Mechanism of the association pathways for a pair of fast and slow binding ligands of HIV-1 protease

Yu-ming Mindy Huang, Mark Anthony V. Raymundo, Wei Chen, and Chia-en A. Chang

Biochemistry, Just Accepted Manuscript • DOI: 10.1021/acs.biochem.6b01112 • Publication Date (Web): 06 Jan 2017

Downloaded from http://pubs.acs.org on January 8, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.
Mechanism of the association pathways for a pair of fast and slow binding ligands of HIV-1 protease

Yu-ming M. Huang1§, Mark Anthony V. Raymundo1, Wei Chen1,2 and Chia-en A. Chang1

1Department of Chemistry, University of California, Riverside, Riverside, CA 92521, USA

2ChemConsulting LLC, Frederick, MD 21704, USA

§Current address: Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA

Corresponding author: Chia-en A. Chang

email: chiaenc@ucr.edu

phone: 951-827-7263
Abstract

Equilibrium constants, together with kinetic rate constants of binding, are key factors in the efficacy and safety of drug compounds, informing drug design. However, the association pathways of protein–ligand binding, which contribute to their kinetic behaviors, are little understood. In this work, we used unbiased all-atom molecular dynamics (MD) simulations with an explicit solvent model to study the association processes of protein–ligand binding. Using the protein–ligand systems of HIV protease (HIVp)–xk263 and HIVp–ritonavir as cases, we observed that ligand association is a multi-step process involving diffusion, localization, and conformational rearrangements of the protein, ligand and water molecules. Moreover, these two ligands preferred different routes of binding, which reflect two well-known binding mechanisms: induced-fit and conformation selection models. Our study shows that xk263 has a stronger capacity for desolvating surrounding water molecules, thereby inducing a semi-open conformation of the HIVp flaps (induced-fit model). In contrast, the slow dehydration characteristic of ritonavir allows for gradual association to the binding pocket of HIVp when the protein’s flap conformation is fully open (conformation selection model). By studying the mechanism of ligand association and understanding the role of solvent molecules during the binding event, we can obtain a different perspective on the mechanism of macromolecule recognition, providing insights for drug discovery.
Introduction

Many diseases can be treated clinically by delivering drugs to specific protein targets. The thermodynamic properties of protein–ligand recognition are important predictors of ligand efficacy\(^1-4\), but the importance of binding kinetics in biological processes has not been recognized until recently \(^5-7\). Binding kinetics may be associated with different binding mechanisms. For decades, scientists have discussed various theories to explain protein–ligand binding such as lock and key, conformation selection and induced-fit. Understanding the underlying mechanism of binding kinetics expands our ability to design ligands.

For some ligands and pharmaceutical targets, binding kinetics could play a significant role in \textit{in vivo} efficacy\(^5,6,8-11\). To illustrate this, for the systems without intermediates, the equilibrium constant (\(K_{eq}\)) is the balance between the two rate constants, \(K_{eq} = k_{on}/k_{off}\)^\(^{12-14}\). To enhance binding affinity, a good drug candidate is expected to have fast \(k_{on}\) and slow \(k_{off}\), corresponding to longer residence time\(^15\). However, in some cases, \(K_{eq}\) is not correlated with \textit{in vivo} efficacy; rather, \(k_{off}\) is better correlated\(^16\). Therefore, understanding the details of binding kinetics is critical to drug discovery, for optimizing the drug efficacy and reducing medical attrition\(^17-19\). However, the physical basis behind fast or slow binding ligands is still unclear.

Revealing the details of the entire binding pathway of a drug via traditional methods is extremely difficult, if not impossible. The available structures for free and bound states of protein systems from X-ray crystallography and NMR studies provide limited information about binding pathways. Fortunately, molecular modeling provides an alternative to gain insight into the details of ligand association. However, one of the major challenges with modeling is the time scale of binding events. Typically, to investigate a ligand diffusing toward and binding to a
protein, numerous simulations with timescales greater than 1 µs are required. To overcome computational costs, previous studies performed in our lab and in other labs used coarse-grained models to study binding kinetics. These studies resulted in valuable information about the diffusion steps of proteins binding with ligands, including xk263 and ritonavir\textsuperscript{20-22}. Recently, modeling a non-specific association of xk263 via Brownian dynamics (BD) with the GeomBD2 program further added to prior work by showing the probability of where the ligand will most likely be found on the surface of HIV-1 protease (HIVp) during diffusion\textsuperscript{23}.

Because of current advances in technology, nowadays, researchers start having a chance to use unbiased, all-atom molecular dynamics (MD) to understand the insights that may affect the binding kinetics of drug–protein systems in a microsecond timescale. For instance, a recent study of G-protein–coupled receptors and Src kinase examined the role of water molecules upon association of the drug. The same study revealed multiple metastable intermediates throughout the entire binding process\textsuperscript{24, 25}. A different study analyzed the water density in the binding interfacial gap. The de-wetting process was found to guide the ligand into the binding pocket and accelerate the approach to the final bound state, especially when binding on a hydrophobic surface\textsuperscript{26}. Thus, hydrophobic dehydration can be considered an essential driving force for assembly of a ligand and a non-polar surface\textsuperscript{27, 28}. Another study involving Barnase protein showed that a hydrogen bond network on the binding interface stabilizes the transition complex with the surrounding water molecules, thereby facilitating protein–protein binding\textsuperscript{29}. Setny et al., showed that fluctuations in hydration and stochastic motions of a ligand were intimately coupled, thus increasing the entrance friction, which decelerated the association kinetics\textsuperscript{30}. Finally, free energy calculations of a protein–ligand complex suggested that water molecules moving from the binding site to the solvent acts as a rate-determining step in ligand association\textsuperscript{31, 32}. In technique
and in choice of protein systems, these studies greatly differ. However, their results and
collections highlight common themes during the binding process: overall, transitions of
hydrogen bonds (H-bonds) and the movement of water molecules during the binding process are
important factors governing the rate of ligand association\textsuperscript{24, 33, 34}. Although the above studies
provided some insights into the mechanism of ligand binding, most of the work did not discuss
changes in protein conformations during ligand entry. However, this does affect the binding
kinetics of drugs; for example, pioneer work done by Shan and co-workers showed that drug
imatinib binding kinetics may provide a probe of the DFG loop transition in the Abl kinase
system\textsuperscript{35}. Also, the mechanisms of fast or slow binding kinetics are still unclear, either.

To demonstrate the events that may lead up to the binding pathway of protein–ligand
association, we selected HIV\textsubscript{p} and two ligands, xk263 and ritonavir, as our study systems. HIV\textsubscript{p}
plays an integral role in the HIV life cycle and is one of the major targets in anti-AIDS
treatments\textsuperscript{36}.

The HIV\textsubscript{p} structure features a $C_2$ axis of symmetry with two flexible glycine-rich flaps
that behave as a gate to control ligand access\textsuperscript{37}. Three conformations of the flaps were detected
in the native state: open, semi-open and closed (Figure 1). Experimental and computational
studies revealed that the semi-open conformation is the dominant form in the ligand-free state\textsuperscript{38, 39}; however, with ligand binding, the flaps favor the closed conformation\textsuperscript{40-42}. Two types of
handedness orientation, semi-open and closed, are associated with the ligand-free and -bound
states, respectively (Figure 1)\textsuperscript{42, 43}. Since the transition from one flap conformation to another of
HIV\textsubscript{p} was observed from early studies, we should expect the flaps to undergo large
conformational changes when the protein switches from the free to bound state.
Most experimental and computational studies focused on the mechanism of flap opening and closing. Two proposed binding models have attempted to explain the association process. The induced-fit model hypothesizes that a binding event occurs when a ligand induces conformational changes and interrupts the native movement of a protein; a ligand could approach HIVp in its semi-open flap conformation, disrupt the flap–flap interactions to induce flap opening, and enter the binding pocket to form a complex. The other model, conformation selection, proposes an opposite assumption: a ligand directly binds to a pre-organized protein conformation, such as HIVp with an open flap conformation. After the ligand binds and stays in the binding site, the flaps close. A previous study suggested that both of these models can explain the HIVp–ligand association but was unable to elucidate the reason for the fast or slow kinetics of a ligand.

We used as an example two HIVp ligands to demonstrate the cases of fast and slow binding: the cyclic urea compound xk263, a fast binder, with $k_{on} \approx 10^9 \text{M}^{-1}\text{s}^{-1}$; and the peptide-mimic compound ritonavir, a slower binder, with $k_{on} \approx 10^6 \text{M}^{-1}\text{s}^{-1}$ (Figure 1). Despite a difference of three orders of magnitude in association rate constants between the two ligands, they bind to HIVp tightly, with $K_d \approx 10^{-9} \text{M}$ (approximately -11 kcal/mol)

With knowledge gained from prior work, we aimed to use unbiased MD simulations to focus on the events leading up to binding and understand the changes in dynamics that result from a ligand binding. We believe that these events can further explain the fast and slow kinetics of binding reported from experimental studies. Overall, we studied different poses of protein–ligand complexes as well as the dynamics of water molecules by using conventional MD simulations for HIVp–ligand systems. We illustrate potential binding mechanisms for ligands with different association kinetics and explain how water molecules affect the rate of binding.
Materials and Methods

Molecular systems

We selected the structure of the ligand-free HIVp from the protein data bank (PDB) code 1HHP (resolution 2.70 Å) to study the motions of the free protein. The conformations of HIVp binding to xk263 and ritonavir were taken from the PDB codes 1HVR (resolution 1.80 Å) and 1HXW (resolution 1.80 Å), respectively. Of note, xk263 has significantly more ligand hydrophobicity than ritonavir. The LogP values were computed by using different software packages. Details are in the Text S3.

MD simulations

We used the Amber 12 package with an efficient GPU implementation for MD simulations of the free HIVp and ligand binding pathways. Amber 99SB and General Amber Force Field (GAFF) were applied to HIVp and the two ligands, respectively. By checking the unperturbed charge of the system, the counter-ion Cl−, was placed to maintain a neutral system. Minimization on the hydrogen atoms, side chains and the entire protein complex was applied for 500, 5000 and 5000 steps, respectively. After being solvated with a rectangular TIP3P water box, the edge of the box is at least 12 Å away from the solutes. The system went through a 1000-step water and 5000-step system minimization to correct any inconsistencies. Next, we relaxed the system by slowly heating it during an equilibrium course of 10 ps at 50, 100, 150, 200, 250 and 300 K. The long-range electrostatic interactions were computed by the particle mesh Ewald method beyond 8 Å distance. The time step of the simulations was 0.002 ps with a non-bonded cutoff of 12 Å. We collected the resulting trajectories every 1 ps, with time step 2 fs, in an isothermic-isobaric (NPT) ensemble. The Langevin thermostat, with a damping constant
of 2 ps⁻¹, was used to maintain a temperature of 300 K. Finally, the SHAKE algorithm was used to constrain hydrogen atoms during the MD simulations⁵⁹.

Because free HIVp has multiple flap conformations, we performed a 250-ns simulation on the free protein to obtain a variety of different protein conformations; four snapshots of the simulation showing different orientations and handedness of the flaps (Figure 2) were selected for molecular docking; MD simulations were based on 58 initial protein–ligand conformations obtained from molecular docking (see the next section for details). We first performed 200-ns of MD simulations on each initial conformation by using cpu parallel processing and local gpu machines. Then, conformations with ligands close to the binding site (a result of 17) were selected for further study by using local gpu machines (timescale 300~1000 ns). To further investigate the association pathway, we chose two promising representative models for xk263 and three for ritonavir and performed 3~14-µs MD simulations with a special-purpose computer, Anton⁶⁰, ⁶¹. Finally, the trajectories were collected and analyzed at intervals of 1 ps, 100 ps or 1 ns, depending on the length of the simulation.

Docking protocol

Autodock with the Lamarckian genetic algorithm was used to dock ligands on designated regions of HIVp⁶². We placed the docking box 20~30 Å away from the center of the HIVp binding site; naturally, ligands could approach HIVp from any direction, so the docking box was placed at the top, left, right and front regions of the protein (Figure 2). We did not consider the back region because of the symmetry of the protein. The placement of the docking box reflected areas of high probability that the ligand would most likely diffuse toward, as determined from previous work by our group and others²⁳, ³⁴. In a ligand-free HIVp, the dominant conformation is
the semi-open form (37, 38, 45). To reflect this form and other open-like forms, we used four different conformations — semi-open, slightly open, open with one flap curled and wide-open — for ligand docking (Figure 2). Combining different protein conformations with different placements of the ligands to areas to which it would most likely diffuse should sufficiently reflect the events leading up to ligand binding. Vcharge was used to assign partial charges for the ligand atoms. The Autogrid 4.0 algorithm was used to create a 22.5-Å dimension cubic box with grid spacing of 0.375, and the ligand was placed 20-30 Å away from HIVp. We performed 10 trials for each ligand-docking simulation, with one million energy evaluations for each trial. We considered only the conformation that had the lowest free energy from each set of trials. Notably, one can randomly place a ligand in the cubic box to set up initial structures for MD simulations. We used molecular docking in this study solely for establishing a more systematic procedure to create initial conformations. Overall, 58 conformations with different flap orientations and ligand positions were generated for sampling ligand association pathways.

Gaussian Accelerated Molecular Dynamics (gaMD) protocol

Because it takes an excessively long time for a ligand to locate a final bound state shown in a crystal structure, we used gaMD in AMBER 14 for more efficient conformational search of pathways that could lead the ligands to the final bound form. Snapshots of ligands that were 5-7 Å away from the crystal bound state were extracted to continue simulations with gaMD for thorough conformational samplings. After multiple runs of gaMD, we took snapshots of ligands that reached a root mean square deviation (RMSD) of 3-4 Å of their crystal structure position. These poses were reseeded by using conventional MD (cMD). Overall, we ran 20 gaMD, from 50 to 100 ns, and 44 cMD simulations (100 ns each) for ritonavir, and ten 100-ns gaMD simulations and 35 cMD simulations (from 100-350 ns) for xk263.
Post-MD Analysis

(1) RMSD and MM energy

Root-mean-square deviation (RMSD), distance among atoms, and molecular mechanics (MM) energy from electrostatic and van der Waals contributions were calculated by using VMD \(^6\). Reference coordinates for RMSD calculations were from the crystal structures (i.e., PDB code 1HVR and 1HXW for HIVp with xk263 and ritonavir, respectively) \(^50, 51\).

(2) Criteria for a “bound” ligand

To compare the ligand-bound structures from our MD simulations with the crystal structures, we first aligned the protein backbone. Next, the center of mass (COM) of selected ligand atoms was calculated (see Text S1). Then, the distances between the COM from the MD snapshots and crystal structures were measured. In this study, we considered that a binding event occurred if the COM distance between the ligand-bound structure and the ligand in the crystal structure was less than 7 Å for more than 5 ns. From previous Brownian dynamics (BD) studies, this cut-off reflects a high probability of successful binding \(^21\).

(3) Removal of free energy of water molecules

The free energy needed to remove a bound water molecule from a protein in solution can be considered the energy needed to move the water molecule into bulk solution, leaving an empty cavity with the same shape as that of the removed water. Thus, the desired water-removal free energy was computed as:
\[ \Delta G^\circ_{\text{wr-liquid}} = \mu^0_{PL} + (\mu^0_{W-gas} + \Delta G_{gl}) - \mu^0_{PLW} \]

where \( \mu^0_{PLW} \) is the standard (C\( ^0 = 1\)M) chemical potential of a protein–ligand–water complex, \( \mu^0_{PL} \) is the standard chemical potential of the protein–ligand complex without the water (but with the water-shaped cavity), and \( \mu^0_{W-gas} \) is the standard chemical potential of the water in gas phase (no GB or PB solvation). \( \Delta G_{gl} \) is the gas-to-liquid transfer of free energy of the water.\(^{66}\) The details for the calculations of \( \mu^0_{PLW}, \mu^0_{PL}, \Delta G_{gl} \) and \( \mu^0_{W-gas} \) are in Supporting Materials.

The removal free energies of water molecules were calculated for each of 21 frames saved every 5 ps for 100 ps, which corresponds to the average life span of the bridge waters in both ritonavir and xk263 complexes. The bridge waters of interest in the 21 frames were chosen by using in-house software for the water removal analysis.

**Results and Discussion**

**Overview of ligand binding processes**

Using all-atom MD simulation, many computational studies have attempted to elucidate and understand the underlying factors that govern kinetic behaviors of various ligands. Our goal was to understand the factors involved in fast and slow binding by analyzing the events leading up to binding and understand the changes in dynamics that result from a bound ligand by using unbiased MD simulations. The diffusion steps for ligands far from HIVp have been reported\(^{21, 67, 68}\), so we focused on a ligand’s association with HIVp when it is close to the target, 20~30 Å away from the center of the HIVp active site. Because we knew that the ligands can diffuse from all directions surrounding the protein, ligands were docked at the top, left, right, front and bottom regions around the HIVp as starting conformations in MD simulations (see Figure 2). We then
performed continuous MD simulations from the 58 different starting conformations (29 for each ligand), ranging from 200 ns to 14.0 µs (55.1 µs in total), in an explicit solvent model. Among all simulations performed, xk263 successfully entered the binding site in 6 of the MD simulations, half from the semi-open conformation (see Table S1). Similarly, ritonavir successfully entered the binding site in 8 of the MD simulations, half from the widely open conformation (see Table S1). We used these 14 simulations to observe and analyze the association pathways for both ligands. The figures reported in this paper reflect a representative pathway that was found in common with others. Another 21 of the simulations showed that a ligand (xk263 or ritonavir) interacted with residues near the HIVp active site similar to positions found in previous BD studies. However, we did not consider these simulations as capable of reaching the bound form in a timely manner, given our limited computational power. In addition, both ligands stayed in other regions, such as the flap elbow and bottom of HIVp, which was also observed by BD simulations with a coarse-grained model (Figure S1)\(^{39, 69}\). Details about all 58 simulations are in Table S1.

Our unbiased MD simulations suggest that the initial location of the ligand may affect the probability of successful binding. Nearly 28% of the simulations starting from the top region and 25% from the front region (Figure 2) resulted in binding to HIVp. In contrast, less than 10% of the simulations starting from regions peripheral to the protein resulted in a bound state. These results agree with prior study of the diffusion of both ligands to HIV-1 protease. Since the flaps of the protein are located at the mid-front and top region, previous BD simulations with a coarse-grained model showed that the chance of a binding event occurring from regions below or peripheral to the enzyme was low\(^ {21}\). In some simulations, when a ligand stayed at the elbow or bottom regions of the HIVp, given enough time, it may diffuse away from the protein and rebind.
at a different region, or remain bound in its general vicinity. Although we can model the probability of successful binding from diffusion via the protein surface, significantly longer simulation time is needed (perhaps in the scale of hundreds of µs to ms for each run), which is not possible with the current computer power. Thus, we did not pursue simulations in which the ligand stayed bound at the bottom or the elbow regions.

Figure 3 shows an example of a successful ligand binding. All successful simulations suggested that xk263 and ritonavir first contacted one of the flaps or one of the loops containing Pro79/Pro178 (Figure 3(ii)). Furthermore, the ligands could be clamped between the two loops via non-polar interactions (Figure 3(iii)B and 3(iii)D). Even after xk263 or ritonavir entered the binding pocket, the flaps kept switching between the open and closed form. This switching allowed for adjusting the orientation of the ligand and the handedness of the flaps until they were in the closed conformation. During this rearrangement period, our simulations showed multiple protein-complex intermediates that were not reported in other studies.

Previous studies of apo HIVp with all-atom MD and coarse-grained BD simulations suggested that HIVp flaps stayed in a closed or semi-open conformation for ~400 ns, on average, and then opened for ~40-50 ns (38, 45). Therefore, in this study, if the conformation of the complex persisted for longer than 500 ns, we considered this a local intermediate state (Figure 3(iii))34, 70. If the flaps opened from a semi-open or closed conformation within 200 ns while a ligand approached, we considered the motion an induced flap motion caused by the presence of the ligand.

*Sampling the crystal bound conformation*
By using the distance between the COM of ligand positions obtained from our MD simulations and experimental crystal structures, our longest MD trajectory of 14 µs (xk263) and 8 µs (ritonavir) was less than 2.5 Å away from the bound crystal structure. However, our MD simulations did not sample ligand conformations that exactly reproduce the crystal structures. Ritonavir was in a curled conformation, whereas the naphthyl rings of xk263 did not rotate to its crystal conformation. Notably, when performing experiments to determine a bound state for kinetic measurement, a ligand must have a bound form identical to a crystal structure. Additionally, a molecular system takes a long period of time for conformational rearrangements to occur when generating a crystal structure. Thus, a significant amount of time is expected for the ligands to reorient themselves to sample the crystal structure conformation. The specialized Anton machine allowed us to sample up to 14 µs, which, however, is not long enough for the molecular system to rearrange solute and solvent conformations to sample the form shown in the crystal structure. Therefore, we combined accelerated gaMD and cMD to widen the range of sampling to find conformations close to the final bound state of the crystal structure (Figure 4).

The combined gaMD and cMD simulations showed that the average distance for xk263 was ~0.45 Å away from the bound state of xk263 as compared with the crystal structure (1HVR) (Figure 4A) and for ritonavir was ~ 2.70 Å away as compared with the crystal structure (1HXW) (Figure 4B). Ritonavir may have greater distance than xk263 because the crystal bridge water located between ritonavir and HIVp flaps has not been successfully sampled yet. In addition, ritonavir has more rotatable bonds and polar groups than does xk263, so it may need a longer simulation time for thorough sampling. With fewer hydrogen bonds to overcome and no need to recruit bridge waters in the complex, xk263 may more easily sample the space seen in the crystal structure. Even though the distance for bound ritonavir is not extremely close to the crystal
structure, the orientation and conformation is in good agreement with the crystal structure (Figure 4B). Overall, this study suggests that use of only unbiased conventional MD to model an entire association pathway may not lead to a conformation that can reproduce the crystal structure; enhanced simulations or sampling techniques are required.

Different binding mechanisms for fast and slow binding ligands

Unlike some proteins whose ligands can reach the binding site via a well-defined channel\textsuperscript{24, 71}, HIVp has an open binding site and flexible gating. Thus, the mechanism for binding is complicated. Early fluorescence and NMR studies of HIVp proposed that ligands binding to it experience both an induced-fit and conformation selection binding mechanism\textsuperscript{72, 73}. As well, from our simulation results, the fast (xk263) and slow (ritonavir) ligands have strikingly different preferences in binding to different protein initial conformations; thus the two ligands present different association mechanisms. Among 26 sampled paths with the ligands diffused to HIVp in a semi-open flap conformation, successful binding occurred in 18.8% and 10.0% of the simulations for xk263 and ritonavir, respectively. Conversely, among 22 sampled paths with the ligands diffused to HIVp in an open flap conformation (including open with a curl flap and wide-open), successful binding occurred in 38.5% of the simulations for ritonavir as compared with 11.1% for xk263. If we further consider the simulations of the HIVp conformation with wide-open flaps, the successful ligand binding rate was 57.1% for ritonavir but 0% for xk263. Therefore, xk263 and ritonavir feature more efficient binding to HIVp with semi-open and wide-open conformations, respectively. Thus, we infer that the fast (xk263) and slow (ritonavir) binding ligands to HIVp use the induced-fit and conformation selection mechanism, respectively. These results prompted us to investigate why the fast and slow binding ligands prefer different
pathways and mechanisms, and what factors contribute to these preferences. The difference of
the binding paths of the two ligands is also discussed in the following sections.

*Induced-fit mechanism: ligands binding to the semi-open-flap HIVp*

We examined the binding processes and protein dynamics as both ligands bind to the
most common conformation of HIVp in its free state; the major conformation of free HIVp is
semi-open flaps with semi-open handedness. The MD simulations for the free HIVp show that
this dominant flap conformation could change to wide-open flaps with semi-open or closed
handedness after 100-200 ns (*Figure S2 and S3*). Although both xk263 and ritonavir can enhance
the conformational flexibility of the flaps, as observed by RMSD calculations (*Figure S4*), the
tendency for the two ligands to alter the flap conformation and handed orientation differs. In all,
90% of the simulations involving xk263 showed induced conformational changes in the flap
region, where the flaps switched a conformation from the dominant form to closed flaps with
closed handedness within 150 ns (*Figure S2 and S3*). In some simulations, xk263 induced the
flap motions to wide open but did not completely bind inside the pocket. Conversely, only 50%
of MD runs with ritonavir showed similar observable flap motions. Ritonavir formed
significantly more transient H-bonds with HIVp than xk263 to stabilize intermediate states,
which is consistent with our previous MD simulations with an implicit solvent model

During the binding process, the flaps need to open up to some extent for ligand access. Once the
ligands are bound, HIVp needs to switch from the semi-open to closed handedness to be one step
closer to reach the conformation reproduced in the crystal structure. Therefore, ligands such as
xk263, which can interact with HIVp in a semi-open conformation can easily induce the flaps to
change their conformations within 200ns, allowing xk263 to bind to the protein quickly (*Figure
5*).
Role of removing transient bridge waters in the induced-fit mechanism

Our simulations revealed that transient bridge waters between HIVp and a ligand stabilize an intermediate state and prevent direct protein–ligand contact. Transient bridge waters act as a buffer zone that weakens interactions between the ligand and HIVp. Our simulations suggest that the flaps show considerable motion only when a ligand contacts HIVp directly without any bridge waters between them. As illustrated in Figure 6, any bridge waters associated with xk263 were quickly stripped out, and xk263, now in direct contact with the flaps, explored several intermediate states. In contrast, water molecules were frequently found between ritonavir and HIVp, which resulted in an intermediate state that lasted longer than 500 ns.

To quantitatively examine the stability of transient bridge waters in the intermediate states of HIVp–xk263 compared to HIVp–ritonavir, we calculated the free energy (FE) cost to move these water molecules from the interaction sites to bulk solvent. From our MD simulations, we focused on the transient bridge waters that stayed between the protein and ligand for longer than 100 ps (Figure 6). Unlike the HIVp–ritonavir complex, with transient bridge waters staying longer than 100 ps, for the xk263 complex, most transient bridge waters stayed for considerably less time (i.e., 15±5 ps), which implies that bridge waters with xk263 were less stable (Figure S5). The computed FE for water removal was saved every 5 ps during the 100-ps run. For xk263, the water molecules we studied had removal energies ranging from 1.5 to 2.0 kcal/mol, which demonstrates their loose binding ability in the HIVp–xk263 system. A close inspection of the conformations revealed that water molecules needed to form a hydrogen bond network to bridge the ligand and protein, a network that seems to be vulnerable in the HIVp–xk263 system. For ritonavir, transient bridge waters had a removal energy of > 5.0 kcal/mol in a 100-ps MD run (Figure 6 and S6). This water molecule directly bridged the thiazole moiety of ritonavir and
several residues of HIVp. Compared to xk263, ritonavir has more polar functional groups that may contribute to the larger FE and longer dwell time of the transient bridge waters in the HIVp–ritonavir intermediate states. These polar functional groups provide more opportunities to form a network of hydrogen bonds with the surrounding water molecules. Because of those stable bridge waters, as compared with xk263, ritonavir has weaker interactions with HIVp and is thus less likely to induce flap motions. Additionally, the time-consuming dehydration process slows the rate of binding for ritonavir.

*Conformation selection mechanism: ligands binding to the open-flap HIVp*

On investigating the association process with ligands binding to HIVp in its open state, we found that ritonavir prefers binding to the protease in an open-flap conformation. Unlike the semi-open conformation, this pathway does not involve ligand-induced protein conformational change because the flaps are already open. Binding to the open-flap conformation forces the ligand to pass through a funnel-like region containing a large number of water molecules to reach the binding cavity (“channel” in Figure 7A). If the flaps open, ritonavir needs to go through the channel to reach the binding site. Thus, gradual dehydration is necessary for entering the protein (Figure 7). When the ligand is located in the bulk solvent (Figure 7A), ~200 water molecules surround ritonavir within the second hydration shell (6.0 Å from ritonavir). The number of water molecules continues to decrease as ritonavir approaches the binding site (Figures 7B–7F). When the ligand binds to the active site and the flaps of HIVp are nearly closed, ~100 water molecules are retained within the second hydration shell (Figure 7G). Similar results from another MD simulation are shown in Figure S7. The HIVp–ritonavir complex may sample multiple intermediate states after ritonavir enters the binding site. During this time, the complex adjusts its conformations, and the flaps are slightly re-opened. When this occurs, nearby water
molecules quickly re-solvate the binding cavity (Figure 7H), but the system is dehydrated again, which allows the flaps to close. Also, a rearrangement of the handedness of the flaps occurs before reaching the closed conformation. Our MD simulations indicate that > 20 ns is required to close the flaps from an open conformation. Whether the solvent facilitates the closing of the flaps or the attractions between ritonavir and HIVp repel water molecules is still unclear.

We observed that xk263 diffused faster toward the gate of the protein (~1 ns) than did ritonavir (~10 ns) before directly contacting HIVp. Because of the highly non-polar character of xk263, it can quickly dehydrate surrounding waters and interact with non-polar residues of the protease. The flap regions of HIVp contain non-polar residues giving it a hydrophobic characteristic (Figure S8). Overall, the hydrophobic effect plays a role here in that xk263 quickly formed non-polar interactions with the residues on the surface of the flaps instead of moving on to the binding pocket. As a result, xk263 was found near the entrance of the binding site or was trapped between two flaps (Figure 8 and S9). The strong non-polar attractions between the flaps and xk263 also led to flaps closing before xk263 could enter and settle into the active site. In addition, xk263 can diffuse along the non-polar surface of one flap, thereby moving away from the active site of the enzyme (Figure 8). The re-opening of the flaps may provide another chance for the ligand to enter the site, but this would extend the timescale of the ligand binding process. Overall, our simulations suggest that fast dehydration of the solvent by xk263 does not help the ligand enter the protease with an open-flap conformation.

We calculated protein–ligand and water–ligand MM energies to study the energy contributions, especially electrostatic and van der Waals energies, which drive ligand binding. Non-polar interactions are typically considered the principal driving force in molecular recognition. However, our calculations show that ritonavir binds to HIVp via strong electrostatic
attractions, which is contrary to the binding of xk263 (dominated by non-polar interactions) (Figure 9A and C). Columbic and van der Waals energy had the same contribution to total energy during a 200-ns simulation (Figure 9C). Therefore, both polar and non-polar atoms in ritonavir play a role in forming a protein–ligand complex.

A simple continuum solvent model cannot account for the explicit transient waters located between HIVp and its ligands. Thus, we calculated interaction energies between the ligands and explicit waters from MD snapshots. Although the total interaction energy between the ligand and waters fluctuated, ritonavir generally formed strong electrostatic interactions with water molecules (Figure 9B and D), which suggests that before establishing a stable HIVp–ritonavir complex, ritonavir could interact with many water molecules for an extended time.

Implications

In this work, we report the continuous processes (mechanisms) of a fast and slow binding ligand entering the binding site of HIVp. Our work contrasts with early studies of HIVp, which used docking tools or Markov state models to search for potential binding poses or energies of the steady states in the ligand-binding pathways. We focus on the pathways and mechanism regarding how ligands approach the protein active site and the corresponding protein conformational changes during ligand binding. By studying 14 successful binding paths of xk263 and ritonavir, we show that both ligands can locate the binding pocket and after approaching the binding site, sample multiple local intermediate conformations, thereby extending the time for a ligand to reach one of the global energy minimum shown in a crystal structure. For example, in our 14-µs simulation of xk263, the ligand reached the binding site within 1.5 µs. The remainder of the time was spent rearranging its conformation during the simulation (Figure 3(i), 3(iii) and
This result contrasts with early reports by D. E. Shaw Research Company staff for different protein systems. In their unbiased MD simulations, they studied ligands binding to Src kinases and G-protein–coupled receptors (GPCRs) and found that the ligands moved around the protein surface for an extended period of time. However, once the ligands reached the entrance of the binding site, the steady-state conformations or the final bound state could be approached in a few microseconds\textsuperscript{24, 25}. One difference with our study could be due to the structure of the ligands we studied, which had more atoms and rotatable bonds. For example, the inhibitors in the Shan et al and Dror et al. studies had about 20 heavy atoms and 4-8 rotatable bonds\textsuperscript{24, 25}, as compared to xk263, with 46 heavy atoms and 8 rotatable bounds, and ritonavir, with 50 heavy atoms and 19 rotatable bonds. Another difference is the nature of the binding site between HIV\textsubscript{p} and Src kinase and/or GPCRs. The asymmetrical motions of the HIV\textsubscript{p} flaps in addition to the wide range of points from where its ligands can associate and enter the active site significantly contrasts with those for Src kinase and GPCRs; GPCRs have a well-defined binding pocket limiting the direction from which their substrate can enter. As the differences mentioned above, the HIV\textsubscript{p} ligands present different binding paths and mechanisms compared to the existing systems studied. In addition, our results also show the challenge of modeling ligand association of HIV\textsubscript{p}, which has two unique and flexible flaps. We note that ligand binding mechanisms may highly depend on different protein systems; more studies in ligand binding kinetics should reveal the differences among various classes of proteins in the future.

Conclusions

The goal of our study was to gain insights that would better explain the fast and slow kinetics of binding by studying the entry and association processes of two different ligands. By using unbiased MD simulations with an explicit solvent model, we provide atomistic insights
into protein–ligand association pathways. We studied the ligands xk263 and ritonavir binding to HIVp. Our simulations showed that binding events were more successful when initially diffused from the top and front regions of HIVp. In addition, the two ligands studied use different binding mechanisms. The induced-fit model of binding hypothesizes that a binding event occurs when a ligand induces conformational changes and interrupts the native motions of a protein \(^{48}\). The other mechanism of binding, conformation selection, proposes that a ligand directly binds to a pre-organized protein conformation \(^{75,76}\). The fast binder, xk263, with a predominantly non-polar property, preferentially binds to HIVp with a semi–open flap conformation, which reflects the induced-fit mechanism. This binding is accomplished by dehydration of surrounding water molecules, thereby allowing xk263 to directly interact with the protein and induce flap motion to allow xk263 access to the active site. The slower binder, ritonavir, successfully binds to HIVp with an open-flap conformation, which reflects the conformation selection mechanism. Because of the more polar features of ritonavir, the hydration shell surrounding it decreases ritonavir’s ability to directly interact with the protease and induce flap motion; thus, ritonavir must wait for HIVp to sample an open conformation for favorable binding. In addition, ritonavir forms more stable transient hydrogen bonds and retains more bridge water molecules than xk263 when binding to the protein. Although these bridge waters weaken the protein–ligand interactions, so that the induced-fit mechanism is less likely to occur, the waters help ritonavir smoothly reach the active site of HIVp in an open-flap conformation. Computed free energies for the removal of transient waters between the ligands and HIVp show that the bridge water molecules in HIVp–xk263 intermediates are less stable and thus have shorter dwell time than those in HIVp–ritonavir. Because of this de-wetting nature of xk263, the induced-fit mechanism is the preferable route for binding. Overall, our study suggests potential binding mechanisms for
ligands with different association kinetics and explains how water molecules affect the binding rates.

**Supporting Material**

A table listing all MD simulations and 5 figures is available at WEBPAGE.

**Author Contributions**


**Acknowledgements**

We are grateful to Dr. Christopher Roberts for the energy decomposition script. This work was supported by the US National Institutes of Health (NIH; no. R01GM109045) and US National Science Foundation (NSF; no. MCB-1350401). Anton computer time was provided by the National Center for Multiscale Modeling of Biological Systems (MMBioS) with a grant (P41GM103712-S1) from NIH and the Pittsburgh Supercomputing Center (PSC). The Anton machine at PSC was generously made available by D.E. Shaw Research.
References


27


Figure Legends

Figure 1: Schematic representation of HIVp structure. Flaps with open, semi-open and closed conformation are in red, green and blue, respectively. The orientation of the flaps, called handedness, has two forms: semi-open (green) and closed (blue).

Figure 2: The two ligands, xk263 and ritonavir, are docked to a box placed 20-30 Å from HIVp. The four initial HIVp conformations with different flap orientations — wide open (red), open with a curl flap (green), slightly open (yellow) and semi-open (blue) — were used in the docking procedure. The ligands were docked in the top, down, right, left and front regions (gray area) as starting conformations in MD simulations.

Figure 3: The pathway of xk263 binding to semi-open HIVp. (i) A trace of xk263 binding route. One bead indicates a position that the center of mass (COM) for xk263 stays at a specific simulation time. Red, yellow and green represent the MD simulation time zone of about 0-0.5, 0.5-1.5 and 1.5-14.0 \( \mu \text{s} \), respectively. (ii) An alignment of a crystal structure and the predicted conformation from our MD simulation. The crystal structure shown in cyan was taken from PDB code 1HVR. The conformation from the MD simulation at 14 \( \mu \text{s} \) is in pink. (iii) MD snapshots of the transient complex conformations during the binding process. (iv) The flap distance and ligand RMSD by MD simulation time. Notably, the pathway reported is from our longest MD simulation for this molecular system. This pathway is representative of common ones found when comparing several other pathways of xk263 associating with HIVp.

Figure 4: Representative ligand binding modes close to crystal structures. MD snapshot of the lowest center of mass (COM) distance between our simulation (green) and the final bound state (yellow) shown in crystal structures. The COM from MD and crystal structures is
represented by a blue and red bead, respectively. (A) The COM distance for xk263 between a MD snapshot and PDB 1HVR was 0.45 Å. (B) The COM distance for ritonavir between a modeled position and PDB 1HXW is 2.70 Å. Residues 76 to 82 were omitted for viewing purposes.

**Figure 5: Comparison of xk263 and ritonavir binding to native semi-open HIVp.** MD snapshots during the first 200 ns; the conformation of HIVp is initially semi-open flaps and semi-open handedness. Both xk263 and ritonavir interact with the protein at a similar region around 18 ns. Xk263 desolvated water molecules and therefore induced conformational changes of the flaps in 56 ns; the flaps then open with the association of xk263. However, because ritonavir has tighter-bound transient bridge waters (not shown explicitly in this figure), it did not induce significant conformational changes in the simulation. The simulations reported here are representative of common features seen in other simulations.

**Figure 6: Free energy needed to remove a specific water molecule during ligand association.** Red and blue lines in the plot show the free energies needed to remove a bridge water molecule with xk263 and ritonavir binding to HIVp, respectively. xk:A, xk:B and xk:C indicate a snapshot at 0, 14 and 59 ps, respectively. rit:A, rit:B and rit:C indicate a snapshot at 0, 14 and 29 ps, respectively. The snapshots represent the beginning snapshot and snapshots with the highest and lowest removal free energies, respectively.

**Figure 7: Ritonavir experiences multiple transitions during the association.** (i) MD snapshots show ritonavir binding to HIVp with open flaps. Each snapshot represents a major conformation in simulation time zone from A to J. The channel the ligand needs to pass through is in green. (ii) Number of water molecules around ritonavir within its first (3.5 Å) and second
(6.0 Å) hydration shell. Gray and white highlight the change in number of water molecules in different simulation time zones. (iii) The ritonavir RMSD and flap distance during the MD simulation.

Figure 8: Comparison of the binding pathway used by xk263 and ritonavir to approach HIVp with an open-flap conformation. Trace of xk263 and ritonavir binding to the open-flap HIVp. Each bead represents the COM of the ligand at 1 ns, and the total simulation time is 200 ns. The ribbon in blue indicates the initial conformation of HIVp, and the cyan tubes indicate the HIVp conformations during a 200-ns MD simulation.

Figure 9: Computed interaction energies between the ligand and water molecules. Coulombic, van der Waals and total energy are shown in black, blue and red lines, respectively.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Figure 7
Figure 8
Figure 9
Graphic for the Table of Contents