the human T-cell leukemia virus: past, present, and future," *F1000Research*, vol. 8, p. 228, 2019.

5. R. F. Edlich, J. A. Arnette, and F. M. Williams, "Global epidemic of human T-cell lymphotropic virus type-I (HTLV-I)," *J. Emerg. Med.*, vol. 18, no. 1, pp. 109–119, 2000.

6. M. Hatanaka and S. H. Nam, "Identification of HTLV-I gag protease and its sequential processing of the gag gene product," *J. Cell. Biochem.*, vol. 40, no. 1, pp. 15–30, 1989.

7. S. B. Shuker, V. L. Mariani, B. E. Herger, and K. J. Dennison, "Pii: S1074-5521(03)00002-4," vol. 10, pp. 1–9, 2003.

8. J. T. Nguyen *et al.*, "Truncation and non-natural amino acid substitution studies on HTLV-I protease hexapeptidic inhibitors," *Bioorganic Med. Chem. Lett.*, vol. 18, no. 1, pp. 366–370, 2008.

9. T. Satoh, M. Li, J. T. Nguyen, Y. Kiso, A. Gustchina, and A. Wlodawer, "Crystal structures of inhibitor complexes of human T-Cell leukemia virus (HTLV-1) protease," *J. Mol. Biol.*, vol. 401, no. 4, pp. 626–641, 2010.

10. M. Kuhnert, H. Steuber, and W. E. Diederich, "Structural basis for HTLV-1 protease inhibition by the HIV-1 protease inhibitor indinavir," *J. Med. Chem.*, vol. 57, no. 14, pp. 6266–6272, 2014.

11. M. Kuhnert, A. Blum, H. Steuber, and W. E. Diederich, "Privileged Structures Meet Human T-Cell Leukemia Virus-1 (HTLV-1): C2-Symmetric 3,4-Disubstituted Pyrrolidines as Nonpeptidic HTLV-1 Protease Inhibitors," *J. Med. Chem.*, vol. 58, no. 11, pp. 4845–4850, 2015.

12. Y. Niu, S. Li, D. Pan, H. Liu, and X. Yao, "Computational study on the unbinding pathways of B-RAF inhibitors and its implication for the difference of residence time: Insight from random acceleration and steered molecular dynamics simulations," *Phys. Chem. Chem. Phys.*, vol. 18, no. 7, pp. 5622–5629, 2016.

13. J. Rydzewski and O. Valsson, "Finding multiple reaction pathways of ligand unbinding," *J. Chem. Phys.*, vol. 150, no. 22, 2019.

14. P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, "Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 5, pp. E386–E391, 2015.

15. A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 5, pp. 826–843, 2011.

16. P. Tiwary, J. Mondal, and B. J. Berne, "How and when does an anticancer drug leave its binding site?," *Sci. Adv.*, vol. 3, no. 5, 2017.

17. D. Sabbadin, V. Salmaso, M. Sturlese, and S. Moro, "Supervised molecular dynamics (SuMD) approaches in drug design," *Methods Mol. Biol.*, vol. 1824, no. Md, pp. 287–298, 2018.

18. F. T. J. C. C. 2004 UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, "UCSF Chimera."

19. A. W. Sousa Da Silva and W. F. Vranken, "ACPYPE - AnteChamber PYthon Parser interfacE," *BMC Res. Notes*, vol. 5, pp. 1–8, 2012.

20. F. Hofer, J. Kraml, U. Kahler, A. S. Kamenik, and K. R. Liedl, "Catalytic Site p

KaValues of Aspartic, Cysteine, and Serine Proteases: Constant pH MD Simulations," *J. Chem. Inf. Model.*, vol. 60, no. 6, pp. 3030–3042, 2020.

21. M. J. Abraham *et al.*, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1–2, pp. 19–25, 2015.

22. E. Monvall, "Statsbudgeten: de stora reformerna gäller arbetslivet," *Tidskr. Sver. Sjukskot.*, vol. 43, no. 2, pp. 54–58, 1976.

23. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

24. B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A Linear Constraint Solver for molecular simulations," *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997.

25. T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.

26. G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, vol. 126, no. 1, 2007.

27. M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.

28. D. Sabbadin and S. Moro, "Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale," *J. Chem. Inf. Model.*, vol. 54, no. 2, pp. 372–376, 2014.

29. 25-28. Kraus, D. (2014) Consolidated data analysis and presentation using an open-source add-in for the Microsoft Excel® spreadsheet software. Medical Writing, 23, "Daniel's XL Toolbox."

30. 2007. J. D. Hunter, “Matplotlib: A 2D Graphics Environment”, Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, “Matplotlib.”

31. F. Sohraby and H. Aryapour, “Comparative analysis of the unbinding pathways of antiviral drug Indinavir from HIV and HTLV1 proteases by supervised molecular dynamics simulation,” *PLoS One*, vol. 16, no. 9, p. e0257916, 2021.