UNIVERSITY OF CALIFORNIA
RIVERSIDE

Investigation of Ligand-HIV Protease Dissociation: Pathway, Interactions, Mechanism, and Free Energy Landscape

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Jianan Sun

December 2022

Dissertation Committee:
Dr. Chia-En Chang, Chairperson
Dr. Seán O'Leary
Dr. Min Xue

The Dissertation of Jianan Sun is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I would like to express my earnest, deepest, and sincerest appreciation to my advisor, Dr. Chia-En Chang. Her professionalness and insights have encouraged me to keep dive deeper for each and every scientific problem. Chia-En's support and guidance are indispensable to the development of critical thinking and scientific training for my doctoral degree. I would also like to thank my committee members, Dr. Seán O'Leary, Dr. Min Xue, and Dr. Jeffery Perry. Their questions are valuable for me to think of projects from multiple aspects. I would like to thank my group members for their discussions and advice for my projects. My thanks also go to collaborators for the opportunities that I can apply computational chemistry to multidisciplinary projects. Finally, I want to thank my family and friends for their support.

Published materials are used as follows:

Chapter 2: Sun, J.; Raymundo, M.A.V.; Chang, C.-E.A. Ritonavir and xk263 Binding-Unbinding with HIV-1 Protease: Pathways, Energy and Comparison. Life 2022, 12, 116. https://doi.org/10.3390/life12010116

ABSTRACT OF THE DISSERTATION

Investigation of Ligand-HIV Protease Dissociation: Pathway, Interactions, Mechanism, and Free Energy Landscape

by

Jianan Sun

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, December 2022
Dr. Chia-En Chang, Chairperson

Drugs with desired kinetic properties have better efficacy. Non-covalent small molecule drugs can bind/unbind from their targets, which normally are macro biomolecules. Experimental methods can measure kinetics parameters at ensemble level, but they are unable to track the exact drug binding/unbinding pathways or explain whether drugs show different kinetic behaviors from different pathways. Investigation of ligand binding/unbinding pathways can deepen our understanding of ligand-protein molecular recognition. In this work, possible pathways of ligands-protein were sampled and post-analysis were conducted to investigate ligand binding kinetics (residence time), free energy landscape of ligand unbinding, and key factors that affect binding kinetics.

We are the first utilizing unbiased molecular dynamics to sample a pair of ligands, ritonavir and xk263, unbinding from HIV protease comprehensively, and classify the

unbinding pathways based on the contacts between ligand-protein regions during dissociation. I identify key residues that form hydrogen bond with ligands that results in a meta-stable state for ligand-protein complex. Four distal mutation sites were observed forming interactions with ligands during unbinding, which explains why distal mutations in HIV protease affect drug binding affinity. Molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) was performed to calculate free energy landscape during ligands dissociation. HIV protease backbone conformations during ligands binding/unbinding were compared based on root-mean-square deviation (RMSD). High similarity of protein conformation during ritonavir binding/unbinding suggests that ritonavir follows conformation-selection mechanism during dissociation. In the contrast, low similarity of protein conformation during xk263 binding/unbinding suggests that xk263 follows induced-fit mechanism during dissociation.

Reconstruction of real free energy landscape of ritonavir-HIV protease dissociation was performed using Binding Kinetic Toolkit (BKiT) package. First two principal components (PC) of alpha carbon of protein along with heavy atoms in ritonavir were used as reaction coordinates to guide dissociation path on PC space. Energy barriers on free energy landscape under different unbinding pathways were explained by the molecular recognition of ritonavir-HIV protease. Furthermore, free energy landscape using ligand RMSD were computed, giving a much more accurate binding free energy and residence time approximation but in the lack of detailed ligand-protein interactions.

Table of Contents

FIGURES

CHAPTER 1 INTRODUCTION

1.1 OVERVIEW

Molecular recognition is the fundamental study of intermolecular interactions between multiple molecules or intramolecular interactions within one molecule through physical interaction, such as Van del Waals, hydrogen bonds, enthalpy/entropy etc. Molecular recognition provides insights into protein-protein or drug-protein interactions; hence it's widely applied in biology, chemistry, pharmacology to help people understanding binding mechanism and binding affinity. Computational study of chemical and biological systems explains molecular recognition at atomic level, and it have been ongoing for decades. [1–5] Vast development of computer hardware greatly reduced computational cost and allowed us to sample biomolecular events at nanosecond or event microsecond scale. Enhanced sampling alters free energy landscape to reduce the height difference between barriers and wells which encourages system to evolve much faster when compared to conventional molecular dynamics. Simulation of molecular recognition processes, such as ligand binding/unbinding from its target, can explain factors that determine drug residence time, which couldn't be explained by experimental approaches. This work uses computational chemical theory and technology to deepen the understanding of the kinetic processes involved in the ligand-protein unbinding. Following sections detail the computational and chemical theory that allow for the computational study of molecular dissociation.

Understanding the dissociation process of ligand-protein system requires us to obtain such process using molecular dynamic simulation tools. For a protein with binding

pocket that is exposed to solvent, the ligand binding/unbinding pathway is hard to predict due to the possible directions that ligand can move towards to. an understanding of the chemical and physical behavior at multiple spatial and temporal scales. Studying ligand-protein unbinding pathways can help us identify amino acids that form key interactions to promote molecular recognition between ligand and protein. Chapters 2 details possible ligand dissociation pathways from a protein with open binding pocket and ligand unbinding mechanisms.

1.2 MOLECULAR MECHANICS

1.2.1 Molecular mechanics and potential energy

Molecular mechanics utilizes classical mechanics to sample the motion, interaction, and recognition for molecular systems at atomic level. [6–9] Molecular mechanics typically treat atoms in a biomolecular system as explicit parameterized hard spheres. In all-atom molecular mechanics models, the property of an atom is not only a rigid sphere with a preset radius, but also has an electronic charge which was computed based on atom connections. A potential energy function describes the state of a system at a given time and predict the movement of atoms in the system at next time step. The potential energy function of an atomistic system is defined generally as follows:

$$E = E_{bonded} + E_{nonbonded} \quad \text{eq. 1.1}$$

where E_{bonded} and $E_{nonbonded}$ are defined as follows:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \quad \text{eq. 1.2}$$

$$E_{nonbonded} = E_{electrostatic} + E_{Van\ der\ Waals} \quad \text{eq. 1.3}$$

The potential energy of a system can be divided into five parts: bond energy between two covalently-bounded atoms, angle energy among three covalently-bounded atoms, dihedral energy among four covalently-bounded atoms, electronic energy of an atom and all surrounding atoms, and Val der Waals energy of an atom and all its surrounding atoms. In molecular mechanics, the covalent bond between two atoms can be described using classical harmonic springs with a reference bond length, bond strength and real time distance between such two atoms. The functional form is defined as:

$$E_{bond} = 0.5*k(l_0 - l)^2 \qquad \text{eq. 1.4}$$

where l_0 is the reference bond length, l is the real time distance between such two atoms, and k is the force constant. Force constant for the bond potential energy is majorly determined by the atom types of two covalently-bounded atoms.

The angular potential energy is defined for three neighbor atoms connected by covalent bonds. Angular terms can be expressed as following:

$$E_{angle} = 0.5*k(\theta_0 - \theta)^2 \qquad \text{eq. 1.5}$$

where θ_0 is the reference angle, θ is the real time angle among three atoms, and k is the force constant. Force constant for the angular potential energy is majorly determined by the hybridization and geometry of the center atom of the three atoms set.

The dihedral angles are defined for four atoms connected by covalent bonds. For the four atoms defined by labels a, b, c and d, each dihedral angle Φ is defined as the angle between the two planes defined by atoms a, b, and c and atoms b, c, and d (Figure 1.1).

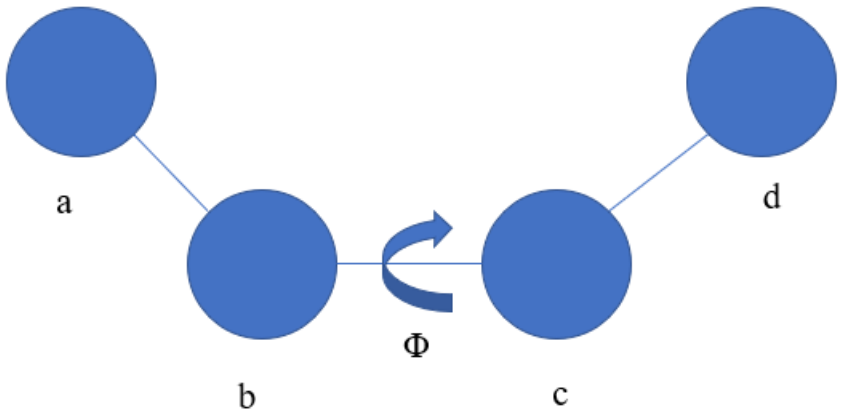


Figure 1.1. Dihedral torsion angle Φ defined between a set of four sequentially bonded atoms a, b, c, and d.

Dihedral angles are critical in determining the three-dimensional geometry of molecules, especially the tertiary structure of protein. The dihedral angular potential can be expressed as the following:

$$E_{dihderal} = 0.5*V_d(1 + \cos(n\Phi_0 - \Phi))^2 \quad \text{eq. 1.6}$$

where V_d is the force constant, n is the periodicity, defining the number of potential minima and maxima with a 360 degree rotation, Φ_0 is the real time dihedral angle among four atoms, and Φ defines the angle at which minima and maxima occur.

Improper dihedral angle, describing the angle of one atom deviates from a plane defined by three other atoms. For four atoms a, b, c and d in which atom b shares a covalent bond with each of a, c, and d, an improper dihedral angle Φ is defined as the angle between the plane of atom a, b, and d and the plane of atoms b, c, and d (Figure 1.2).

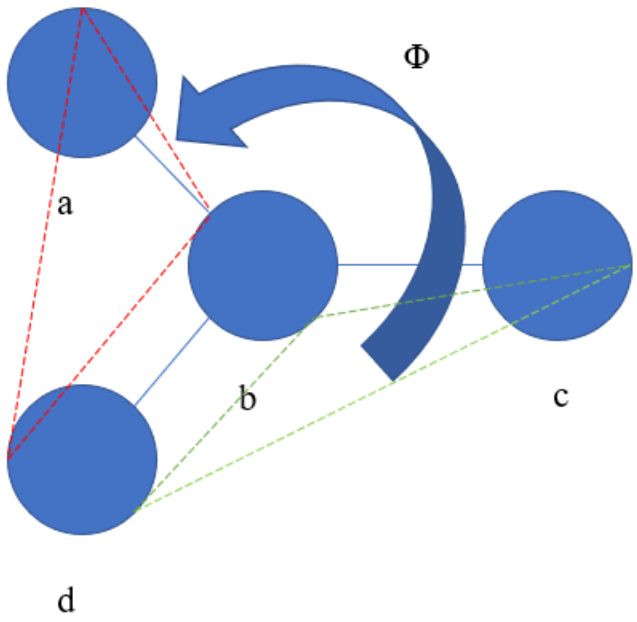


Figure 1.2. Improper dihedral angle Φ between the plane of atom a, b, and d and the plane of atoms b, c, and d.

The improper dihedral potential can be expressed as a harmonic potential as following:

$E_{dihderal,improper} = 0.5*k(1 + \cos(\Phi - \pi))^2$ eq. 1.7

where k is the force constant and Φ is the real time improper dihedral angle between two planes.

Electrostatic and van der Waals forces are considered as nonbonded potentials. Charges are assigned to each atom at the center of atom sphere and each atom only has a partial charge which is obtained from quantum mechanical calculations. [8–10] Electrostatic potential can be expressed using Coulomb interaction potential:

$$E_{electronic} = \frac{q_a * q_b}{4\pi\varepsilon_0 \, r} \quad \text{eq. 18}$$

where q_a and q_b are the partial electron charges of atoms a and b, r is the real time distance between the two atoms, and ε_0 is Coulomb's constant.

Van der Waals potential describes the non-polar component of the nonbonded forces in molecular mechanics. The Van der Waals potential can be expressed as following:

$$E_{Van\ der\ Waals} = \varepsilon_{ab}[(\frac{r_{0ab}}{r_{ab}})^{12} - 2 * \left(\frac{r_{0ab}}{r_{ab}}\right)^{6}] \quad \text{eq. 1.9}$$

where ε_{ab} is the potential energy between two atoms at most stable state, r_{0ab} is the reference atom distance at potential well, r_{ab} is the real time atom distance.

1.2.2 Force fields

Force fields are parameter files containing the specific constants in aid to compute bond, angle, dihedral, electrostatic, Van der Waals and improper potential energies. Parameters are obtained from both experimental data and quantum calculations. Through different methods, the parameters are computed differently and may result in different accuracy of simulation. Common parameter files include General Amber Force Field, ff14Sb, GROMACS, CHARM, OPLS etc. [13–18]

1.2.3 Solvent modeling

Most biomolecular events occur in aqueous environment. Hence providing an environment that mimic natural situation is helpful to achieve accurate simulation results. Solvent model includes explicit water model, implicit water model and organic solvent models. Explicit water model utilizes all-atom method for water molecules, where oxygen and hydrogen atoms are treated as hard spheres. Other than traditional TIP3P water model, which uses 3 points of electronic charges for one water molecule, TIP4P and TIP5P are also useful in simulation process and provide more realistic environment under certain simulation conditions. [22,23] Explicit water model can display the atomistic interactions between model system and water molecules. However, due to the

large number of water molecules required for explicit water model, the computation time may be costly. Poisson-Boltzmann (PB) and Generalized-Born (GB) are most commonly used implicit solvent models. [25-29] Implicit water model applies a force on the model system to mimic the solvent influence on the model system, including the Brownian dynamics collision between water molecules and solutes as well as the electrostatic interactions between solvent and solutes. Implicit solvent model greatly reduced computational time, but in the lack of solute-solvent interactions at atomistic level.

1.3 MOLECULAR DYNAMICS

1.3.1 System preparation

A carefully prepared initial structure is the first step of a successful molecular dynamics simulation. For systems with existing experimental structural data, either from crystallization or NMR structure determination, starting with these experimentally confirmed structure would be best to obtain molecular dynamic simulations results. However, if the structure of a system is unknown, homology modelling such as SWISS-model or AlphaFold is available for structure prediction. [30]

To start a molecular dynamics simulation, we must build a file that contains necessary bond, angle etc. parameters for the potential energy calculation. tLEAP is a program built in AMBER package to help to create prmtop file, which was AMBER specific parameter file. Depending on the molecule type, AMBER has different force field file for protein, DNA, RNA, organic molecules etc. We should choose the correct force field file depending on the system. Choosing correct pH is also crucial in this step because the

protonation state of the side chains of amino acids is dependent on both environmental pH and side chain pKa. Carefully examine the protonation state of polar side chains must be conducted during system preparation.

1.3.2 Minimization

The initial ligand-protein structure may contain bond, angle etc. parameters that are far from equilibrium state. Hence, minimize the system to reach a good structure before simulation is crucial. The steepest descent is the most common algorithm to make the system move along the most negative derivative of potential energy surface to reach a local minimum. It takes significantly long time to reach the lowest point of energy minimum because the derivative approaches to zero as the system reaching local minimum. Another minimization method, conjugate gradient, is applied in system minimization after steepest descent. Conjugate gradient is an iterative algorithm to quickly reach local minimum when the start point is close to the minimum. Combining steepest descent and conjugate gradient is essential for a system to reach local energy minimum quickly and accurately.

1.3.3 Equilibrium

After system solvation, the whole system has only coordinates, not atom velocities. The whole system is still at crystal-packed low energy state. To fully relax the system, we need to slowly increase the temperature by running simulations at different temperatures to bring the system to 298 K or any other desired temperature. At each temperature, we assign random initial atom velocity based on the time clock, a.k.a random number seed, Maxwell-Boltzmann distribution and zero total momentum, followed by molecular

dynamic simulation to bring the system to a stable state at such temperature by reading the system energy in mdout file.

1.3.4 Molecular Dynamics

Molecular Dynamics models system based on molecular mechanics. At a given time and location of an atom, the potential energy can be calculated by the bond, angle, dihedral etc. With given potential energy, E, the force on such atom can be obtained by considering the gradient of the potential energy:

$$F = \frac{E}{dr} \quad \text{eq. 1.10}$$

where F and r are vectors. The acceleration of such atom, a, can be obtained by Newton's law of motion:

$$F = m * a \quad \text{eq.1.11}$$

where F and a are vectors. Using random number seed, we assign initial velocities for all atoms in our system. Hence, we can predict the atom's velocity and position with following equation:

$$v_i = v_0 + a * \Delta t \quad \text{eq.1.12}$$

$$p_i = p_0 + v_i * \Delta t \quad \text{eq.1.13}$$

where v_0 is the initial atom velocity, v_i is atom velocity at next time step, Δt is predefined time step, p_0 is the initial atom position, p_i is atom position at next time step. By repeatedly calculating potential energy, atom position and atom velocity, the system will evolve following molecular mechanic rules, giving us a molecular dynamics trajectory with desired time length. In general, the system is controlled under NVE (constant-number (N), constant-volume (V), and constant-energy (E)), NVT (constant

number (N), constant-volume (V), and constant-temperature (T)) or NPT (constant number (N), constant-volume (V), and constant-pressure (P)) conditions. [31-35]

1.4 REFERENCES

1. Ameijden, M. van. Molecular Dynamics Simulation and Visualisation in a Grid Environment. J. Chem. Phys. 1957.

2. Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Phys. Rev. 1967, 159, 98–103.

3. Stillinger, F. H. Improved Simulation of Liquid Water by Molecular Dynamics. J. Chem. Phys. 1974, 60, 1545.

4. Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. J. Chem. Phys. 1980, 72, 2384.

5. Pal, S.; Fichthorn, K. A. Accelerated Molecular Dynamics of Infrequent Events. Chem. Eng. J. 1999, 74, 77–83.

6. Westheimer, F.; Newman, M. Steric Effects in Organic Chemistry. by MS Newman, John Wiley Sons, New York 1956.

7. Engler, E. M.; Andose, J. D.; Schleyer, P. V. R. Critical Evaluation of Molecular Mechanics. J. Am. Chem. Soc. 1973, 95, 8005–8025.

8. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comput. Chem. 1983, 4, 187–217.

9. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. Comput. Phys. Commun. 1995, 91, 1–41.

10. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. J. Am. Chem. Soc. 1984, 106, 765–784.

11. Bondi, A. Van Der Waals Volumes and Radii. J. Phys. Chem. 1964, 68, 441–451.

12. Alvarez, S. A Cartography of the van Der Waals Territories. Dalton Trans. 2013,

42, 8617–8636.

13. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. Comput. Phys. Commun. 1995, 91, 1–41.

14. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. J. Comput. Chem. 2004, 25, 1157–1174.

15. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. A 1999, 103, 3596–3607.

16. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. J. Comput. Chem. 2010, 31, 671–690.

17. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 2009, 30, 1545–1614.

18. Case, D.; Darden, T.; III, T. C. AMBER 12. Univ. California, San ... 2012.

(19) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. J. Comput. Chem. 2005, 26, 1668–1688.

20. Reimers, J. R.; Watts, R. O.; Klein, M. L. Intermolecular Potential Functions and the Properties of Water. Chem. Phys. 1982, 64, 95–114.

21. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79, 926.

22. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J. Phys. Chem. A 2001, 105, 9954–9960.

23. Zielkiewicz, J. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and TIP5P Models of Water. J. Chem. Phys. 2005, 123, 104501.

24. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. Proc. Natl. Acad. Sci. U. S. A. 2001, 98, 10037–10041.

25. Gruziel, M.; Grochowski, P.; Trylska, J. The Poisson-Boltzmann Model for tRNA: Assessment of the Calculation Set-up and Ionic Concentration Cutoff. J. Comput. Chem. 2008, 29, 1970–1981.

26. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. J. Phys. Chem. A 1997, 101, 3005–3014.

27. Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. Annu. Rev. Phys. Chem. 2000, 51, 129–152.

28 Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. Biopolymers 56, 275–291.

29. Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. J. Comput. Chem. 2002, 23, 1297–13

30. Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-204.

31. Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. Mol. Phys. 1983, 50, 1055–1076.

32. Melchionna, S.; Ciccotti, G.; Lee Holian, B. Hoover NPT Dynamics for Systems Varying in Shape and Size. Mol. Phys. 1993, 78, 533–544.

33. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. J. Chem. Phys. 1994, 101, 4177.

34. Sturgeon, J. B.; Laird, B. B. Symplectic Algorithm for Constant-Pressure Molecular Dynamics Using a Nosé–Poincaré Thermostat. J. Chem. Phys. 2000, 112, 3474.

35. Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-Isobaric Molecular Dynamics Using Stochastic Velocity Rescaling. J. Chem. Phys. 2009, 130, 074101.

CHAPTER 2 RITONAVIR AND XK263 BINDING-UNBINDING WITH HIV-1 PROTEASE: PATHWAYS, ENERGY, AND COMPARISON

2.1 INTRODUCTION

HIV type 1 (HIV-1) garnered enormous attention in the 1970s because it can at-tack CD4 cells and weaken the immune system, eventually causing acquired immuno-deficiency syndrome (AIDS) if not suppressed in vivo. HIV protease (HIVp) is one of the essential proteins in the HIV life cycle, responsible for cleaving premature protein and producing fully functional enzymes [1-3]. Protease inhibitors (PIs) can disrupt HIVp function and stop HIV replication via competitive inhibition [4,5]. However, PI-selected mutations decrease the HIVp susceptibility to inhibitors and induce cross-resistance among PIs [6-8].

HIVp is a good model system to study ligand–protein binding/unbinding because HIVp is a homodimer with well-defined regions. Flap regions in HIVp must open in order to bring the peptide into the binding pocket [9,10]. Then flaps close. Four resi-dues, Asp25, Asp124, Ile50 and Ile149, are essential for holding the peptide in place, with Asp25 and Asp124 catalyzing proteolysis [11]. Such well-defined structural be-havior and ligand–protein interaction features are informative in investigating protein conformational change during ligand–protein binding/unbinding.

Non-covalently binding drugs can reversibly bind and unbind from their targets. Many drugs forfeit their pharmacological effect once they leave the target site. In vivo and in silico studies suggest that pharmacological activity depends on the drug life-time [12-14]. Binding (kon) and unbinding rate constants (koff) are 2 factors used to de-scribe the

kinetic process of the drug–protein interaction. Ideally, we want a drug with large kon (faster binding) and small koff (slower leaving). Residence time, the reciprocal of the unbinding rate constant, describes the lifetime of a ligand in the bound state with its target protein. Developing drugs with long residence time can enhance the pharmacological effect. All-atom simulations of ligand–protein dissociation can pro-vide mechanistic information at the atomistic level [15,16]. Investigating the interactions between the drug and protein during unbinding may provide insights into drug design and altering drug residence time.

Several recent studies have focused on the binding kinetics of drug–HIVp sys-tems [17-20]. Enhanced sampling methods, such as steered molecular dynamics (SMD), metadynamics and accelerated MD (aMD), have been used to sample ligands binding or unbinding from HIVp [21-24]. Binding kinetics explain the dynamic processes of ligand binding/unbinding, not merely the difference between the bound state and free state. Two binding mechanisms, induced-fit and conformation-selection, have been proven in experimental and computational studies [25-27]. Different ligands show distinct binding kinetics [28,29]. For example, ritonavir requires HIVp conformational change to bind, whereas xk263 can bind to HIVp without obvious protein mo-tions [27]. The free energy profile was constructed computationally to reveal relative free energies between various intermediate, free and ligand bound states for ligand–HIVp binding or unbinding [30-32]. However, because of the complex dynamic nature of HIVp when binding to ligands with diverse properties, we still lack comprehensive studies explaining the possible unbinding pathway(s) and protein conformational change during ligand unbinding from HIVp.

Here, we studied processes of unbinding of a pair of ligands, ritonavir and xk263, from HIV-1 protease. Investigation into how differently these 2 ligands interact with HIVp provides insights into factors that affect the unbinding processes. We classify unbinding trajectories and discuss ligand unbinding pathways, intermolecular H-bond networks and HIVp conformations during the dissociation processes. Root-mean-square deviation (RMSD)-based analysis was used to identify overlapping HIVp conformations between xk263 or ritonavir binding and unbinding. We also dis-cuss the ligand unbinding mechanisms and compare these with ligand–protein binding models.

2.2 MATERIALS AND METHODS

2.2.1 Target molecular systems

We selected Protein Data Bank entry 1HXW and 1HVR to study ritonavir and xk263 unbinding from HIVp, respectively (Fig. 1b, c, d) [33,34]. 1HXW contains a ritonavir–HIVp crystal structure in the bound state, with a single protonation state applied for Asp25. 1HVR contains an xk263–HIVp crystal structure in the bound state, with a double protonation state for Asp25/124. We applied distinct protonation states for the 2 systems based on the results of the lowest interaction energy between the ligand and 4 different protonation states of HIVp [35].

2.2.2 MD simulations

We performed 100-ns conventional MD (cMD) simulation for ritonavir–HIVp and xk263–HIVp followed by aMD simulations of 25 seeds for each system with a re-seeding approach, as described below.

The Amber18 package with GPU implementation was used for unbiased all-atom MD simulations [36]. The Amber FF99SB and General Amber Force Field (GAFF) were used for HIV protease and ligands, respectively [37]. VCharge was used to assign partial charges for the ligand atoms for better charge accuracy of resonance structures of phenyl groups in ritonavir and phenyl and naphthalenyl groups in xk263 [38]. Five chloride ions and 6 chloride ions were placed to maintain a neutral system for ritonavir–HIVp and xk263–HIVp, respectively. Minimization of hydrogen atoms, the side chains and the entire system was applied for 1000, 1000 and 3000 steps, respectively. The system was solvated in a rectangular TIP3P water box and the edge of the box was at least 10 Å away from the solutes. The system went through 1000 steps of water minimization and 5000 steps of whole-system minimization to correct any inconsistencies. Then the system was slowly heated to 50, 100, 150, 200, 250 K for 200 ps at each temperature and 9 ns at 300K. Then 100-ns cMD was performed in an isothermic−isopressure (NPT) ensemble to ensure that the system reached equilibrium. Langevin Thermostat along with a damping constant of 2 ps-1 was used to maintain a temperature of 310K. Particle mesh Edward (PME) was used to compute the long-range electrostatic interactions > 10 Å [39]. Frames were saved every 10 ps with a time step of 2 fs. Finally, the SHAKE algorithm was used to constrain the covalent bonds involving hydrogen atoms [40].

2.2.3 Accelerated MD (aMD) simulation

Biomacromolecules can have high energy barriers when a molecule is moving around the energy surface, which prevents efficient conformation sampling in cMD. Therefore, we used the final frame from the cMD run as the initial conformation for aMD simulations

with a different starting velocity for each seed. aMD enhances the conformational sampling by adding a continuous non-negative bias or "boost" potential ΔV(r) to the potential energy V(r) whenever the potential energy takes values below some predetermined threshold E [41]. A bias may also be applied to specific components of the potential energy, for example, the dihedral energy V_D (r). Energy boosts to both the total potential energy and the dihedral energy are applied here. The overall bias potential is given by

$$\Delta V(\boldsymbol{r}) = \frac{\left(E_P - V(\boldsymbol{r})\right)^2}{\alpha P + \left(E_P - V(\boldsymbol{r})\right)} + \frac{\left(E_D - V_D(\boldsymbol{r})\right)^2}{\alpha D + \left(E_D - V_D(\boldsymbol{r})\right)} \quad \text{eq. 2.1}$$

where Ep and ED are the average total potential energy threshold and average dihedral energy threshold based on cMD simulation. αP and αD are boost factors for the total potential and dihedral potential, respectively. The actual values of Ep, ED, αP and αD are listed in supporting information (Table S1). By increasing the value of the tuning parameter α, we can reduce the energy boost, ΔV(r), thus allowing the conformational change of protein to evolve faster than in cMD yet maintain the protein's secondary structure. Protein backbone dihedral angle change directly leads to protein conformational change. Hence, we applied both dihedral and total potential energy boost to our model systems. AMD input parameters can be found in Supporting Information.

2.2.4 Re-seeding approach

Ritonavir and xk263 are both tight binders to HIVp. To efficiently explore their unbinding pathways, we applied a re-seeding approach under aMD simulation. Re-seeding refers to generating multiple MD simulation production runs from the same initial ligand–protein conformation but with different atom velocities. The last frames

from cMD simulations of ritonavir–HIVp and xk263–HIVp were used to provide an initial conformation for 25 ritonavir-HIVp first re-seedings and 25 xk263-HIVp first re-seedings. Each first re-seeding is a 400-ns-long aMD simulation. Five 25-ns-long second re-seedings were applied with the last frame of the first re-seeding as an initial conformation if unbinding did not occur in the 400-ns aMD simulation. The third re-seeding was performed to generate five 25-ns-long aMD simulations using the frame from the second re-seeding in which the ligand reached the largest RMSD as compared with the fully relaxed ligand position in the bound state. The following re-seedings were performed with the same method as generating third-layer re-seeding. If the selected initial frame contained a ligand–HIVp conformation that presented a strong tendency to dissociate, up to twenty 25-ns-long aMD simulations would be generated from such frame. A detailed example of the re-seeding approach can be found in Supplementary Information Figure S1. Re-seeding was repeated until successful ligand dissociation, after the 17th re-seeding, or protein unfolding/distortion. Table S2 lists re-seeding attempts for ritonavir and xk263 dissociations.

2.2.5 Hydrogen bond analysis

Each hydrogen bond between the ligand and protein may contribute up to a few kcal/mol to interaction energy [42]. To hydrolyze the peptide bond in premature protein, Asp25/124 establishes hydrogen bonding with the peptide backbone in the substrate, whereas Ile50/149 stabilizes the substrate with a water bridge [35]. To better understand which residues may form H-bonds with ritonavir/xk263 during ligand dissociation, we analyzed and plotted H-bond versus time by using the CPPTRAJ program [36,43]. The

criteria for H-bonding are 1) the distance between donor D and acceptor A < 3 Å and 2) the D-H-A angle, where H is the shared hydrogen, at least 150°.

2.2.6 MM/PBSA interaction energy

To investigate the interaction energy between the ligand and HIVp, we used the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method [44]. Interaction energy, ΔE, is calculated by

$$\Delta E = E_{PL} - E_P - E_L \quad \text{eq. 2.2}$$

where E_"PL" ,E_"P" and E_"L" represent enthalpy of the ligand–protein complex, enthalpy of protein and enthalpy of ligand, respectively.

$$E_{MM/PBSA} = E_{gas} + G_{solvation} \quad \text{eq. 2.3}$$

E_gas is gas-phase energy calculated from the molecular mechanical force field, G_solvation is the solvation energy calculated by solving the Poisson-Boltzmann equation. E_(MM/PBSA) can be further decomposed as

$$E_{MM/PBSA} = E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{elect} + (G_{pb} + G_{np}) \quad \text{eq. 2.4}$$

where E_bond is bonded energy, E_angle is angle energy, E_torsion is dihedral energy, E_vdW is van der Waals (vdW) energy, E_elect is electrostatic energy, G_pb is polar solvation energy and G_np is non-polar solvation energy. The bonded terms were canceled in interaction energy calculations because E_"PL" ,E_"P" and E_"L" used same frame of ligand–protein conformation. We can rewrite

$$E_{MM/PBSA} = E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{elect} + (G_{pb} + G_{np}) \quad \text{eq. 2.5}$$

$$= \Delta E_{polar} + \Delta E_{non-polar} \quad \text{eq. 2.6}$$

where ΔE_polar is polar interaction energy (ΔE_elect + ΔG_pb) and ΔE_(non-polar) non-polar interaction energy (ΔE_vdW + ΔG_np). Entropy contributions are neglected in MM/PBSA calculations. Polar, non-polar and total interaction energies are calculated with a 1-ns interval. Energies are averaged using 3 computed energy values, 1 ns before and after a reference frame (average 3 ns). The average energy at 0 ns is the mean of 0 ns and 1 ns and the average energy at the last timestep is the average of the last and second last timestep.

2.2.7 RMSD-based dissociation-association trajectory comparison

If the backbone RMSD of 2 HIVp conformations are within 2.5 Å, we define them as similar conformations or overlapping conformations. Under such criteria, we iteratively compared HIVp conformations in dissociation trajectories with association trajectories. Because pathway A is the most popular dissociation path for both ligands, we selected one ritonavir dissociation trajectory and one xk263 dissociation trajectory under pathway A that are discussed in the following context. We selected 8 ritonavir association trajectories and 6 xk263 association trajectories from previous work [27]. Frames were saved every 1 ps in our original trajectories, and we re-saved a frame every 0.1 ns for each trajectory. Frames from the representative trajectory from Pathway A were re-saved every 1 ns as reference structures (every 10 frame) to ensure that the frame count figure was succinct.

Using regions of HIVp from the crystal structure (i.e., regions containing residues 1 to 41, 61 to 130 and 160 to 198, flap regions excluded) from PDB entry 1HXW as a reference frame, we aligned all ritonavir association/dissociation trajectories to clearly

observe flap motions. Using 1-ns interval frames in ritonavir dissociation as reference, we calculated HIVp backbone RMSD for all frames in each of 8 association trajectories, then plotted a bar graph, similar frame count versus each reference frame, for visualizing the comparison result. The same analysis was conducted for the xk263–HIVp system using partial 1HVR as a reference frame for alignment.

2.3. RESULTS

Figure 2.1a illustrates a schematic of non-covalent binding free energy, with 2 mol-ecules binding and unbinding when a system reaches equilibrium. Although binding and unbinding may share highly similar paths, ligand unbinding may not need to fol-low the same path as it binds to the protein target. Here we first classified unbinding pathways for 2 ligands with different chemical properties and then compared the as-sociation and dissociation processes to further understand the binding and unbinding processes.

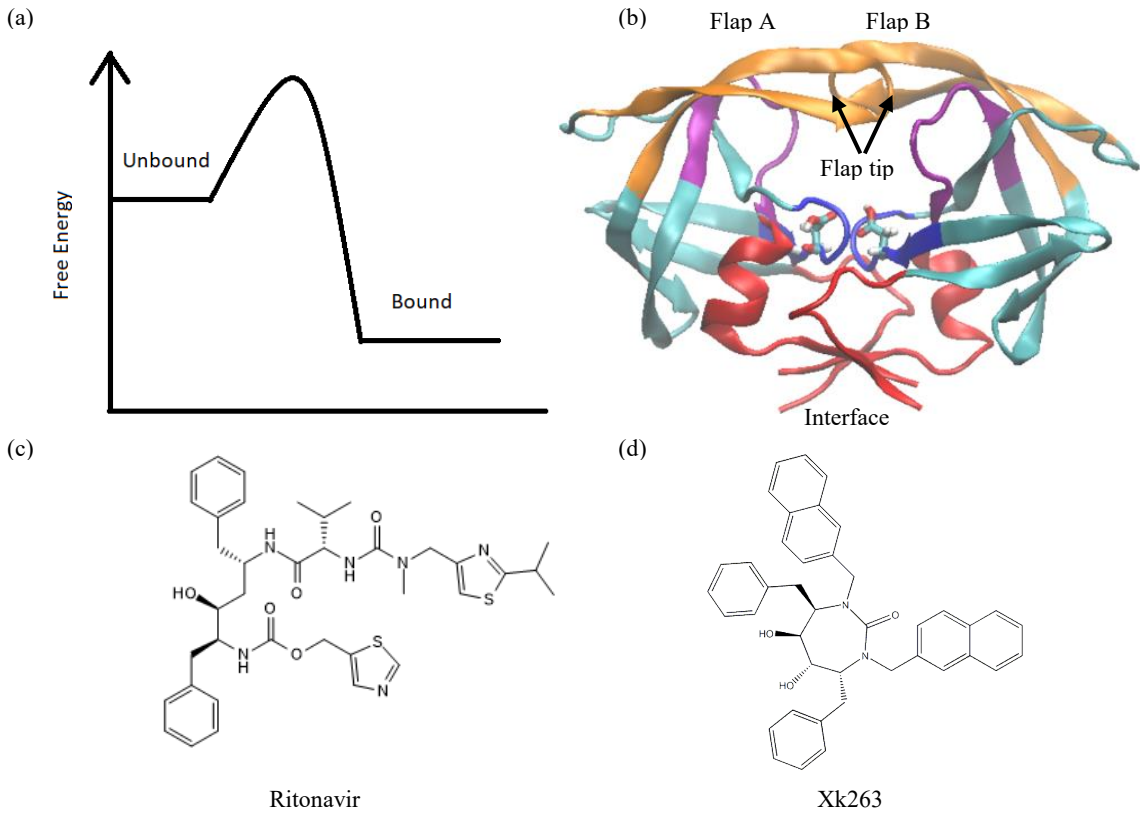


Figure 2.1. (a) Free energy difference between ligand–protein unbound state and bound state. (b) HIV protease (HIVp) structure with color-coded regions: flap (orange, flap a: residues 42-60, flap b: residues 141-159, flap tips: residues 49-52 and 148-151), loop (pink, residues: 75-85 and 174-184), catalytic triads (blue, residues: 23-29 and 122-128) and dimer interface (red, residues: 1-9, 86-99, 100-108 and 185-198). Catalytic residues, Asp 25 and Asp 124 are shown in licorice. Two-dimensional structure of (c) Ritonavir and (d) xk263.

Using the re-seeding approach described in Methods, we obtained 20 and 15 un-binding trajectories for ritonavir and xk263, respectively. We grouped the trajectories into 4 unbinding routes based on where a ligand diffuses after leaving the binding pocket of HIVp─pathway A unbinding between the flap and loop region; pathway B: surface diffusion through the flap region; pathway C: surface diffusion through the in-terface region; and others (Table 1 and Fig. 2). Both ritonavir and xk263 preferred to dissociate

from HIVp under pathway A because in 11 of 20 dissociation trajectories and 8 of 15 dissociation trajectories, ritonavir and xk263 followed the flap/loop un-binding pathway, respectively. We reported binding pocket openness and flap motions using the distance between Cα of Ile 50-Ile 149 (flap tip), Asp 124-Ile 50 (flap A open-ness), Asp 25-Ile 149 (flap B openness). Figure 3 also shows ligand RMSD (ritonavir/xk263) and RMSF of HIVp for easy comparison between molecular motions, pathways and different ligands. We performed H-bond analysis and intermolecular energy calculation by using MM/PBSA during ligand unbinding [43]. We also exam-ined conformations from our dissociation pathways with reported association pathways to reveal popular HIVp conformations sampled during binding/unbinding.

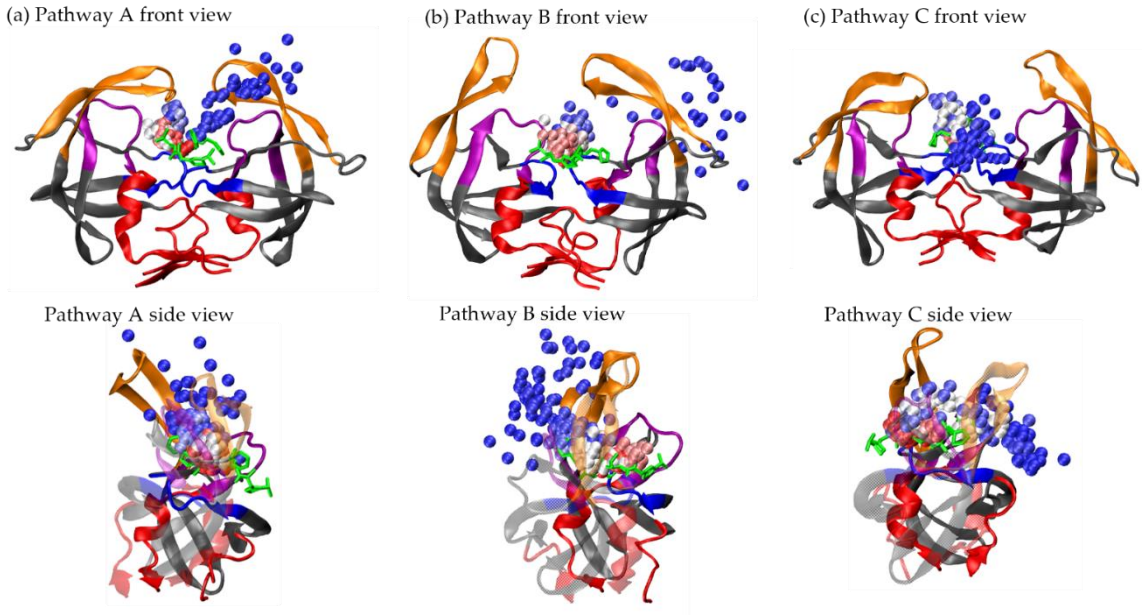


Figure 2.2. Dissociation pathways of HIVp and ritonavir sampled by accelerated molecular dynamics (aMD). Each bead represents a position of center of mass of ritonavir with 1-ns interval during dissociation, with total simulation time of 248.7 ns, 304.9 ns and 543.1 ns for pathways A (a), B (b), and C (c), respectively. Color beads present center of mass of ritonavir taken from frames in the beginning (red), middle (white) and near the end (blue) of the trajectory. Ritonavir's initial position is green licorice. Flap region (orange), loop region (purple), catalytic triads (blue) and interface re-gion (red) are colored for better visualization. The HIVp conformation is taken from the final frame in each dissociation trajectory.

2.3.1 Ligand unbinding pathways

Here, we discuss the pathways in detail to understand the similarities and differences between ritonavir and xk263 during unbinding.

2.3.1.1. Pathway A: Dissociation between flap/loop region

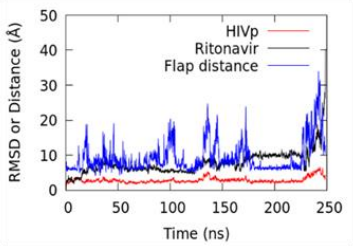
Pathway A is the major unbinding pathway for both ligands because 11 of 20 ritonavir dissociation trajectories and 8 of 15 xk263 dissociation trajectories presented unbinding through the flap/loop region of HIVp (Fig. 2.2a). Flaps opened and closed repeatedly during ligand dissociation because the distance between Asp 124 and Ile 50 of flap A and

Asp 25 and Ile 149 of flap B fluctuated constantly. Flap motions were asymmetrical during dissociation because flap A and flap B opened at different times of the simulation. For example, at 140 ns, flap A motion was correlated with increased ritonavir RMSD, and at 165 ns, the ritonavir dissociation was followed by flap B open-ing (Fig. 3A,i). Previous studies showed that free HIVp has flaps that open/closed spontaneously, and the same behavior was preserved when in the ritonavir bound complex [27,45,46]. For example, flap B opening at 100 ns did not directly result in ri-tonavir displacement or rearrangement. At the near end of ritonavir or xk263 dissocia-tion, the protein underwent a wide-open handedness, which allowed the ligand to un-bind from HIVp. RMSF analysis of HIVp during ligand dissociation revealed that the flap region experienced larger motion with ritonavir than xk263 dissociation (Fig. 3A, iii). H-bonds between both ligands and the loop region were essential to stabilize lig-ands during unbinding, and yet a strong H-bond may be a disservice for ritonavir dis-sociation. As shown in Figure 4b, ritonavir can form an H-bond with Thr 80 in the loop region for a long simulation time (Fig. S2), which strengthens the ritonavir-HIVp in-teraction and might prevent dissociation, whereas the H-bond between xk263 and Thr 80 was rarely observed. However, it took much longer for xk263 to leave the binding pocket because of the strong H-bond network. Xk263 must break the H-bond network formed from both the flap tip and catalytic triads to unbind. The H-bond network re-striction and the rigidity of xk263 led to smaller ligand motion on unbinding with ~5 Å RMSD from 100 ns to 400 ns as compared to ~8 Å RMSD for ritonavir from 20 ns to 215 ns (Fig. 3A,i,ii). When we restarted simulations using reseeding strategy, xk263 and ritonavir frequently returned to

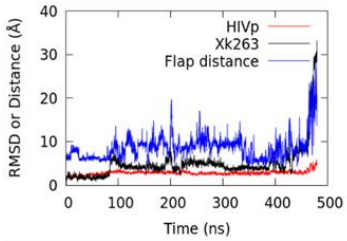
the binding pocket even when already in the flap/loop region. Only when the ligand passed the gap between the flap/loop, it quickly solvated and left HIVp.

(A) Pathway A

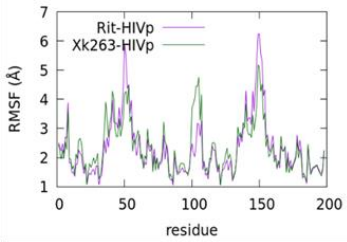
(i) Ritonavir



(ii) Xk263

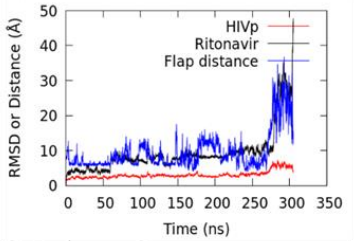


(iii) RMSF of HIVp

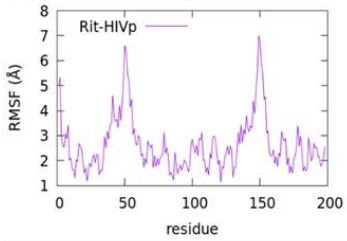


(B) Pathway B

(i) Ritonavir

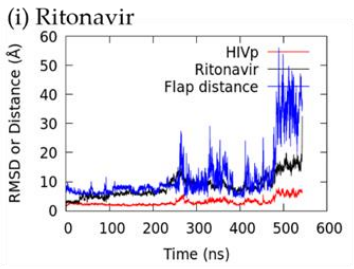


(ii) RMSF of HIVp

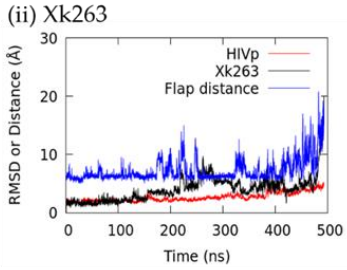


(C) Pathway C

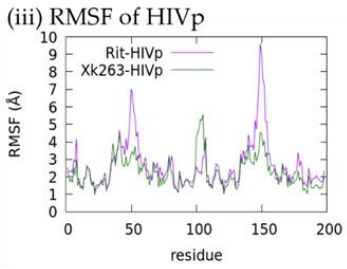
(i) Ritonavir



(ii) Xk263

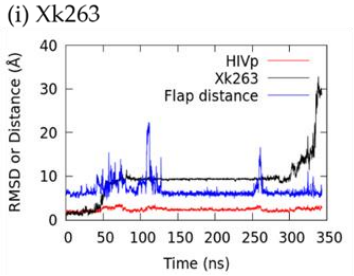


(iii) RMSF of HIVp



(D) Other Pathway

(i) Xk263



(ii) RMSF of HIVp

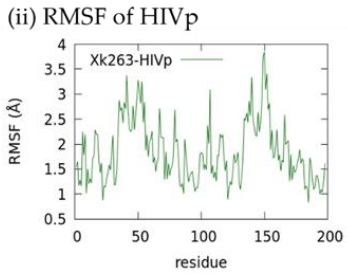


Figure 2.3. Binding pocket openness, ligand RMSD and HIVp RMSD ((A,i and ii), (B,i), (C,i and ii) and (D,i)) as well as HIVp RMSF((A,iii), (B,ii), (C,iii) and (D,ii)) under different pathways. We use flap tip distance (αC Ile 50 to αC Ile 149) to describe flap behavior and binding pocket openness. Dis-tance, RMSD and RMSF are in angstroms. Here we computed the interaction energy to further understand the dissociation processes and the intermolecular attractions. Although ritonavir has more H-bond donor/acceptor atoms, the drug does not have stronger polar attraction with HIVp as compared with xk263. The polar interactions consider both electrostatic attraction and PB solvation free energy. Our calculations show that non-polar attraction is the dominant energy term during ligand dissociation for both ritonavir and xk263 (Fig. 4c, Fig. 5g, Fig. 6f, Fig. 7e, Fig. 8f and Fig. 9f).

In the beginning of the aMD run (Fig. 4a), the interaction energy between ritonavir and HIVp was -76.11 kcal/mol (non-polar: -113.46 kcal/mol, polar: 37.35 kcal/mol). Notably, we focused on intermolecular interactions. The solute conformational energy and the entropic effects are not included here. As ritonavir rearranged and moved toward one side of the flap/loop region of HIVp, the system reached local energy minima of -107.48 kcal/mol (non-polar: -124.43 kcal/mol, polar: 16.95 kcal/mol) at 34 ns because of increased contact area and the H-bond formation between hydroxyl groups in Thr 80 and ritonavir (Fig. 4b). The H-bond length was 1.67 Å at 34 ns and remained so until 228 ns (Fig. S2), and flap A and B opened and closed repeatedly. Ritonavir also formed an H-bond with Asp 29 and Arg 107 from 1 ns to 170 ns, which stabilized ritonavir, and the RMSD of ritonavir was maintained within 8 Å as compared with its bound-state position (Fig. S2, Fig. 3a). At 172 ns, HIVp rearranged to a wide-open conformation, which kept the ritonavir–Thr 80 H-bond (1.77 Å) but significantly weakened the non-polar intermolecular interactions, thus resulting in an energy barrier of -68.72 kcal/mol (non-polar: -91.49 kcal/mol, polar: 22.77 kcal/mol) (Fig. 4c). As ritonavir migrated from chain A (Fig. 4c) to chain B (Fig. 4d), an energy barrier of -51.05 kcal/mol (non-polar: -63.16 kcal/mol, polar: 25.62 kcal/mol) occurred because of the reduced contacts between the 2 molecules. Once ritonavir contacted flap B at 234 ns (Fig. 4e), it moved to the gap between the flap/loop region and was temporarily stabilized with interaction energy of -52.76 kcal/mol (non-polar: -70.88 kcal/mol, polar: 18.12 kcal/mol). Then ritonavir moved outward and completely re-solvated. The final interaction energy at 250 ns was 0.65

kcal/mol (non-polar: 0.53 kcal/mol, polar: 0.12 kcal/mol), indicating complete ritonavir dissociation from HIVp.

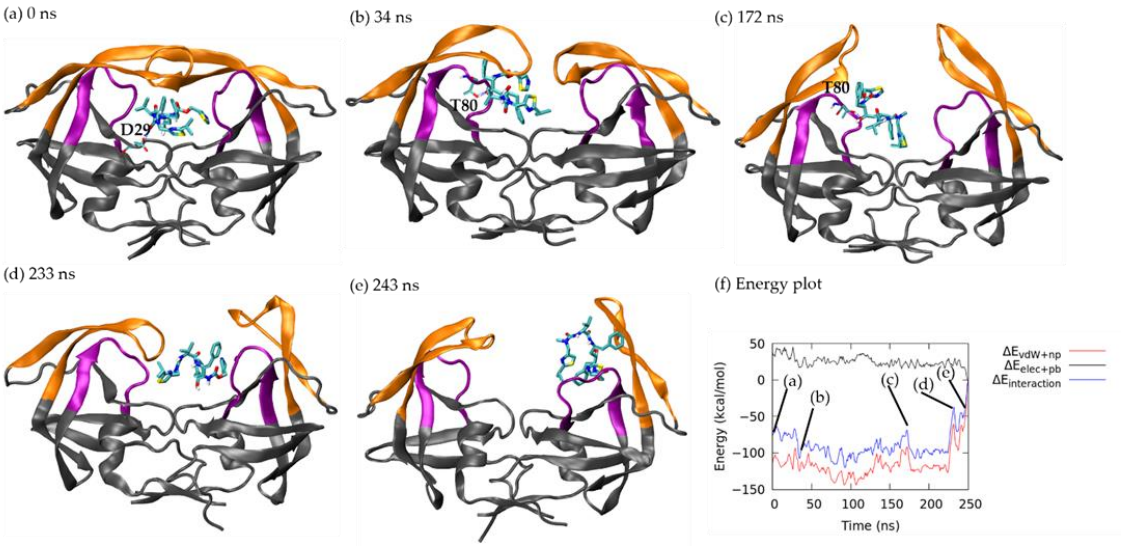


Figure 2.4. Ritonavir dissociation under pathway A. Flap and loop are in orange and purple, respectively. Ritonavir is in licorice. Hydrogen bond (H-bond) is shown in dashed blue line with the corresponding residue labelled. (a) 0 ns. (b) 34 ns. (c) 172 ns. (d) 233 ns. (e) 243 ns. (f) Non-polar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between ritonavir and HIVp during unbinding.

Different from ritonavir interacting with the loop region during dissociation, xk263 spent most of the time in the bound state, owing to the strong H-bond network with flaps and catalytic triads (Fig. 5a and Fig. S3). Such an intermolecular H-bond network was mainly restricted to residue sAsp 25, Ile 50, Asp 124 and Ile 149 because the H-bond length was 1.85 Å for xk263–Asp 25 and 1.78 Å for xk263–Ile 50. Interaction energy between xk263 and HIVp at 0 ns was -92.03 kcal/mol (non-polar: -123.27 kcal/mol, polar: 31.24 kcal/mol), which is 15.92 kcal/mol stronger than that of ritonavir in the bound state. During unbinding, xk263 tilted at 86 ns, which led to decreased interaction between xk263 and HIVp as flap A lifted and the diol group was released from the catalytic triads,

thus resulting in an interaction energy barrier of -83.84 kcal/mol (non-polar: -107.58 kcal/mol, polar: 24.15 kcal/mol) (Fig. 5b). Even though xk263 moved back to the binding pocket at 104 ns (Fig. 5c), the flap handedness was disrupted, with only one flap contacting xk263 and the ligand RMSD increased (Fig. 3A,ii). Xk263 was trapped inside the pocket until 440 ns, when xk263 moved toward chain B and formed an H-bond with Pro 180 and Val 181 with interaction energy of -62.94 kcal/mol (non-polar: -82.87 kcal/mol, polar: 19.93 kcal/mol) (Fig. 5d) and H-bond length of 1.77 Å and 1.70 Å, respectively. Finally, flap B opened (Fig. 5e), which allowed xk263 to shift into the flap/loop region, followed by both flaps opening and xk263 continuing to unbind (Fig. 5f).

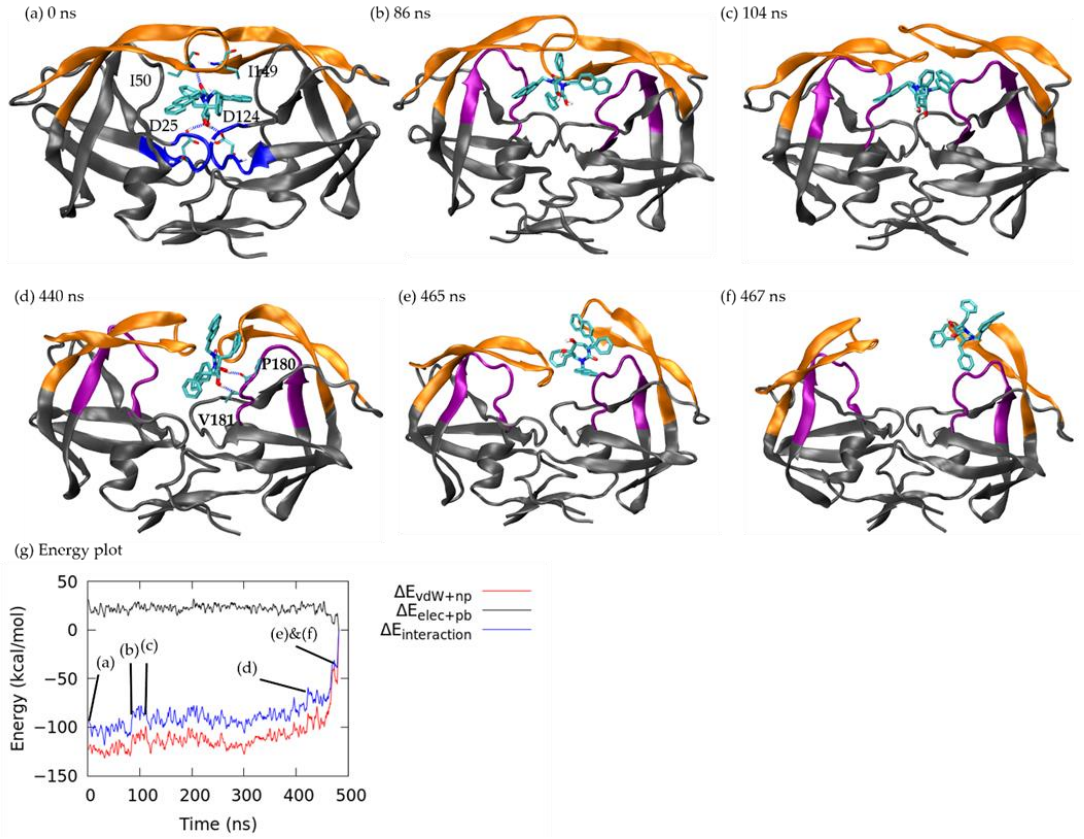


Figure 2.5. Xk263 dissociation under pathway A. Flap and loop are in orange and purple, respectively; xk263 is in licorice. H-bond is shown in dashed blue line with the corresponding residue labelled. (a) 0 ns. (b) 86 ns. (c) 104 ns. (d) 440 ns. (e) 465 ns. (f) 467 ns. (g) Nonpolar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between xk263 and HIVp during unbinding.

2.3.1.2. Pathway B: Dissociation with surface diffusion through the flap region

Pathway B is defined as the ligand unbinding along the flap without contacting the loop region (Fig. 2b). Four of 20 ritonavir and zero xk263 dissociation trajectories used this pathway. Xk263 has 4 aromatic groups, and the hydrophobicity always drives the ligand to contact with the loop region. Similar to pathway A, during ritonavir dissociation under Pathway B, flaps fluctuated spontaneously, which can be irrelevant to the position of ritonavir and not directly induced by ritonavir (Fig. 3B,i). However, ritonavir moved with

one flap as the flaps moved open/closed at ~280 ns, thus resulting in increased ligand RMSD fluctuation. Because it moved with a flap, ritonavir also underwent more noticeable rotation (Fig. 6e) as compared with the conformations found in Pathway A. Nevertheless, ligand rotation did not affect protein motions, and the RMSF of HIVp in both pathways A and B had major fluctuations from the flap region (Figs. 3Aiii and 3Bii).

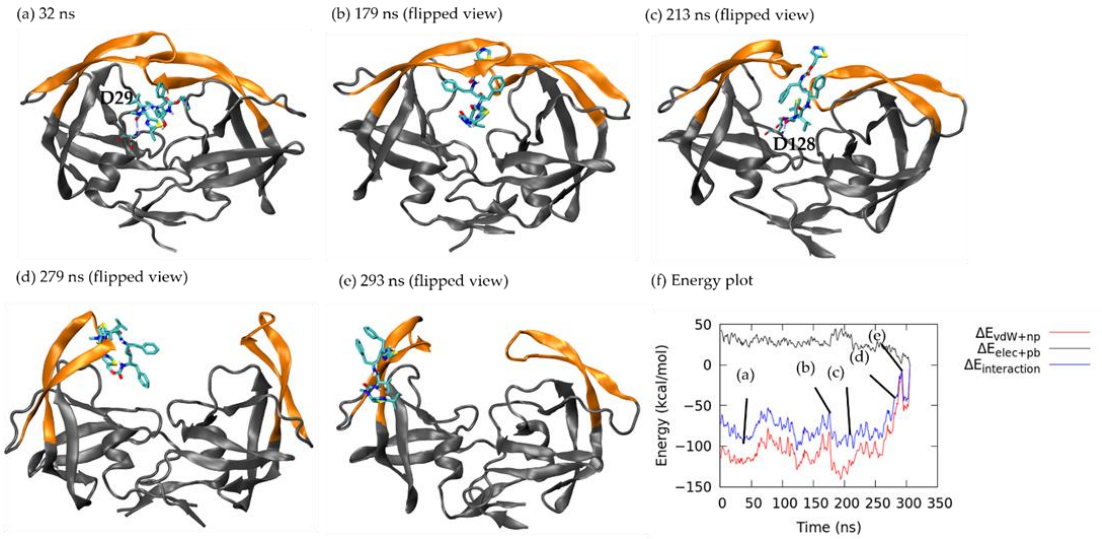


Figure 2.6. Ritonavir dissociation under pathway B. Flap is in orange. Ritonavir is in licorice. H-bond is shown in dashed blue line with the corresponding residue labelled. (a) 31 ns. (b) 179 ns. HIVp and ritonavir were horizontally flipped in (c) 213 ns, (d) 279 ns and (e) 293 ns for better visualization. (f) Non-polar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between ritonavir and HIVp during unbinding.

In the beginning of the simulation (see the same conformation as Fig. 4a), a stable H-bond between ritonavir and Asp 29 (bond length 1.80 Å) remained until 75 ns (Fig. S4). The energy fluctuated when ritonavir stayed close to the crystal structure bound complex (Fig. 3B,i before 55 ns). At 179 ns (Fig. 6b), ritonavir moved in between flaps A and B, where a H-bond with Asp 29 and subsequently with ILE50 was broken, thus resulting in

a slightly increased interaction energy of -81.72 kcal/mol (non-polar: -117.74 kcal/mol, polar: 36.03 kcal/mol). Ritonavir continued wiggling, and at 213 ns, the drug formed a transient H-bond with Asp 128 (length 1.86 Å) to create a local energy minimum by reducing the polar interaction (total energy: -102.52 kcal/mol; non-polar: -121.91 kcal/mol, polar: 19.93 kcal/mol). At 279 ns, the flaps were fully open, and ritonavir maintained contacts with flap A only, which significantly weakened the non-polar interaction energy (total: 57.33 kcal/mol; non-polar: -64.58 kcal/mol, polar: 18.51 kcal/mol) (Fig. 6d). Then ritonavir migrated to the outer side of flap region, partially solvated in water and raised the interaction energy to -5.12 kcal/mol (Fig. 6e) followed by further ligand rearrangement and complete unbinding.

2.3.1.3. Pathway C: Dissociation with surface diffusion through interface region

Pathway C is defined as a ligand diffusing on the interface region and then unbinding from HIVp (Fig. 2c). One unique HIVp motion in this pathway is that the flaps open widely at first and then ligand dissociation occurs. Dissociations under pathway C were observed in 3 of 20 ritonavir trajectories and 3 of 15 xk263 trajectories. It took longer time (543 ns) for ritonavir to unbind from HIVp under pathway C as compared with pathway A (249 ns) or pathway B (305 ns), whereas xk263 spent a similar time dissociating under pathways A and C. Wide-open flap handedness was observed before ligand unbinding in both ritonavir and xk263 dissociation under this pathway (Fig. 7c and Fig. 8d). For example, ritonavir shifted to the catalytic triad and diffused along the interface region when flap tip distance was > 30 Å after 479 ns (Fig. 3C,i). Similarly, xk263 moved to the interface region with noticeable flap openness after 453 ns (Fig.

3C,ii). Even after leaving the binding pocket, ligands may remain in the interface region for a significantly long time, which is consistent with existing studies that this area is a highly favorable non-specific binding site for ligands [47]. As a result, only 1 of 3 ritonavir trajectories and 1 of 3 xk263 trajectories achieved complete dissociation from HIVp under pathway C. Notably, because no stable H-bonds were observed when a ligand stays in the interface region (Fig. S5 and Fig. S6), the non-specific attraction was mostly non-polar interactions.

Initially, the crystal structure bound complex (Fig. 4a) had a stable H-bond between ritonavir and Asp 29, but the H-bond broke at ~40 ns. Two new H-bonds between Ile 50 and Arg 107 formed, which strengthened the ΔE_polar and lasted 50 ns (Fig. 7a and S4). However, ΔE_(MM/PBSA)kept increasing because vdW contacts were reduced when ritonavir gradually moved away from the crystal structured bound form until 53 ns, when ritonavir interacted with 5 non-polar residues (Ile 50, Ile 84, Val 131, Pro 180 and Val 181 in Fig. 7a) to bring the system to a local energy minimum (total energy: -100.25 kcal/mol; non-polar: -119.97 kcal/mol, polar: 19.38 kcal/mol). Ritonavir stayed in the region until 245 ns (Fig. 7b), when the drug curled within the binding pocket and reduced the contacts with HIVp, thus yielding an energy barrier of -61.14 kcal/mol (non-polar: -97.64 kcal/mol, polar: 36.50 kcal/mol). Figure 7c shows that at 434 ns, both flaps opened, and ritonavir arranged to an extended form, thus resulting in a local energy minimum of -100.63 kcal/mol (non-polar: -125.02 kcal/mol; polar: 24.40 kcal/mol). The flaps continued opening widely and ritonavir formed contacts with only the interface region (Fig. 7d), which quickly increased interaction energy to -52.10 kcal/mol (non-polar: -

61.47 kcal/mol, polar: 9.36 kcal/mol). Ritonavir stayed in the area for ~60 ns before it eventually detached from the interface region as interaction energy reached ~0 kcal/mol. In the other 2 ritonavir dissociations under pathway C, ritonavir may attach to the interface region for longer than ~125 ns (equivalent to 5 re-seeding attempts) and yet no complete unbinding was observed.

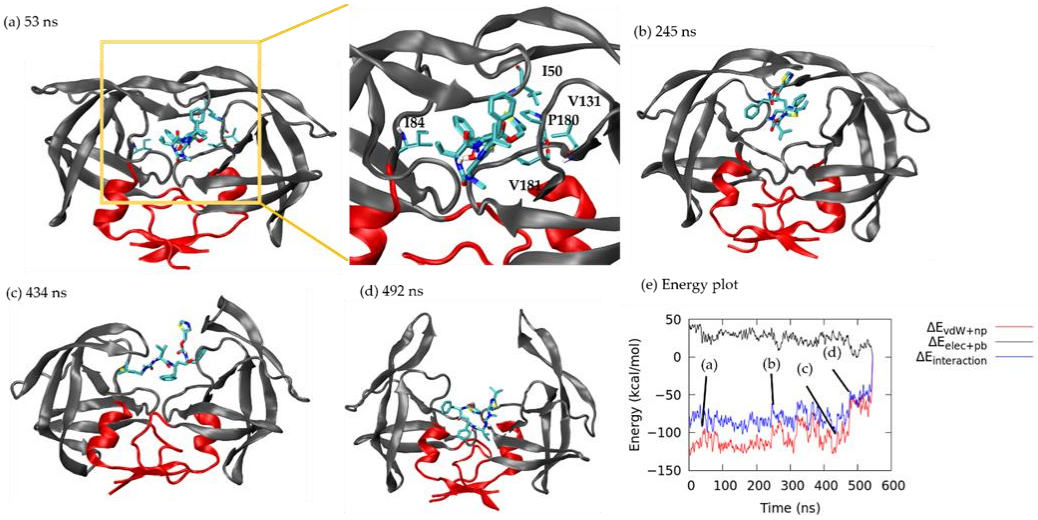


Figure 2.7. Ritonavir dissociation under pathway C. Interface region is in red. Ritonavir is in licorice. (a) 53 ns. (b) 245 ns. (c) 434 ns. (d) 492 ns. (e) Non-polar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between ritonavir and HIVp during unbinding.

Similar to xk263 unbinding from pathway A (see the same conformation as Fig. 5a), the ligand formed a stable H-bond network at the beginning of the aMD run (Fig. S6). Xk263 jiggled inside the binding pocket until it exposed a diol group toward solvent at 222 ns with interaction energy of -79.21 kcal/mol (non-polar: -101.58 kcal/mol, polar: 22.37 kcal/mol) (Fig. 8a), followed by partially sliding out of the binding pocket and partially shifting to the flap/loop region with greatly decreased non-polar interaction energy, -86.14 kcal/mol at 264 ns (Fig. 8b). Flap A opened up at 452 ns, whereas xk263

maintained interactions with flap B and moved back to interact with the interfacial region, which enhanced the non-polar interaction to -110.21 kcal/mol (Fig. 8c). Xk263 moved with flap B to the flap/loop region at 486 ns (Fig. 8d). However, the complex failed to keep strong attractions (total energy: -43.66 kcal/mol; non-polar: -54.01 kcal/mol, polar: 10.35 kcal/mol), and xk263 quickly shifted to the interface region with the diol group contacting HIVp and the ketone group exposed to solvent (Fig. 8e). In this sampled dissociation trajectory, xk263 continued to diffuse on the interface region for longer than 100 ns (equivalent to 4 re-seeding attempts, which are not included here for the trajectory length). If xk263 successfully unbound from HIVp, xk263 spent only ~20 ns contacting the interface region.

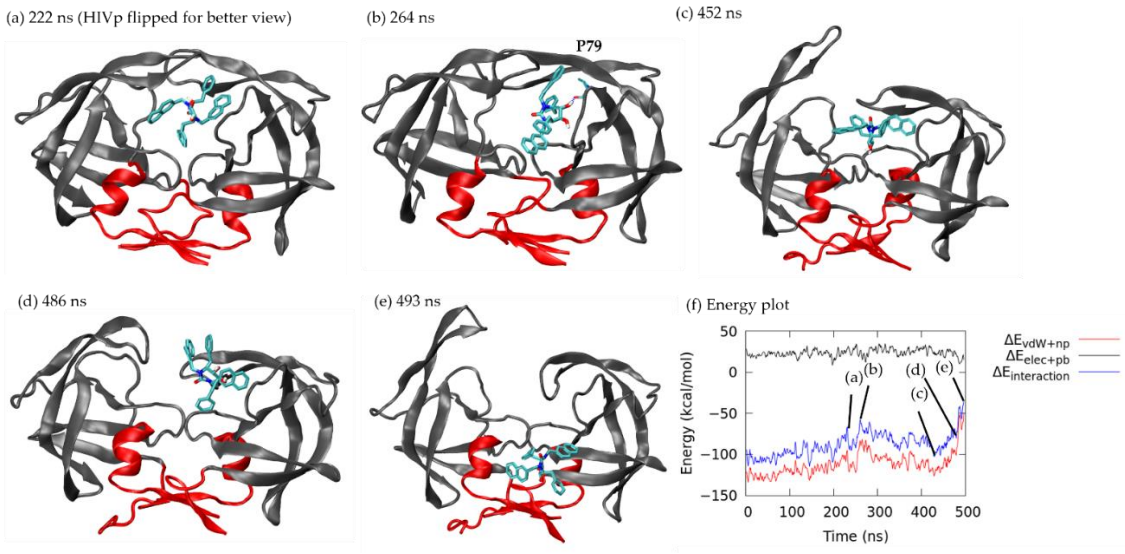


Figure 2.8. Xk263 dissociation under pathway C. Interface region is in red. Xk263 is in licorice. H-bond is shown in dashed blue line with the corresponding residue labelled. (a) 222 ns. (b) 264 ns. (c) 452 ns. (d) 486 ns. (e) 493 ns. (f) Non-polar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between xk263 and HIVp during unbinding.

2.3.1.4. Other pathways

Besides the 3 major dissociation pathways, ritonavir could solvate and unbind when flaps were wide open (observed in 1 trajectory) or attach to the flap tip when the flap opened to dissociate (1 trajectory). Xk263 could unbind through the gap between flaps (1 trajectory), move with the flap tip (1 trajectory) or directly slip out of the binding pocket with the closed-flap HIVp conformation (2 trajectories).

Existing studies showed that xk263 can bind to HIVp with semi-open flaps [11,27]; therefore, here we discuss a case in which xk263 dissociated with slightly-open flaps and closed-flap handedness. Only 2 of 15 xk263 dissociation trajectories showed closed-flap handedness during unbinding. Among them, one trajectory completed unbinding and the other remained on the protein surface when we terminated the run. Before xk263 left the

pocket, HIVp had slightly-open flap handedness during ligand rearrangement inside the binding pocket, but the flaps never fully opened. None of the 20 ritonavir dissociation trajectories had closed flap handedness. Because of the special flap behavior, we classified this pathway under “other pathways” and discussed it in the following text. Flaps remained closed at ~5 Å, but a significant xk263 RMSD increase was recorded starting at 300 ns (Fig. 3,i). HIVp also experienced less fluctuation because the backbone RMSF was < 3.8 Å as compared with 5.4 Å for pathway A (Fig. 3A,iii) and 5.7 Å for pathway C (Fig. 3C,iii). An important precondition for this pathway was that xk263 flipped upside down within the binding pocket when flaps opened at 50 ns and water molecules entered the binding pocket to weaken the interactions between xk263 and HIVp (Fig. 9b). H-bond analysis revealed that only 13 residues formed an H-bond with xk263, but there were 22 residues in pathway A and 19 residues in pathway C, respectively (Fig. S3, S6 and S7).

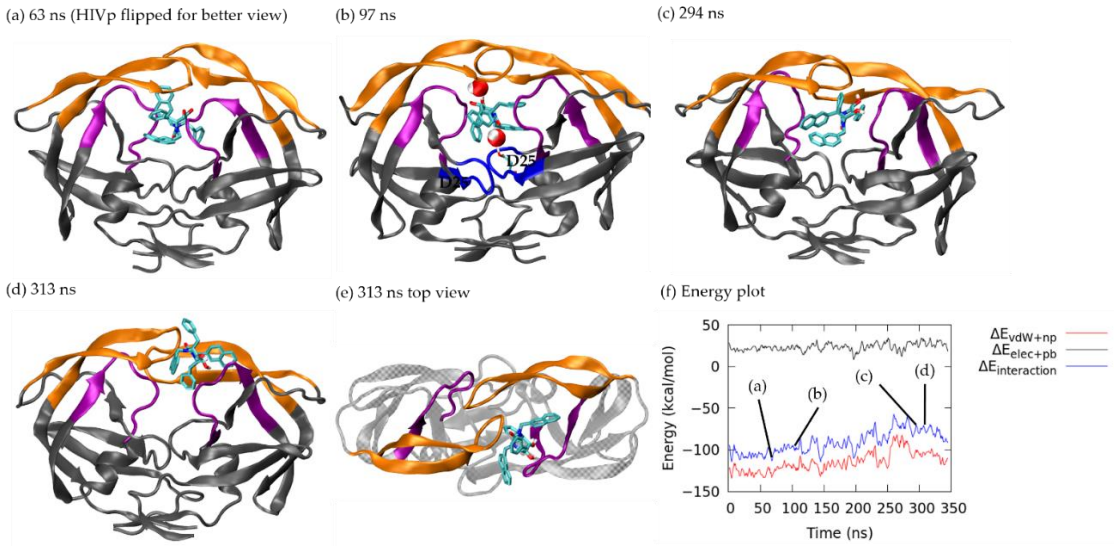


Figure 2.9. Xk263 dissociation with closed-flap HIVp. Flap, loop and catalytic triads are in orange, purple and blue. Xk263 is in licorice. H-bond is shown in dashed blue line with the corresponding residue labelled. (a) 63 ns. (b) 97 ns. (c) 294 ns. (d) (e) 313 ns. (f) Non-polar interaction energy (red), polar interaction energy (black) and total interaction energy (blue) between xk263 and HIVp during unbinding.

Xk263 remained inside the bonding pocket until the flap region slightly opened, which allowed xk263 to flip upside down. During xk263 flipping at 63 ns, the non-polar interaction was slightly weakened, from -130.04 kcal/mol to -120.68 kcal/mol, thus resulting in a small energy barrier of -97.09 kcal/mol. The H-bond between xk263 and HIVp barely changed before and after ligand flipping (Fig. S7) because a carbonyl oxygen in xk263 formed H-bonds with Ile 50/Ile 149 and 2 hydroxyl groups formed H-bonds with Asp 25/Asp 124 before flipping. After xk263 flipping, carbonyl oxygen formed H-bonds with Asp 25/Asp 124 and 2 hydroxyl groups formed H-bonds with Ile 50/Ile 149. The flaps moved from slightly-open to a closed conformation and xk263 remained in the flipped position with bridge waters observed between the flaps and the catalytic triads at 97 ns (Fig. 9b). The 2 bridge water molecules joined the H-bonds

between xk263-HIVp at about 100 ns as the H-bonds (xk263 and Asp 25, Asp 124, Ile 149) broke and reformed (Figure S7). From 50 ns to 280 ns, the RMSD of flipped xk263 remained steady at ~10 Å, which suggests that the bridge waters stabilized xk263 (Fig. 3D,i). At 294 ns, xk263 exposed the diol group to solvent for unbinding (Fig. 9c), then xk263 moved into flap A and the loop region. Because the flaps were closed, xk263 remained in contact with flap B, but the attractions were weak (total energy: -69.77 kcal/mol; non-polar: -99.75 kcal/mol, polar: 29.98 kcal/mol) at 313 ns. Xk263 diffused on the HIVp surface after it passed through the flap/loop region. Because previous studies suggested that the surface diffusion may last longer than 1 us, we terminated the simulation because xk263 completely left the binding pocket.

2.3.2. Association–dissociation trajectories comparison

To examine the similarity of ligand binding/unbinding processes, we first investigated the mutual protein conformations during ligand binding and unbinding. Comparing protein backbone RMSD between different frames from various trajectories is a simple and precise strategy to identify similar protein conformations. We selected the dissociation trajectories of pathway A for ritonavir (Figs. 2a and 4) and xk263 (Fig. 5) and re-saved a frame every 1 ns, yielding a total of 248 and 479 reference frames, respectively. Each frame served as a reference structure to calculate protein backbone RMSD for the association trajectories. In comparison with a reference frame, if the computed RMSD between 2 structures was < 2.5 Å, we considered it an overlapping HIVp conformation with the reference structure.

Both ritonavir and xk263 preferred the flap/loop pathway, dissociation pathway A, when unbinding, as shown above (Table 1). Previous studies also showed that HIVp ligands with different molecular properties prefer different association pathways and/or distinct association rate constants [24,29,48]. However, whether their unbinding processes are simply the reverse process of ligand binding is unclear. Notably, classical binding mechanisms such as conformation selection and induced-fit models mainly describe the conformational fluctuations during ligand binding. For example, ritonavir binds to HIVp with a conformation selection model, whereas xk263 binds with an induced-fit model [27]. Therefore, we wondered whether xk263 could induce the same HIVp motion in both binding and unbinding processes.

In the ritonavir dissociation–association comparison, overlapping conformations were found in 7 of 8 association trajectories (Fig. 10a). Association #2 had no frame similar to any reference frame in the compared dissociation trajectory because HIVp experienced uncommonly large elbow region displacement during ritonavir association (Fig. S8). The population of overlapping frames revealed that HIVp conformations at ~50, ~170 and ~220 ns from the dissociation trajectory were popular among association trajectories (Fig. S9). HIVp flap conformations were semi-open or slightly-open with a flap distance of 6.91 Å, 7.23 Å and 5.96 Å for snapshots at 50 ns, 170 ns and 220 ns, respectively. Ritonavir spent most of the time inside the binding pocket with flaps closed or semi/slightly-open during dissociation. The two-step binding mechanism suggests that the ligand–HIVp spends a long time in conformational change to form a tighter complex [46,49]. Existing studies suggested that the semi-open flap conformation was purposed to

be the most stable intermediate state during flap open/closed position [50,51], which also explains why semi-open/slightly-open flaps were commonly observed in both ligand binding/unbinding in our trajectories. HIVp conformations in ~10 ns, ~100 ns and ~240 ns are not common in association trajectories because the frame count is low or none (Fig. S10). Elbow loop distortion at 10 ns was the main reason why we found no overlapping conformation at this step, even though the flaps were in a slightly-open conformation (i.e., flap tip distance ~6.09 Å). Once the elbow loop recovered from distortion, more overlapping conformations were found starting at ~25 ns. HIVp flap conformations were wide open, with flap distance of 17.56 Å and 27.57 Å for snapshots at 100 ns and 240 ns, respectively. The wide-open conformations were less popular, which agrees with a previous study reporting ~6 kcal/mol higher conformational free energy of substrate-HIVp with open-flap than semi-open flap conformation [46]. Hence, HIVp tends to remain in semi/slightly-open flap conformation.

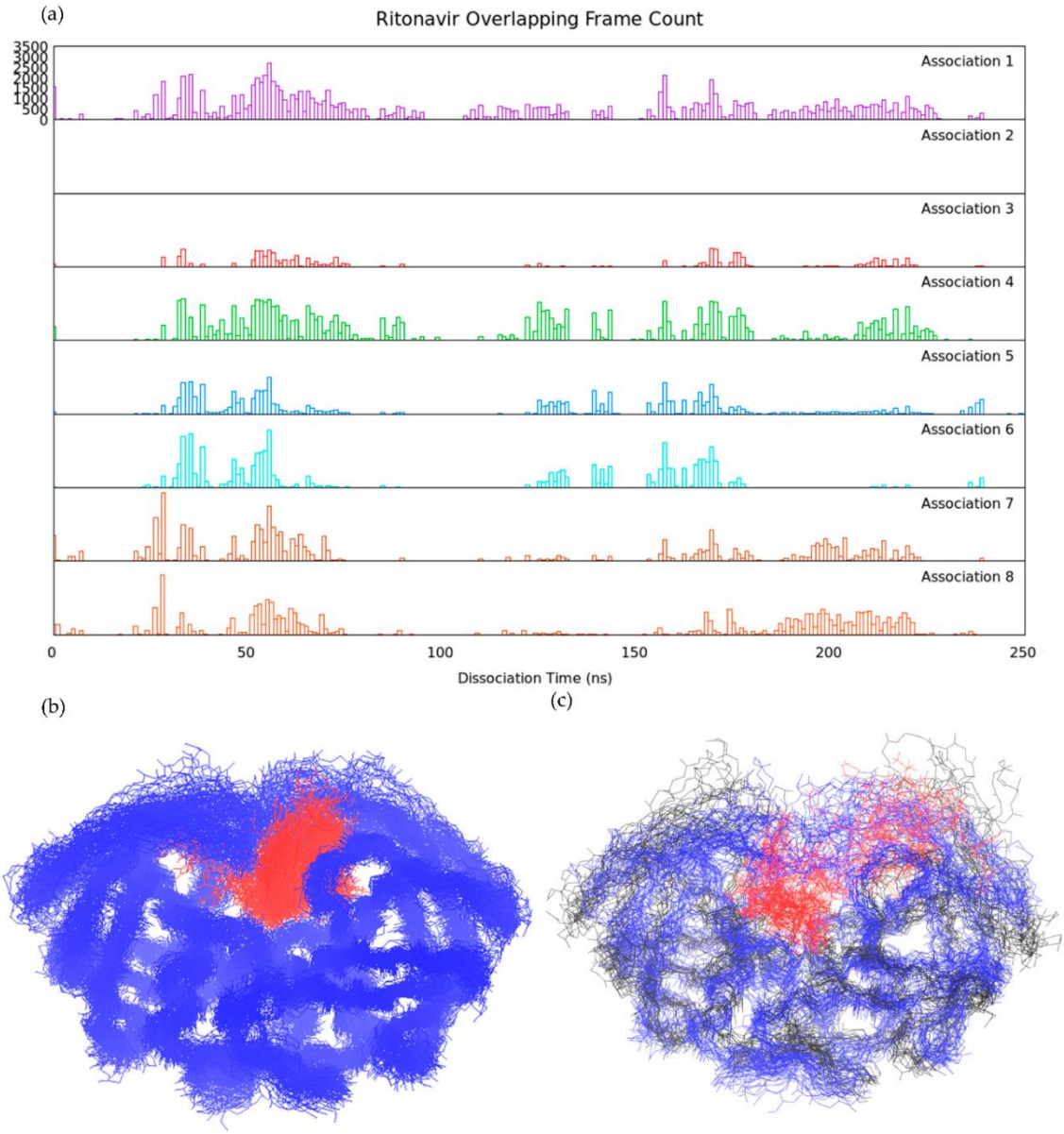


Figure 2.10. Ritonavir association-dissociation comparison under pathway A. (a) Distribution of overlapped frame (< 2.5 Å RMSD) count from 8 ritonavir association trajectories compared to ritonavir dissociation. The reference frame was taken every 1 ns from the dissociation trajectory. (b) Ritonavir-HIVp conformations with overlapped frame in 7 association trajectories. Ritonavir-HIVp was horizontally flipped to better visualize ritonavir's position. (c) Ritonavir-HIVp conformations with no overlapped frame in any association trajectories. HIVp is shown in blue line with backbone. Wide-open HIVp conformations are shown in black line. Ritonavir is shown in red line.

With a total of 248 reference frames in ritonavir dissociation trajectory, 98 frames in dissociation pathway A were found to be overlapping conformations in all 7 association trajectories (Fig. 10b), whereas 31 reference frames had no overlapping frame in any association trajectories (Fig. 10c). In these 98 frames, ritonavir mostly located between the flap/loop region, and HIVp flaps were slightly-open. According to the conformational selection model, these 98 frames are selected popular conformations when ritonavir is entering/leaving the binding pocket. In contrast, when HIVp flaps were wide open, even when ritonavir also located in the flap/loop region, these 9 conformations (black lines shown in Fig. 10c) did not overlap with any frames in association. In the other 22 no-overlapping conformations, ritonavir located inside the binding pocket, attaching to a loop region while HIVp flaps were closed, accompanied by large elbow displacement or open, accompanied by a distorted flap tip. We can conclude that HIVp constantly fluctuates during ritonavir dissociation, yet there are certain HIVp conformations that occur in almost all ritonavir association trajectories. Commonly preferred protein conformations during ritonavir association or dissociation exist, which suggests that the drug utilizes a conformational selection model for both binding and unbinding to HIVp. In contrast to ritonavir, with thousands of overlapping frames during association/dissociation, xk263 association-dissociation shows significantly fewer overlapping conformations (Fig. 11a). The distribution of frame count revealed that HIVp conformations at ~10 ns, ~200 ns and ~440 ns were popular in association trajectories with flap distance 9.03 Å, 11.33 Å and 7.46 Å, respectively (Fig. S11). HIVp in xk263 dissociation favors slightly/semi-open flap conformations than wide-open conformations

because wide-open conformations are at high energy states. Of note, even in the bound state (snapshot at 10 ns), the xk263-HIVp complex had fewer overlapping frames than ritonavir-HIVp, which suggests that xk263 introduced various HIVp conformational changes. Unique HIVp conformations were induced by xk263 unbinding, such as frames between 90 to 190 ns and snapshots at 100 ns, 140 ns and 160 ns shown in Figure S12, where the flaps stacked above xk263, and the elbow loop was largely distorted (Fig. S13). Here flap stacking refers to an asymmetric motion that one flap remains contacting with xk263 while the other flap rotates away from xk263. Flap stacking was only observed in the xk263 dissociation trajectory where the carbonyl group of xk263 tightly binds to one flap tip and the catalytic trials move together with xk263 (Fig. S13). The concerted motions result in the unique conformations in xk263 dissociation, which further demonstrates that xk263-HIVp undergoes an induced-fit mechanism during unbinding. With a total of 479 reference frames, 18 frames in xk263 dissociation were overlapping conformations in all 6 association trajectories (Fig. 11b), whereas 214 frames had no overlapping frame in association trajectories (Fig. 11c). In these 18 frames, xk263 mostly located inside the binding pocket and HIVp flaps were closed. Unlike ritonavir, xk263 quickly moved through the flap/loop region to dissociate; we did not observe popular HIVp conformations with xk263 between the flap/loop region. Figure 5g shows a small energy fluctuation during xk263 unbinding, which suggests a constantly induced small conformational rearrangement during ligand dissociation. The local environment during xk263 binding and unbinding is not the same. As a result, the conformations induced during xk263 association–dissociation are rare.

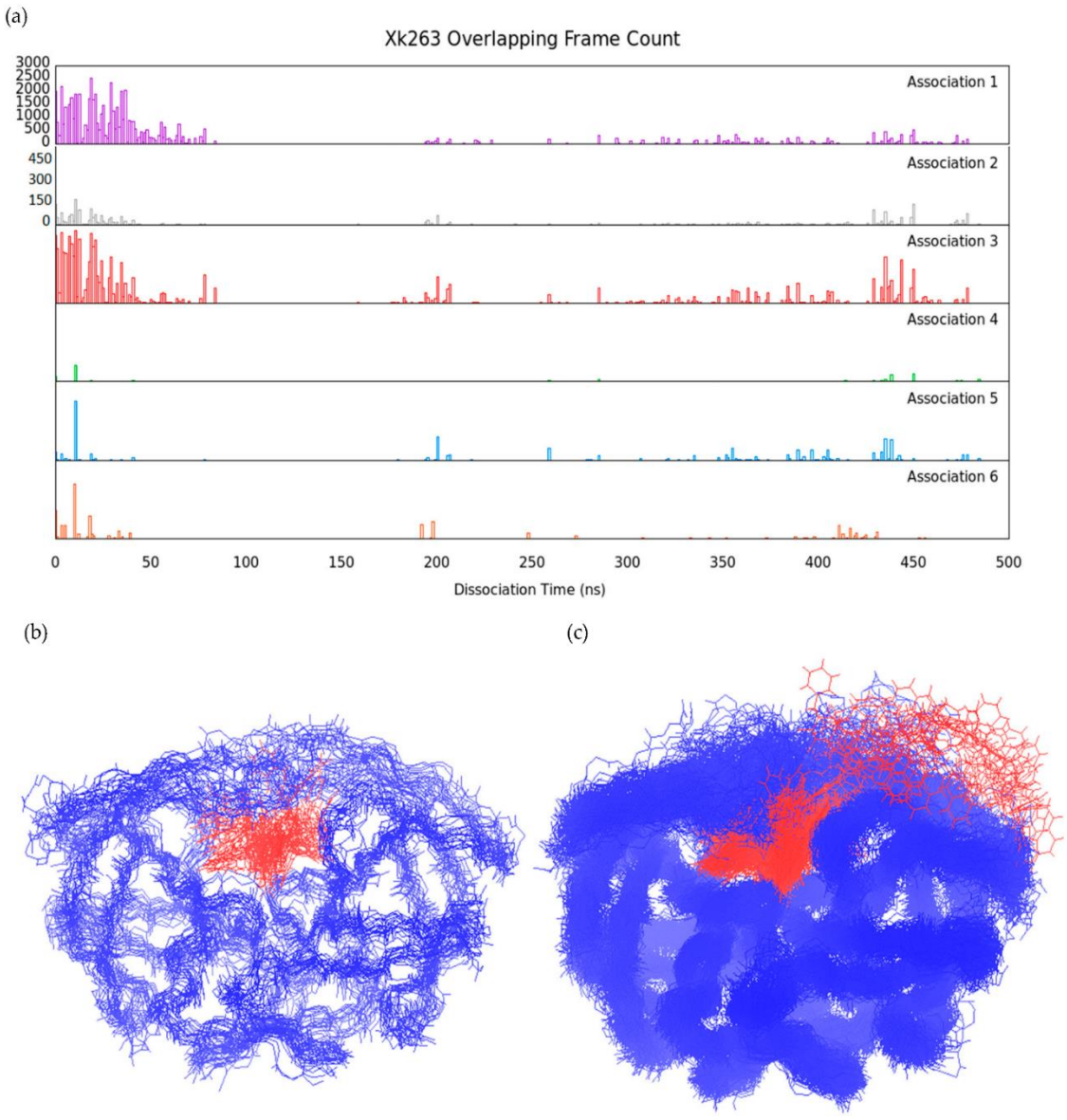


Figure 2.11. Xk263 association–dissociation comparison under pathway A. The y-axis range is 0-3000 for association 1 and 0-450 for the rest of the association trajectories. (a) Distribution of overlapped frame (< 2.5 Å RMSD) count from 6 xk263 association trajectories compared to xk263 dissociation. The reference frame was taken every 1 ns from the dissociation trajectory. (b) Xk263-HIVp conformations with overlapped frame in all association trajectories. (c) Xk263-HIVp conformations with no overlapped frame in any association trajectories. HIVp is shown in blue line with backbone only. Xk263 is shown in red line.

The overlapping HIVp conformation frame count difference between ritonavir and xk263 dissociation supported a conformational-selection mechanism for ritonavir-HIVp and induced-fit mechanism for xk263-HIVp. With ritonavir binding/unbinding from HIVp under pathway A, protein conformations with ritonavir between the flap/loop were observed in both dissociation and association. In contrast, HIVp conformations rarely overlapped during xk263 association/dissociation, which suggests that the conformations were induced during xk263 association/dissociation. The environment during xk263 binding/unbinding is not identical, which results in different induced conformations.

2.3.3. Mutual conformations in association/dissociation

Here, we discuss the pathways in detail to understand the similarities and differences between ritonavir and xk263 during unbinding.

2.3.3.1. Closed flap conformation

HIVp with a closed-flap configuration was the most-observed mutual conformation in the ritonavir dissociation–association comparison, which was expected because all dissociation trajectories started with a closed flap configuration and all association trajectories ended with a closed or semi-open configuration. In contrast, only one reference conformation with a closed flap in xk263 dissociation was found with similar HIVp conformations in xk263 association 1.

Using the initial frame of ritonavir dissociation trajectories as a reference frame (Fig. 4a), we observed the overlapping conformation in 7 association trajectories with abundant frame counts. Figure 4a was the initial frame of aMD simulation, and the protein

conformation is nevertheless highly similar to its crystal structure conformation. Therefore, we selected another closed-flap HIVp conformation in pathway B (Fig. 6a, the frame taken in 32 ns) as a reference structure and found 1283 overlapping frames from our reference dissociation pathway B and 505 overlapping frames from association 1 (Fig. 12a). Hundreds of similar conformations with closed flaps on ritonavir association/dissociation suggested that the pre-existing conformations are selected during ritonavir binding/unbinding processes, which implies use of a conformational-selection binding mechanism.

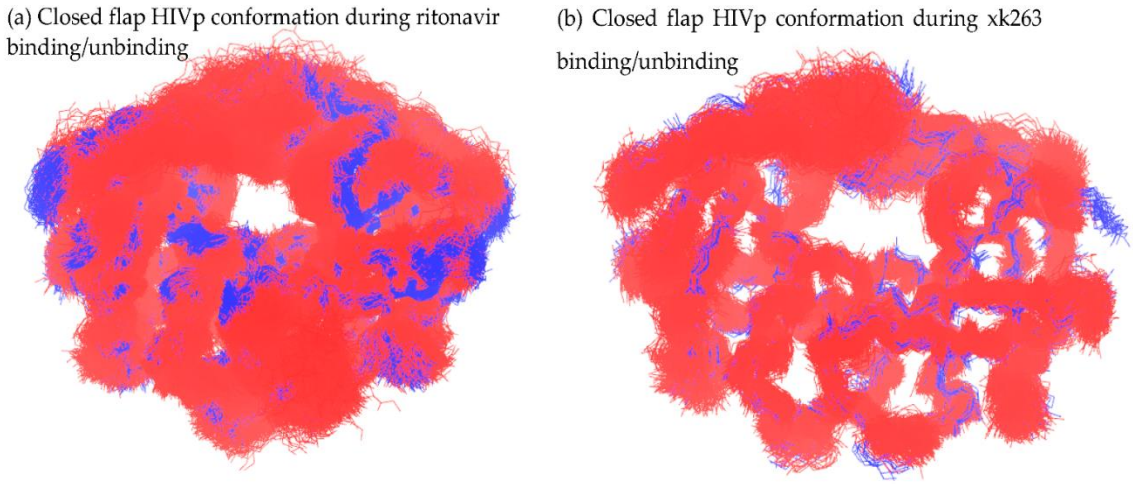


Figure 2.12. Overlapped frames with closed flaps and projections of HIVp configurations on PC space. (a) HIVp conformations in ritonavir dissociation trajectory under pathway B (red) using Figure 6a as a reference frame and corresponding overlapped frames from ritonavir association 1 (blue). (b) HIVp conformations in xk263 dissociation trajectory under pathway C (red) using Figure 8a as a reference frame and corresponding overlapped frames from xk263 association 1 (blue).

Using 8 xk263-HIVp conformations with a closed flap from previous figures (Fig. 5b, 5c, 8a, 8b, 9a, 9b, 9c and 9d) to examine conformation overlap with xk263 association 1, we found that only 27 frames overlapped with conformations (Fig. 8a). No conformations from association 1 were similar to the other 7 closed-flap conformations that appeared during xk263 dissociation. Using the HIVp conformation in pathway C (Fig. 8a) as a reference structure, we found 2710 overlapping frames from a dissociation trajectory with 4932 frames but only 27 overlapping frames from the association 1 trajectory with 3000 frames (Fig. 12b). Although Figure 8a is a highly populated conformation in pathway C, the low similarity of closed-flap confirmations between xk263 dissociation and association raised our interest. The low similarity is due to the induced-fit model with different micro-environments during xk263 association and dissociation precluding the system from inducing similar HIVp conformations.

In the ritonavir-HIVp bound complex, a bridge water molecule connects ritonavir and the flap region with H-bonds [33,35]. The cyclic-urea inhibitors were designed by using the oxygen atom in the carbonyl group of cyclic urea to replace the bridge water [52]. The diol group in xk263 forms H-bonds with catalytic triads, especially with catalytic residues Asp 25 and Asp 124. Hence, the H-bond network consisting of the flap region-xk263-catalytic triads tightly clamped xk263 in a non-polar environment. The H-bond network and non-polar binding pocket also restricted flap motions, as seen in Figure 3, in which flap distance is generally less in xk263 unbinding than ritonavir unbinding. As compared with xk263 association, in which the ligand was binding to a pocket full of water molecules, the ligand was unbinding from a hydrophobic environment, which

induced significantly different HIVp conformations during association–dissociation. Our analysis suggests that an induced-fit mechanism is used for xk263-HIVp dissociation as well.

2.3.3.2. Open flap configuration

HIVp with an open-flap configuration was observed in both ritonavir and xk263 dissociation trajectories. As a ligand moves to one flap/loop region, the ligand formed an asymmetric contact with the flaps, thus resulting in the flaps opening.

Here we used a slightly-open (Fig. 4b, 34ns) and wide-open conformation (Fig 4c, 172 ns) from the most popular ritonavir dissociation pathway A to examine association–dissociation overlapping. We used the flap tip distance (α-C distance between Ile 50-Ile 149) to define slightly (7.70 Å in Fig. 4b) or wide open (14.92 Å in Fig. 4c) conformations. Different flap openness yielded significantly different results. In the association 1 trajectory, 1803 and 3 overlapping frames had a slightly-open (Fig. 13a) and wide-open (Fig. 13b) conformations, respectively. Notably, the trajectory recorded a total of 3000 frames; therefore, this overlapping conformation populated 60.2% of the entire association process. Comparing the entire dissociation pathway A with frame 4b of the same trajectory, this conformation is also highly popular: 12.94% of the entire ritonavir dissociation process. Notably, this slightly-open conformation is also one of the most popular conformations when HIVp is in the ligand-free state, which suggests that ritonavir also selects this conformation during dissociation [27].

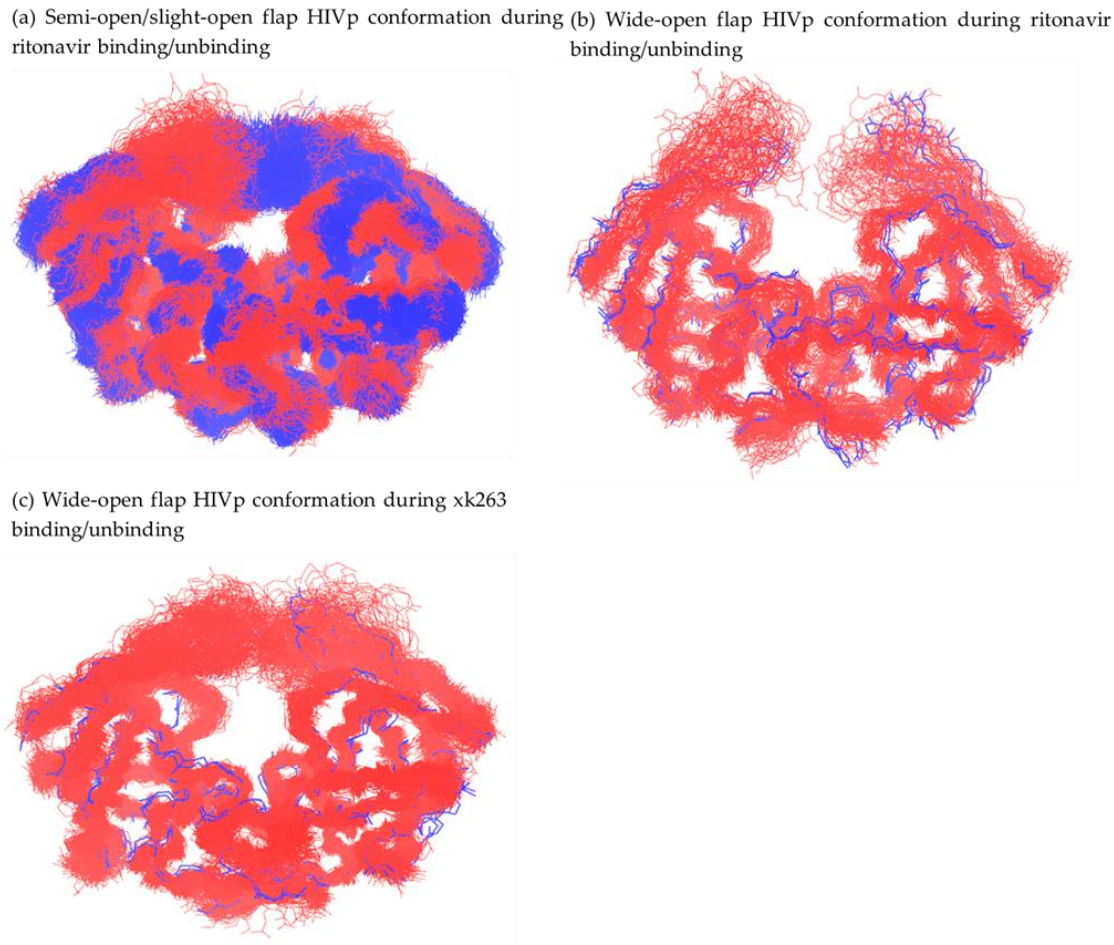


Figure 2.13. Overlapped frames with open flaps and projections of HIVp configurations on PC space. (a) HIVp conformations in ritonavir dissociation trajectory under pathway A (red) using Figure 4b as a reference frame and their corresponding overlapped frames from ritonavir association 1 (blue). (b) HIVp conformations in ritonavir dissociation trajectory under pathway A (red) using Figure 4c as a reference frame and their corresponding overlapped frames from ritonavir association 1 (blue). (c) HIVp conformations in xk263 dissociation trajectory under pathway A (red) using Figure 5f as a reference frame and their corresponding overlapped frames from xk263 association 1 (blue).

Wide-open conformations also pre-exist in the ligand-free state, but they are significantly less populated because of unfavorable conformational free energy. Previous studies showed that conformational free energy for HIVp was ~6 kcal/mol higher for wide-open

than semi-open confirmations [46]. As compared with the wide-open conformation (Fig. 4c) with the entire dissociation pathway A, only 54 frames (2.17 %) overlapped with this wide-open conformation. Because of the free energy cost and highly flexible open flaps, only a few (i.e., 3 in this case) overlapping frames from association 1 to Figure 4c was anticipated.

As for xk263, we also selected a slightly-open (Fig. 5d, 440ns) and wide-open (Fig. 5f, 467 ns) flap to examine conformation overlapping during xk263 association-dissociation. Regardless of flap openness, in the association 1 trajectory, we found 0 and 2 overlapping frames with a slightly-open and wide-open conformation, respectively (Fig. 13c). Although a low overlap possibility with a wide-open conformation during association-dissociation is anticipated, we did not expect that no frames from Association 1 would show the similar slightly-open HIVp conformation observed in dissociation Pathway A. Therefore, we checked the population of the 2 open conformations in Pathway A: slightly-open conformations (Fig. 5d) occupied 41.4% (1989 of 4800 frames) of the entire trajectory and wide-open conformations 12.5% (600 of 4800 frames). Even though the slightly-open conformation was highly popular in dissociation pathway A, the induced flap conformations were unique in xk263 dissociation. As a result, the wide-open conformation with symmetric flap opening is the only common open HIVp conformation during xk263 association–dissociation. Xk263 is highly hydrophobic. Because the pocket of HIVp contains mainly non-polar residues, xk263 prefers contacting with HIVp to induce new protein motions. For example, when xk263 moved to flap B/loop B for dissociation, xk263 could also remain in contact with flap A, as seen in Figure 5d and 5e

for pathway A and Figure 8d for pathway C. As xk263 continued dissociating, xk263–flap A interactions promoted asymmetric flap A movement, thus resulting in a newly induced HIVp flap conformation. Hence, new HIVp conformations were constantly induced as xk263 was unbinding from the pocket. Unlike ritonavir, which can partially re-solvate after flaps open widely (Fig. 4c and Fig. 7d), xk263 always favors contacting with flap/loop regions instead of easily re-solvating into the solvent. We also believe that the hydrophobicity of xk263 prevents the ligand from undergoing pathway B because diffusion on the surface through the flaps exposes most of the ligand to solvent. As suggested in previous papers, ligand properties contribute to the binding mechanisms, and here we showed that the properties affect unbinding mechanisms as well [10,17,27].

2.4. Conclusions

The study used unbiased aMD and a re-seeding approach to sample unbinding pathways of ritonavir-HIVp and xk263-HIVp, which brings a more comprehensive picture for understanding molecular recognition and unbinding mechanisms of ligand and protein. We observed 3 common dissociation pathways: between the flap/loop region, diffusion on the flap region and diffusion on the interface region. Dissociation between flap/loop regions (pathway A) was the major dissociation pathway for both ritonavir and xk263 because of strong non-polar interactions between the ligand and HIVp to bring increased contacts between the ligand and loop region during dissociation. However, the loop region could form transient but strong H-bonds with ritonavir to prevent ritonavir from dissociation. The H-bond between the loop region and xk263 has a short lifetime but helps to stabilize xk263. Diffusion on the flap, pathway B, with the ligand moving to the

solvent following the flap opening, was observed in only ritonavir dissociation. The non-polar property of xk263 keeps xk263 contact with HIVp even after the flaps already opened. Diffusion on the interface region was observed in both ritonavir and xk263 dissociation.

We observed overlapping conformations between dissociation and association trajectories for both ritonavir and xk263 unbinding/binding. However, overlapping HIVp conformations are rarely seen between xk263-HIVp association/dissociation. Because the environments during binding/unbinding differ, significantly different conformations are induced during unbinding. Previous studies suggested a conformational-selection binding mechanism for ritonavir. During unbinding processes, ritonavir revisited these conformations, which suggests the use of a conformational-selection mechanism in unbinding as well. Among overlapping conformations, we observed more closed-flap conformations than slightly-open or wide-open conformations for both ritonavir and xk263. In general, ritonavir requires open flaps to achieve unbinding, whereas xk263 can unbind with a closed-flap conformation. Our study suggests that traditional structure-based drug design that focused on the bound state may be extended to stabilizing the transient conformations during ligand binding/unbinding to prolong drug dissociation and increase the dissociation rate constant.

2.5 REFERENCE

1. Kohl, N.E.; Emini, E.A.; Schleif, W.A.; Davis, L.J.; Heimbach, J.C.; Dixon, R.A.F.; Scolnick, E.M.; Sigal, I.S. ACTIVE HUMAN IMMUNODEFICIENCY VIRUS PROTEASE IS REQUIRED FOR VIRAL INFECTIVITY. Proceedings of the National

Academy of Sciences of the United States of America 1988, 85, 4686-4690, doi:10.1073/pnas.85.13.4686.

2. Tomasselli, A.G.; Heinrikson, R.L. Targeting the HIV-protease in AIDS therapy: a current clinical perspective. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 2000, 1477, 189-214, doi:10.1016/s0167-4838(99)00273-3.

3. Weber, I.T.; Wang, Y.F.; Harrison, R.W. HIV Protease: Historical Perspective and Current Research. Viruses-Basel 2021, 13, doi:10.3390/v13050839.

4. Zuo, X.; Huo, Z.; Kang, D.; Wu, G.; Zhou, Z.; Liu, X.; Zhan, P. Current insights into anti-HIV drug discovery and development: a review of recent patent literature (2014–2017). Expert Opinion on Therapeutic Patents 2018, 28, 299-316, doi:10.1080/13543776.2018.1438410.

5. Voshavar, C. Protease Inhibitors for the Treatment of HIV/AIDS: Recent Advances and Future Challenges. Current Topics in Medicinal Chemistry 2019, 19, 1571-1598, doi:10.2174/1568026619666190619115243.

6. Pawar, S.D.; Freas, C.; Weber, I.T.; Harrison, R.W. Analysis of drug resistance in HIV protease. Bmc Bioinformatics 2018, 19, doi:10.1186/s12859-018-2331-y.

7. Whitfield, T.W.; Ragland, D.A.; Zeldovich, K.B.; Schiffer, C.A. Characterizing Protein-Ligand Binding Using Atomistic Simulation and Machine Learning: Application

to Drug Resistance in HIV-1 Protease. Journal of Chemical Theory and Computation 2020, 16, 1284-1299, doi:10.1021/acs.jctc.9b00781.

8. Nascimento, A.; Fernandes, R.P.; Quijia, C.; Araujo, V.H.S.; Pereira, J.; Garcia, J.S.; Trevisan, M.G.; Chorilli, M. Pharmacokinetic Parameters of HIV-1 Protease Inhibitors. Chemmedchem 2020, 15, 1018-1029, doi:10.1002/cmdc.202000101.

9. Trylska, J.; Tozzini, V.; Chang, C.-e.A.; McCammon, J.A. HIV-1 Protease Substrate Binding and Product Release Pathways Explored with Coarse-Grained Molecular Dynamics. Biophysical Journal 2007, 92, 4179-4187, doi:https://doi.org/10.1529/biophysj.106.100560.

10. Lexa, K.W.; Carlson, H.A. Binding to the open conformation of HIV-1 protease. Proteins-Structure Function and Bioinformatics 2011, 79, 2282-2290, doi:10.1002/prot.23054.

11. Brik, A.; Wong, C.H. HIV-1 protease: mechanism and drug discovery. Org Biomol Chem 2003, 1, 5-14, doi:10.1039/b208248a.

12. Copeland, R.A. The drug-target residence time model: a 10-year retrospective. Nature Reviews Drug Discovery 2016, 15, 87-95, doi:10.1038/nrd.2015.18.

13. Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. Annual Review of Physical Chemistry, Vol 70 2019, 70, 143-171, doi:10.1146/annurev-physchem-042018-052340.

14. Lu, H.; Iuliano, J.N.; Tonge, P.J. Structure-kinetic relationships that control the residence time of drug-target complexes: insights from molecular structure and dynamics. Current Opinion in Chemical Biology 2018, 44, 101-109, doi:10.1016/j.cbpa.2018.06.002.

15. Ribeiro, J.M.L.; Tsai, S.T.; Pramanik, D.; Wang, Y.H.; Tiwary, P. Kinetics of Ligand-Protein Dissociation from All-Atom Simulations: Are We There Yet? Biochemistry 2019, 58, 156-165, doi:10.1021/acs.biochem.8b00977.

16. Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. International Journal of Molecular Sciences 2020, 21, doi:10.3390/ijms21176339.

17. Huang, Y.M.M.; Kang, M.; Chang, C.E.A. Switches of hydrogen bonds during ligand-protein association processes determine binding kinetics. Journal of Molecular Recognition 2014, 27, 537-548, doi:10.1002/jmr.2377.

18. Yu, Y.X.; Liu, W.T.; Li, H.Y.; Wang, W.; Sun, H.B.; Zhang, L.L.; Wu, S.L. Decoding molecular mechanism underlying binding of drugs to HIV-1 protease with molecular dynamics simulations and MM-GBSA calculations. Sar and Qsar in Environmental Research, doi:10.1080/1062936x.2021.1979647.

19. Peng, C.; Wang, J.A.; Xu, Z.J.; Cai, T.T.; Zhu, W.L. Accurate prediction of relative binding affinities of a series of HIV-1 protease inhibitors using semi-empirical

quantum mechanical charge. Journal of Computational Chemistry 2020, 41, 1773-1780, doi:10.1002/jcc.26218.

20. Gupta, S.; Senapati, S. Mechanism of inhibition of drug-resistant HIV-1 protease clinical isolates by TMC310911: A molecular dynamics study. Journal of Molecular Structure 2019, 1198, doi:10.1016/j.molstruc.2019.126893.

21. Li, D.C.; Ji, B.H.; Hwang, K.C.; Huang, Y.G. Strength of Hydrogen Bond Network Takes Crucial Roles in the Dissociation Process of Inhibitors from the HIV-1 Protease Binding Pocket. Plos One 2011, 6, doi:10.1371/journal.pone.0019268.

22. Huang, S.H.; Zhang, D.; Mei, H.; Kevin, M.; Qu, S.J.; Pan, X.C.; Lu, L.C. SMD-Based Interaction-Energy Fingerprints Can Predict Accurately the Dissociation Rate Constants of HIV-1 Protease Inhibitors. Journal of Chemical Information and Modeling 2019, 59, 159-169, doi:10.1021/acs.jcim.8b00567.

23. Ngo, S.T.; Nguyen, M.T. Determination of the absolute binding free energies of HIV-1 protease inhibitors using non-equilibrium molecular dynamics simulations. Chemical Physics Letters 2017, 676, 12-17, doi:10.1016/j.cplett.2017.03.034.

24. Bhattarai, A.; Miao, Y.L. Gaussian accelerated molecular dynamics for elucidation of drug pathways. Expert Opinion on Drug Discovery 2018, 13, 1055-1065, doi:10.1080/17460441.2018.1538207.

25. ES, F.; E, D.S.; KJ, I.; JJ, L.; T, S.; DJ, P. Two-step binding mechanism for HIV protease inhibitors. 1992, 31, 7886-7891, doi:10.1021/bi00149a020.

26. Katoh, E.; Louis, J.; Yamazaki, T.; Gronenborn, A.; Torchia, D.; Ishima, R. A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex. Biological Magnetic Resonance Data Bank 2012.

27. Huang, Y.M.M.; Raymundo, M.A.V.; Chen, W.; Chang, C.E.A. Mechanism of the Association Pathways for a Pair of Fast and Slow Binding Ligands of HIV-1 Protease. Biochemistry 2017, 56, 1311-1323, doi:10.1021/acs.biochem.6b01112.

28. Qu, S.J.; Huang, S.H.; Pang, X.C.; Yang, L.; Mei, H. Constructing Interconsistent, Reasonable, and Predictive Models for Both the Kinetic and Thermodynamic Properties of HIV-1 Protease Inhibitors. Journal of Chemical Information and Modeling 2016, 56, 2061-2068, doi:10.1021/acs.jcim.6b00326.

29. Shuman, C.F.; Hamalainen, M.D.; Danielson, U.H. Kinetic and thermodynamic characterization of HIV-1 protease inhibitors. Journal of Molecular Recognition 2004, 17, 106-119, doi:10.1002/jmr.655.

30. Miao, Y.L.; Huang, Y.M.; Walker, R.; McCammon, J.A.; Chang, C.E. Ligand binding pathways and conformational transitions of the HIV protease. Abstracts of Papers of the American Chemical Society 2018, 256.

31. Acevedo, O.; Ambrose, Z.; Flaherty, P.T.; Aamer, H.; Jain, P.; Sambasivarao, S.V. Identification of HIV Inhibitors Guided by Free Energy Perturbation Calculations. Current Pharmaceutical Design 2012, 18, 1199-1216.

32. Karnati, K.R.; Wang, Y.X. Structural and binding insights into HIV-1 protease and P2-ligand interactions through molecular dynamics simulations, binding free energy and principal component analysis. Journal of Molecular Graphics & Modelling 2019, 92, 112-122, doi:10.1016/j.jmgm.2019.07.008.

33. Kempf, D.J.; Marsh, K.C.; Denissen, J.F.; McDonald, E.; Vasavanonda, S.; Flentge, C.A.; Green, B.E.; Fino, L.; Park, C.H.; Kong, X.P.; et al. ABT-538 IS A POTENT INHIBITOR OF HUMAN-IMMUNODEFICIENCY-VIRUS PROTEASE AND HAS HIGH ORAL BIOAVAILABILITY IN HUMANS. Proceedings of the National Academy of Sciences of the United States of America 1995, 92, 2484-2488, doi:10.1073/pnas.92.7.2484.

34. Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; et al. RATIONAL DESIGN OF POTENT, BIOAVAILABLE, NONPEPTIDE CYCLIC UREAS AS HIV PROTEASE INHIBITORS. Science 1994, 263, 380-384, doi:10.1126/science.8278812.

35. Chen, J.Z.; Yang, M.Y.; Hu, G.D.; Shi, S.H.; Yi, C.H.; Zhang, Q.G. Insights into the functional role of protonation states in the HIV-1 protease-BEA369 complex:

molecular dynamics simulations and free energy calculations. Journal of Molecular Modeling 2009, 15, 1245-1252, doi:10.1007/s00894-009-0452-y.

36. DA, C.; V, B.; JT, B. AMBER 2018. 2018.

37. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. Journal of Computational Chemistry 2004, 25, 1157-1174, doi:doi:10.1002/jcc.20035.

38. Gilson, M.K.; Gilson, H.S.R.; Potter, M.J. Fast assignment of accurate partial atomic charges: An electronegativity equalization method that accounts for alternate resonance forms. Journal of Chemical Information and Computer Sciences 2003, 43, 1982-1997, doi:10.1021/ci034148o.

39. Kholmurodov, K.; Smith, W.; Yasuoka, K.; Darden, T.; Ebisuzaki, T. A smooth-particle mesh Ewald method for DL_POLY molecular dynamics simulation package on the Fujitsu VPP700. Journal of Computational Chemistry 2000, 21, 1187-1191, doi:10.1002/1096-987x(200010)21:13<1187::aid-jcc7>3.0.co;2-7.

40. JEAN-PAUL, R.; GIOVANNI, C.; C., B.H.J. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. 1977, 23, 327-341.

41. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. Journal of Chemical Physics 2004, 120, 11919-11929, doi:10.1063/1.1755656.

42. Chen, D.L.; Oezguen, N.; Urvil, P.; Ferguson, C.; Dann, S.M.; Savidge, T.C. Regulation of protein-ligand binding affinity by hydrogen bond pairing. Science Advances 2016, 2, doi:10.1126/sciadv.1501240.

43. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 2013, 9, 3084-3095, doi:10.1021/ct400341p.

44. Miller, B.R.; McGee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. Journal of Chemical Theory and Computation 2012, 8, 3314-3321, doi:10.1021/ct300418h.

45. Huang, X.; Britto, M.D.; Kear-Scott, J.L.; Boone, C.D.; Rocca, J.R.; Simmerling, C.; McKenna, R.; Bieri, M.; Gooley, P.R.; Dunn, B.M.; et al. The Role of Select Subtype Polymorphisms on HIV-1 Protease Conformational Sampling and Dynamics. Journal of Biological Chemistry 2014, 289, 17203-17214, doi:10.1074/jbc.M114.571836.

46. Karthik, S.; Senapati, S. Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle. Proteins 2011, 79, 1830-1840, doi:10.1002/prot.23008.

47. Tang, Z.Y.; Roberts, C.C.; Chang, C.E.A. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. Frontiers in Bioscience-Landmark 2017, 22, 960-981, doi:10.2741/4527.

48. Li, D.C.; Ji, B.H.; Liu, M.S.; Hwang, K.C.; Huang, Y.G. Coarse-grained Molecular Dynamics of Inhibitors Binding into HIV-1 Protease. In Proceedings of the 11th International Congress of the IUPESM/World Congress on Medical Physics and Biomedical Engineering, Munich, GERMANY, Sep 07-12, 2009; pp. 623-625.

49. Furfine, E.S.; Dsouza, E.; Ingold, K.J.; Leban, J.J.; Spector, T.; Porter, D.J.T. 2-STEP BINDING MECHANISM FOR HIV PROTEASE INHIBITORS. Biochemistry 1992, 31, 7886-7891, doi:10.1021/bi00149a020.

50. Toth, G.; Borics, A. Flap opening mechanism of HIV-1 protease. Journal of Molecular Graphics & Modelling 2006, 24, 465-474, doi:10.1016/j.jmgm.2005.08.008.

51. Sadiq, S.K.; De Fabritiis, G. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. Proteins-Structure Function and Bioinformatics 2010, 78, 2873-2885, doi:10.1002/prot.22806.

52. Lam, P.Y.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; et al. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 1994, 263, 380, doi:10.1126/science.8278812.

CHAPTER 3 FREE ENERGY LANDSCAPE OF RITONAVIR-HIV PROTEASE DISSOCIATION UNDER DIFFERENT PATHWAYS

3.1 ABSTRACT

Investigation of drug binding kinetics can help us better understand molecular recognition between drugs and proteins. Here, we construct free energy landscape of ritonavir unbinding from its host, HIV protease, under three dissociation pathways. We utilized Binding Kinetics Toolkit (BKiT) to project dissociation trajectories onto principal component (PC) space and use PC as guidance to assign unbinding indexes which represent distinct ligand-protein interactions during dissociation. Transitions between unbinding indexes are sampled by running numerous short molecular dynamics. Using milestoning theory, we computed binding free energy landscape and drug residence time. We also connected ritonavir-HIV protease atomistic interactions with free energy barriers or wells. The results of binding free energy and drug residence time were improved by using ligand root-mean-squared-deviation (RMSD) as reaction coordinates, giving us a much more accurate approximation to the experimental values. Our study provided close estimation of ligand binding free energy and demonstrated the importance of sampling ligand-protein dissociation pathways in understanding ligand binding KINETICS.

3.2 INTRODUCTION

Human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS) by attacking CD4 cells and weakening human immune system. [1-4] HIV protease (HIVp) cleavages premature peptide and produces functional enzyme, which

makes it indispensable for virus replication. [5-6] Among 23 FDA-approved anti-HIV drugs, 5 of them are HIVp inhibitors. Investigation of drug-HIVp interactions not only improves drug binding affinity, but also benefits drug discovery for other proteases. [7-10]

Transient states during ligand−protein binding/unbinding provide insights into ligand binding kinetics, binding mechanisms, and ligand-protein molecular recognition. [11-15] Understanding ligand-protein interactions at transient states can help us optimize drugs with desired kinetic properties. [16-20] Enhanced sampling have been applied to achieve ligand dissociation because of the high energy barrier during ligand binding/unbinding which are not sampled efficiently in conventional molecular dynamics (cMD). [21-28] Reconstructing the real free energy profile was crucial among enhanced sampling methods. Free energy landscape, not merely the free energy difference between ligand-protein crystal-bound and complete unbound states, is worth studying because free energy landscape connects atomic details of drug action mechanism with transient states. [29-33]. Even though the binding pocket of HIVp is well defined, the possible ligand unbinding pathways were not investigated comprehensively until recent work.

Drug residence time can better predict drug efficacy when compared to drug binding affinity. Association rate constant, dissociation rate constant, and dissociation constant are the major parameters to describe ligand binding kinetics. Experimental kinetic data is the ensemble result of many individual ligand binding/unbinding events. Studying the free energy landscape of one ligand dissociating under multiple pathways elucidates molecular recognition.

Here, we investigated the free energy landscape of ritonavir unbinding from HIVp under different pathways using trajectories sampled by aMD. The first two principal components of each trajectory were used as the reaction coordinates to guide us to assign the unbinding indexes. We ran numerous short cMD trajectory to obtain transition information. We computed the free energy landscape using BKiT package, which is a python toolkit embedded in Jupyter Notebook. [34-35] Moreover, we also computed the landscape using RMSD as reaction coordinate in BKiT package, giving an improved estimation of drug binding free energy and residence time.

3.3 METHODS

3.3.1 Ritonavir dissociation trajectory and short MDs

Ritonavir-HIVp dissociation trajectories sampled by aMD were taken from published work[cite] which used Amber14SB and GAFF for protein and ligand, respectively. [28] Initial frames for short MDs were taken every 100 ps for the first 200ns, 275 ns, and 500 ns for ritonavir dissociation under pathway A, B, and C, respectively. For the rest of the trajectories (200.02 ns to 248.50 ns, 275.02 ns to 304.90 ns, 500.02 ns to 543.10 ns for pathway A, B, and C, respectively), we took initial frames every 20 ps to better sample ritonavir-HIVp motions. We ran 10 replicas of short MDs for each initial frame. All short MDs were 100 ps long unbiased conventional MD with TIP3P water model, Amber14SB for protein and GAFF for ritonavir at 298K. Frames in short MDs were saved every 100 fs.

3.3.2 Procedures of running BKiT

A brief description of procedures in obtaining ligand-protein dissociation free energy landscape using BKiT is listed below:

i. Obtain ligand-protein dissociation trajectory. Due to the high energy barriers in macromolecular systems, enhanced sampling is encouraged in obtaining ligand-protein dissociation trajectories.

ii. Compute principal components of selected atoms in ligand-protein system to reduce the degree of freedom. Then project whole trajectory onto the PC space and assign unbinding indexes to represent particular ligand-protein interactions during dissociation. Assigned unbinding indexes should be parallel to each other to better distinguish between system conformations.

iii. Run short MDs. System transitioning between unbinding indexes are computed by analyzing system behavior in numerous short MDs. We selected frames from dissociation trajectory as initial frames for short MD runs. As ligand leaves binding pocket, we decreased the interval of initial frame selection to better cover the PC space.

iv. Obtain transition matrix. We provide two methods of constructing transition matrix, PC based or ligand RMSD based. For the PC based method, all frames from short MD are projected onto PC space. The system movement in each short MD is examined and judged if it crossed any two consecutive unbinding indexes. If a transition occurred, two neighbor unbinding indexes and frames are recorded to compute the time needed in such transition. The ligand RMSD

based method is similar to PC based method but using ligand RMSD as unbinding indexes.

v. Free energy landscape calculation and residence time approximation. Free energy of each unbinding index can be calculated by:

$$F = -k_B T ln(q_i t_i) \quad \text{eq. 3.1}$$

where q_i is the stationary flux of unbinding index i and t_i is the averaged life time of unbinding index i. Residence time of system transit from unbinding index 1 to unbinding index n can be estimated by:

$$MFPT = \sum_1^n q_i t_i / q_n \quad \text{eq.3.2}$$

3.4 RESULTS AND DISCUSSION

Unbinding free energies of ritonavir-HIV protease under pathway A, B, and C are -9.4 kcal/mol, -9.5 kcal/mol, and -6.1 kcal/mol, respectively, when using PC as reaction coordinates. Residence time of ritonavir-HIV protease under pathway A, B, and C are 483 us, 2 ms, and 4.5 us, respectively, when using PC as reaction coordinates. Using ligand RMSD as reaction coordinates, the computed unbinding free energies are -14.1 kcal/mol, -15.4 kcal/mol, and -14.9 kcal/mol for pathway A, B, and C, respectively. The residence time are 2.5 s, 41.6 s, and 29.6 s for pathway A, B, and C, respectively.

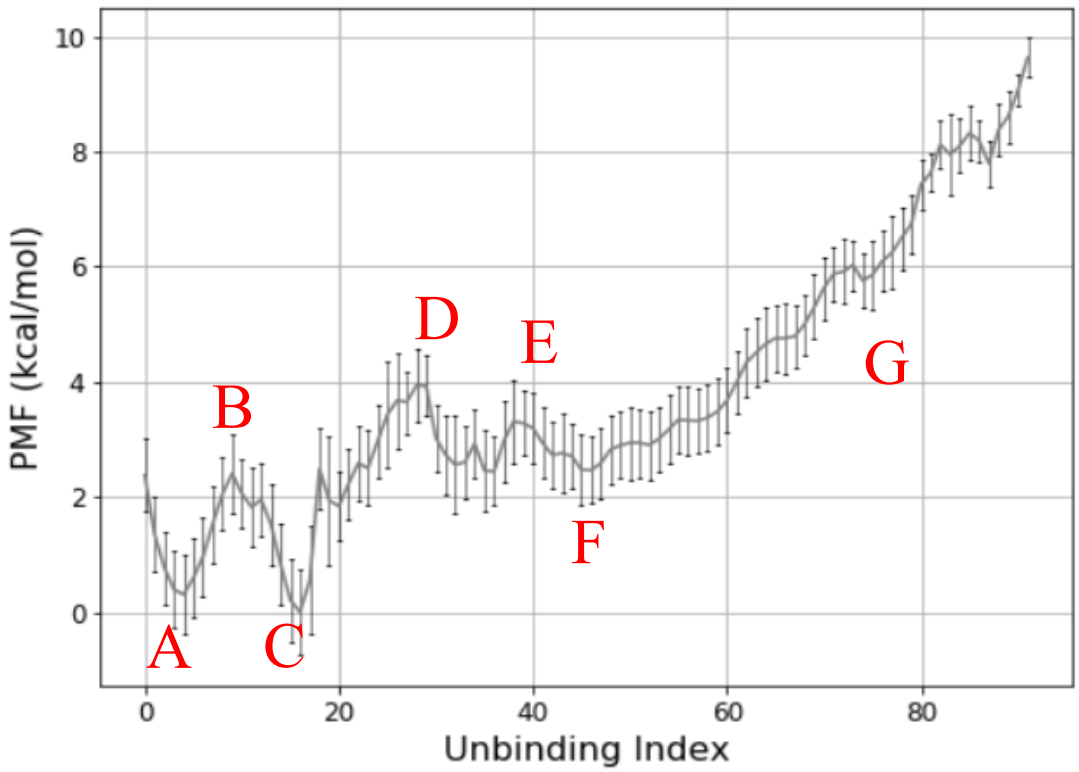


Figure 3.1 Free energy landscape of ritonavir unbinding from HIV protease under pathway A. Major energy barriers/wells are label.

3.4.1 Free energy landscape under pathway A

Using BKiT, we computed unbinding free energy and residence time of ritonavir dissociating under pathway A. We identified four major energy wells and three energy barriers during ritonavir unbinding from HIV protease under pathway A. Energy minima A is the crystal bound state of ritonavir-HIV protease (Figure 3.2). BKiT can correctly capture the crystal bound state. By visualizing ritonavir-HIV protease interactions at energy minima A, we found hydrogen bond between ritonavir and Asp 25, Asp 29, as well as pi-stacking between ritonavir and Arg 107.

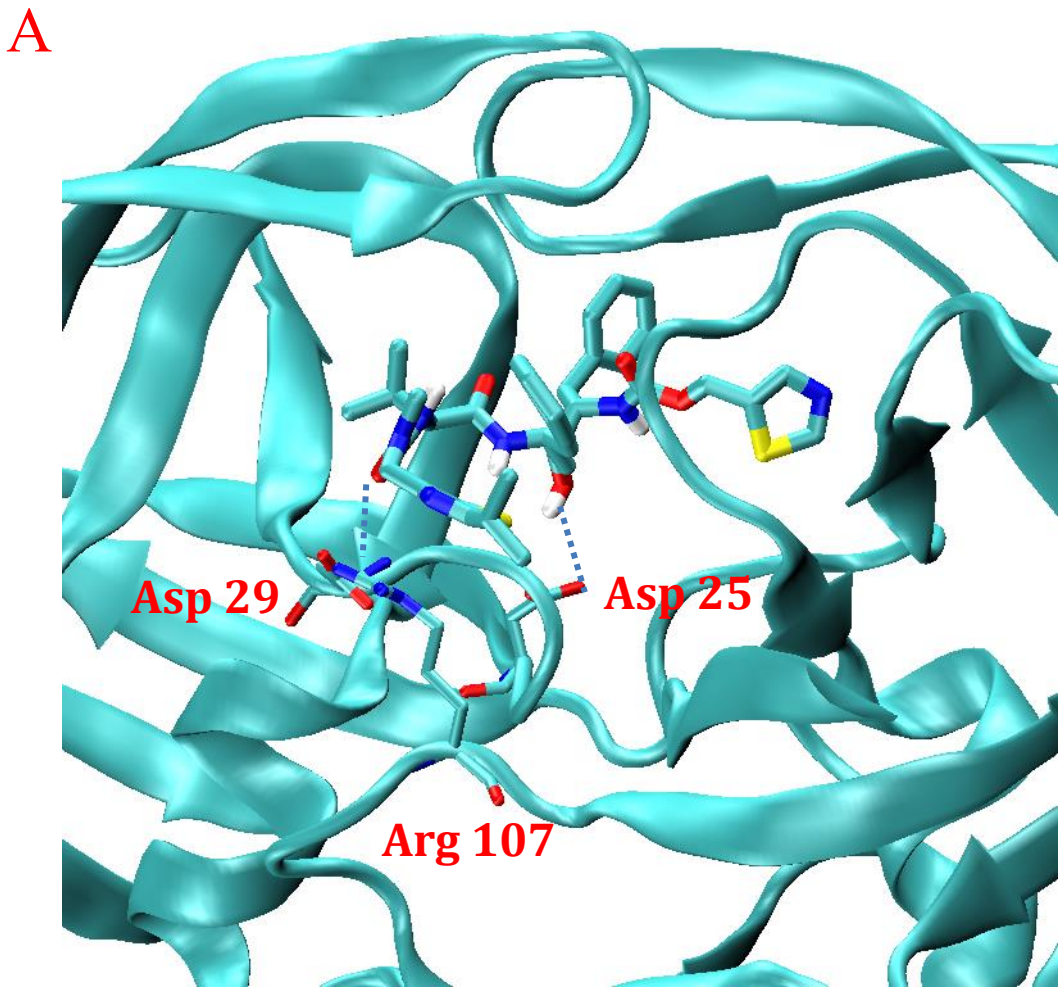


Figure 3.2 Molecular interactions of ritonavir and HIV protease at energy minima A under pathway A.

As system evolved, interactions between ritonavir and HIV protease were broken, giving us an energy barrier B, as shown in Figure 3.3

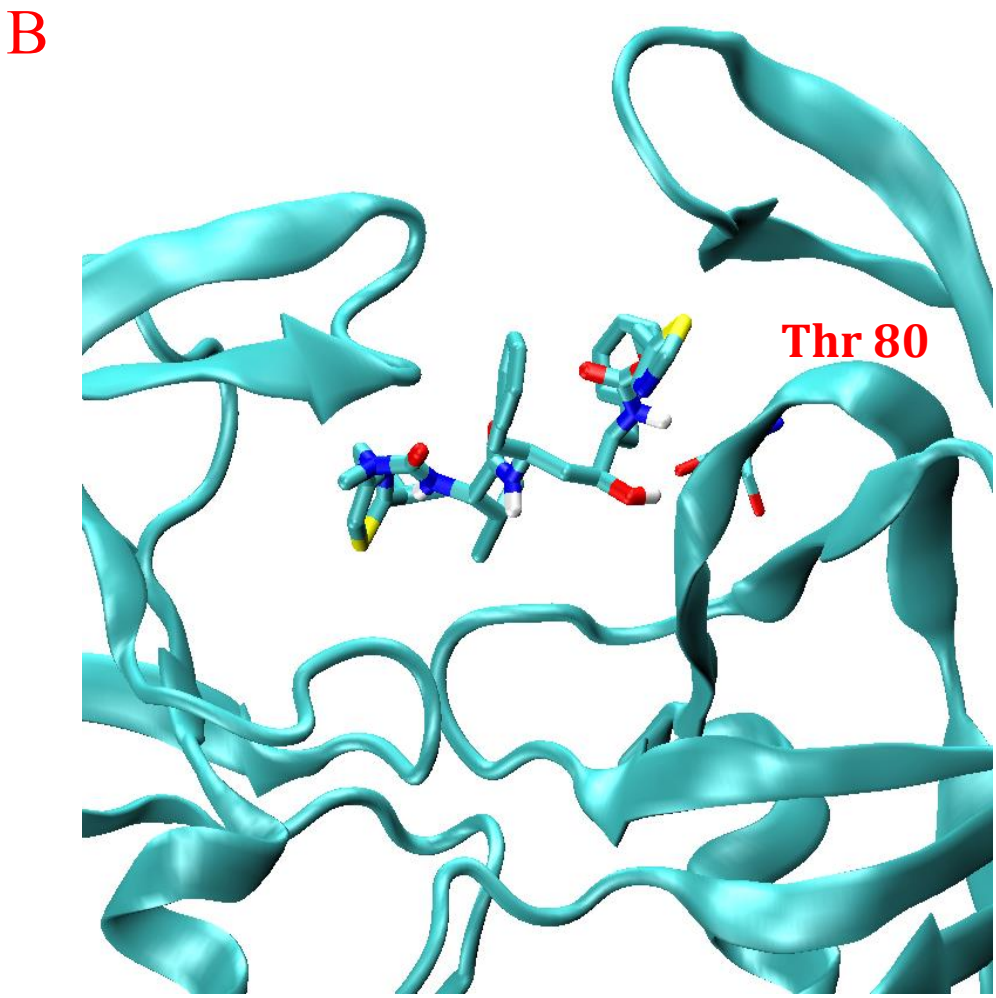


Figure 3.3 Molecular interactions of ritonavir and HIV protease at energy barrier B under pathway A.

As ritonavir approaches Thr 80, a strong hydrogen bond of 1.8 angstrom between the hydroxyl groups of ritonavir and Thr 80 was formed, which further attracted ritonavir to loop region of Thr 80 (Figure 3.4). More interactions were followed by such hydrogen bond. Gly 150 and Ile 149 formed hydrogen bond with ritonavir as well. Non-polar attraction between Ile 146 and ritonavir was also observed.

C

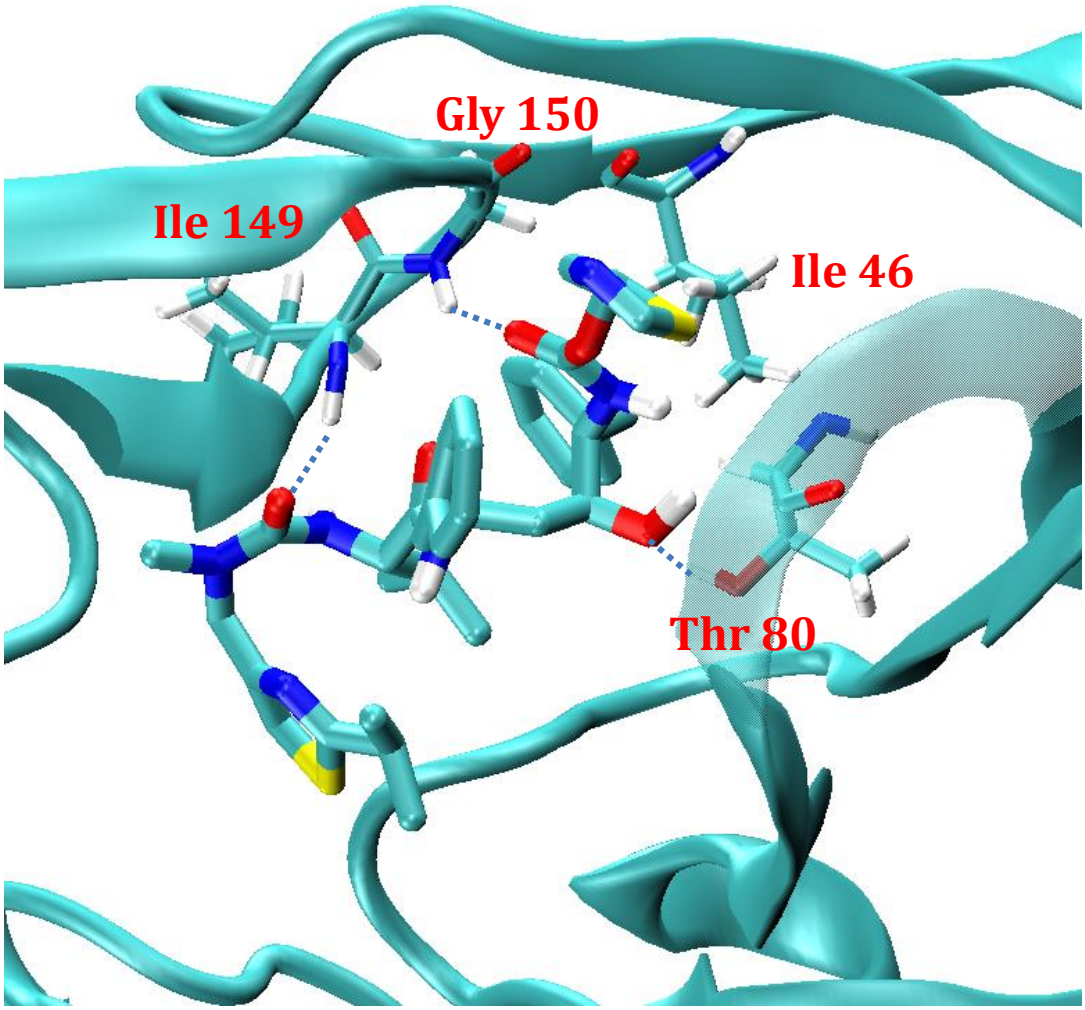


Figure 3.4 Molecular interactions of ritonavir and HIV protease at energy minima C under pathway A.

Flap regions of HIV protease opened, exposed ritonavir to solvent, and reduced contact with ritonavir, giving us an energy barrier D (Figure 3.5). Non-polar attraction was the major interactions between ritonavir and HIV protease.

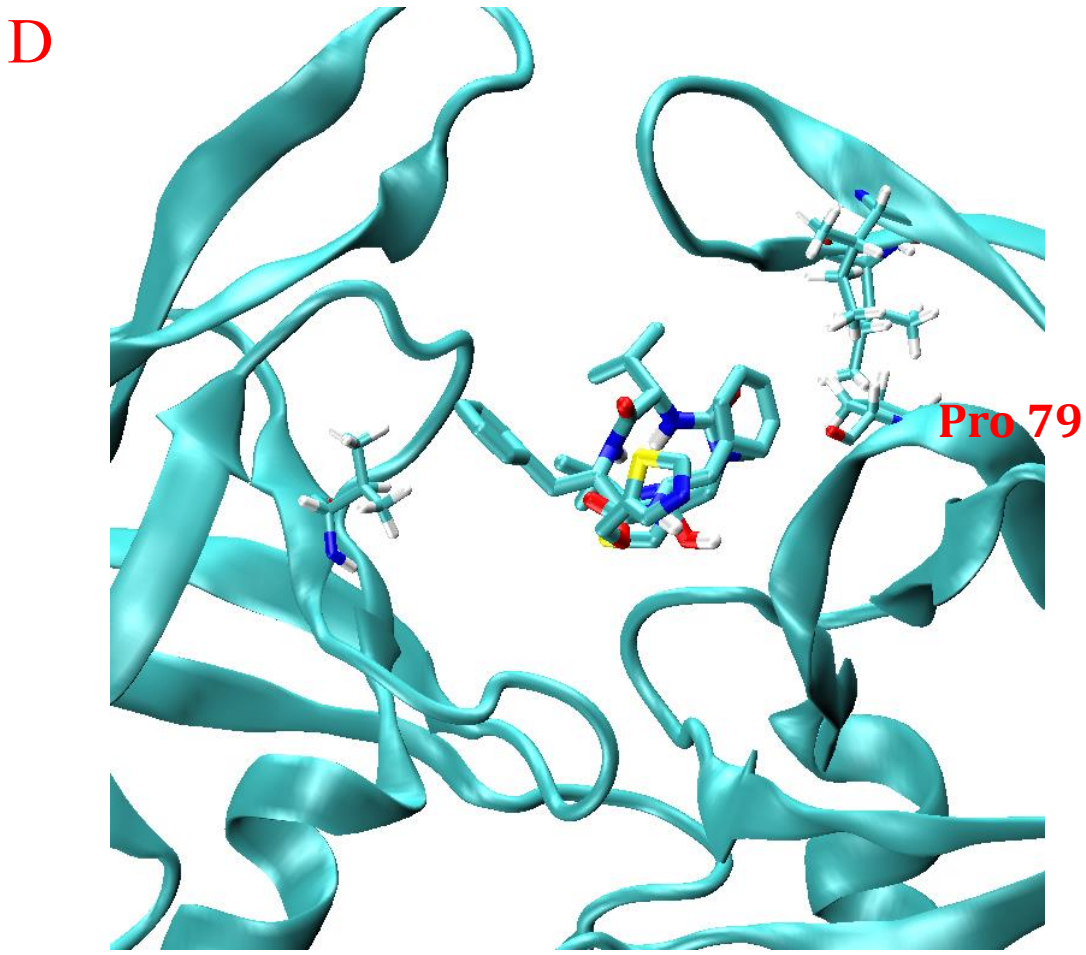


Figure 3.5 Molecular interactions of ritonavir and HIV protease at energy barrier D under pathway A.

As ritonavir keeps moving to left side of HIV protease, flap region on the right side

closed, resulting in a slight lower energy barrier compared to barrier D (Figure 3.6).

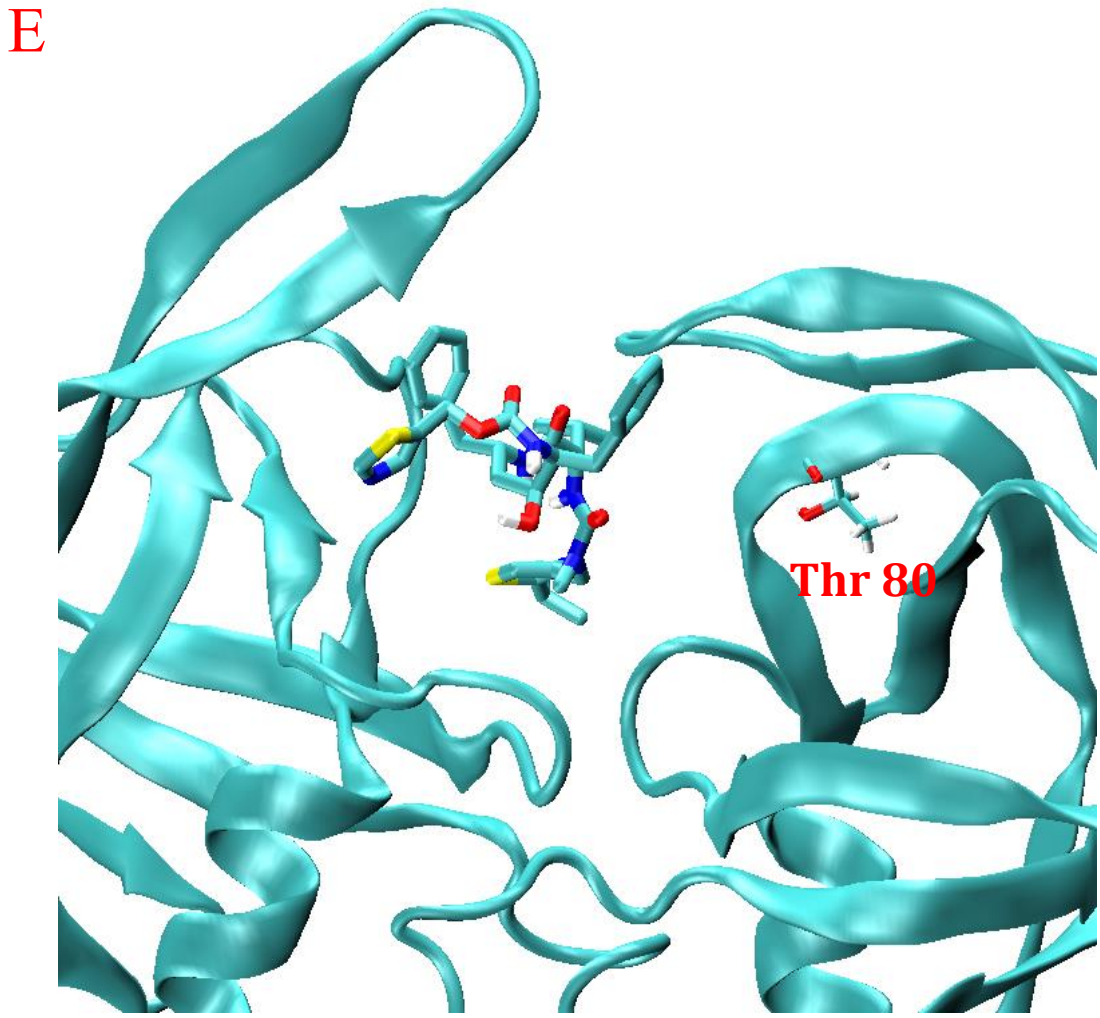


Figure 3.6 Molecular interactions of ritonavir and HIV protease at energy barrier E under pathway A.

Once ritonavir moved to the gap between flap and loop region, a local energy minimum was observed at energy well F (Figure 3.7). Increased contact between ritonavir and HIV protease was the major reason for the stronger attraction.

F

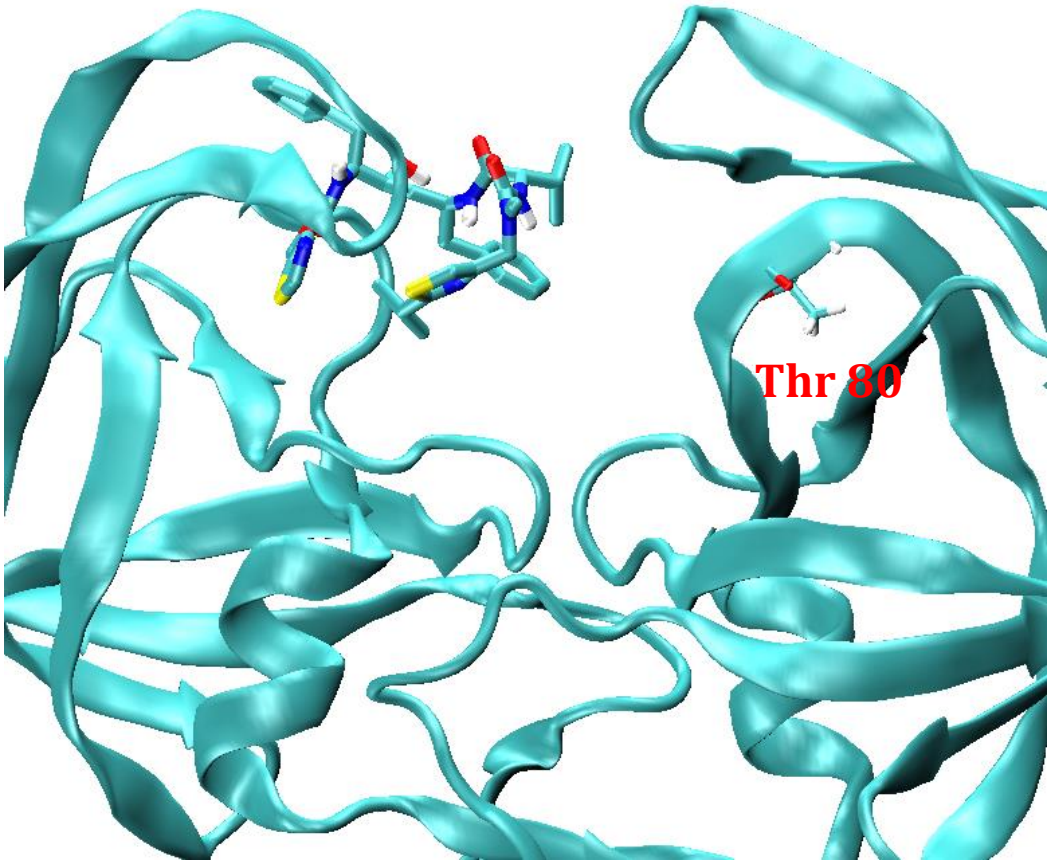


Figure 3.7 Molecular interactions of ritonavir and HIV protease at energy minima F under pathway A.

As ritonavir kept moving along unbinding pathway A, it temporarily formed hydrogen bond with Lys 154, giving us a tiny energy minimum at G (Figure 3.8).

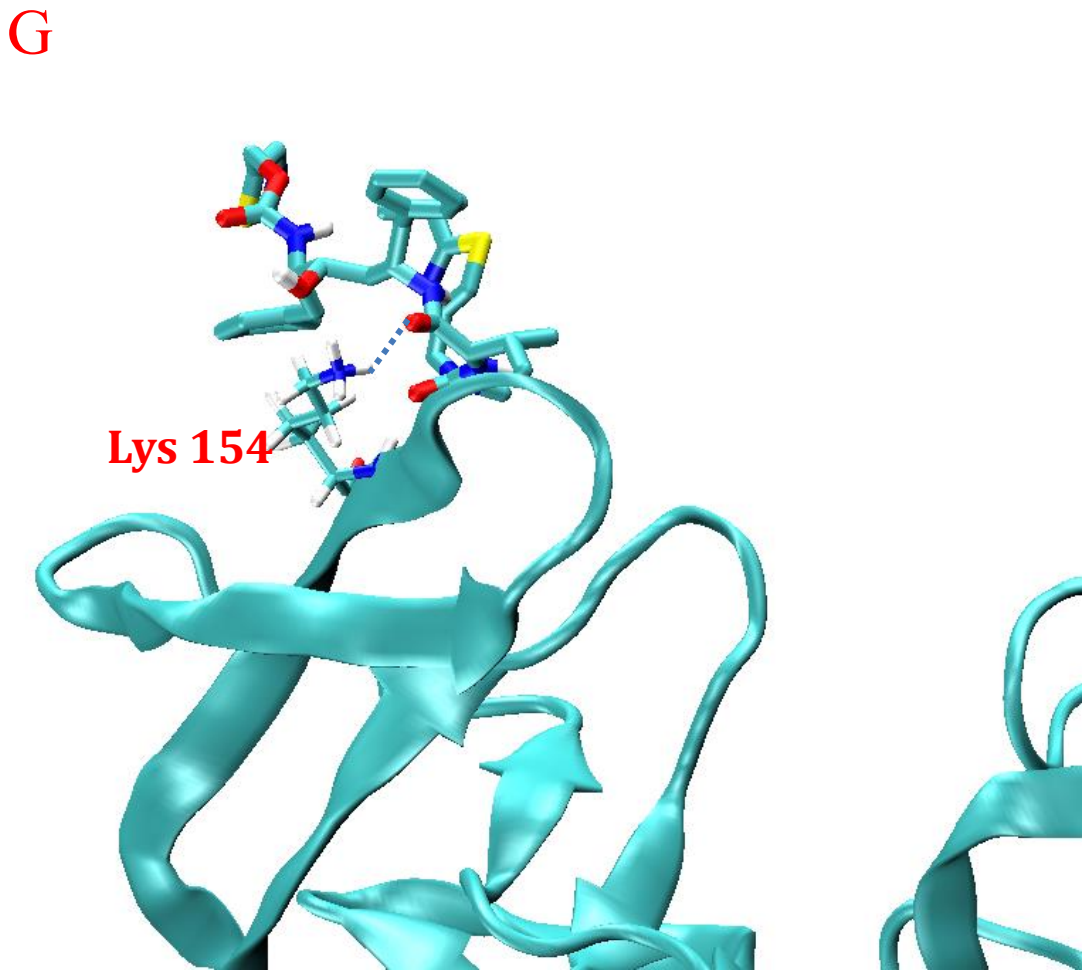


Figure 3.8 Molecular interactions of ritonavir and HIV protease at energy minima G under pathway A.

3.4.2 Free energy landscape under pathway B

We further constructed the free energy landscape of ritonavir unbinding from HIV protease under pathway B (Figure 3.9). Detailed intermolecular interactions are discussed below.

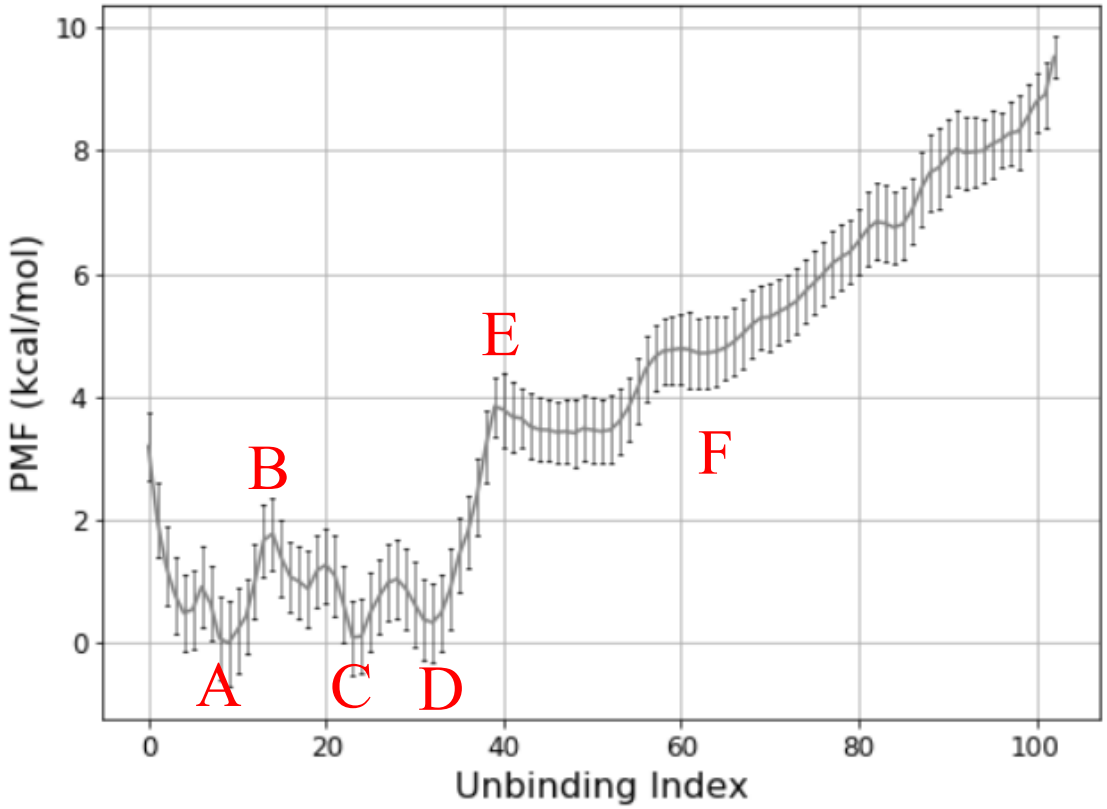


Figure 3.9 Free energy landscape of ritonavir unbinding from HIV protease under pathway B. Major energy barriers/wells are label.

BKiT correctly captured the crystal bound state at energy minimum A and when these interactions were broken at energy barrier B. Since pathway B is ligand diffusion on flap region, minimal interactions between ritonavir and loop region were observed. At energy minimum C, ritonavir formed hydrogen bond with Asp 128, Gly 51 and Gly 148 (Figure 3.10).

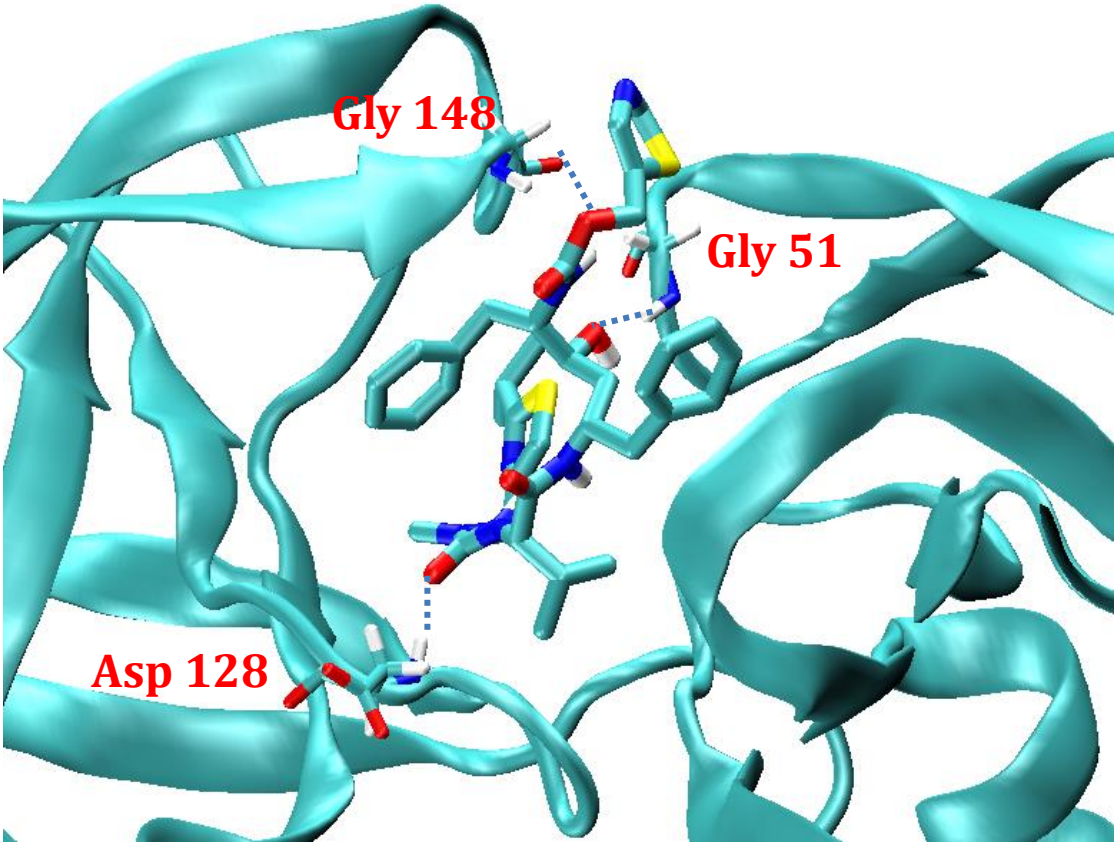


Figure 3.10 Molecular interactions of ritonavir and HIV protease at energy minima C under pathway B.

Energy well D was observed with hydrogen bond between ritonavir and Asp 128, Ile 149

(Figure 3.11).

D

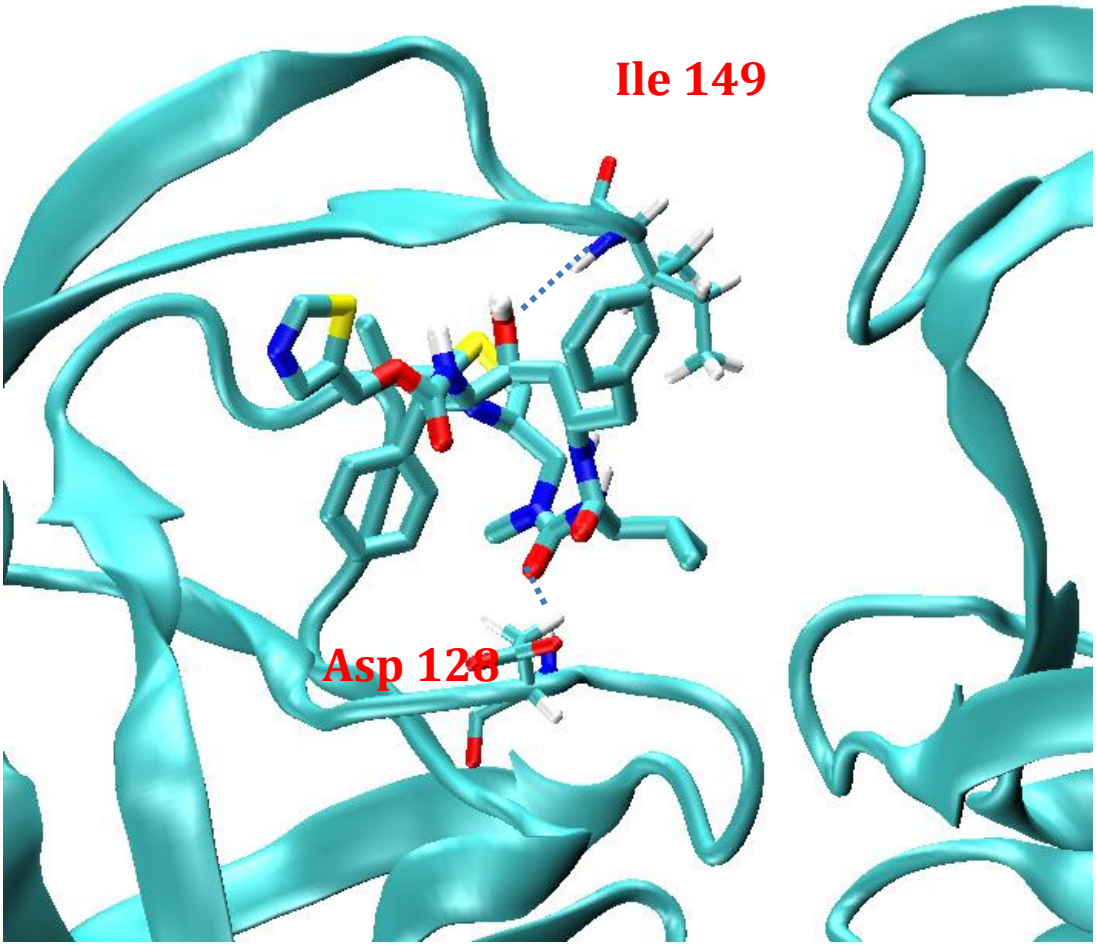


Figure 3.11 Molecular interactions of ritonavir and HIV protease at energy minima D under pathway B.

As flap regions opened widely and exposed ritonavir to solvent, reduced contact between ritonavir and HIV protease, and broken hydrogen bonds, a steep energy barrier showed up from D to E, where minimal contact between ritonavir and HIV protease presented at energy barrier E (Figure 3.12).

E

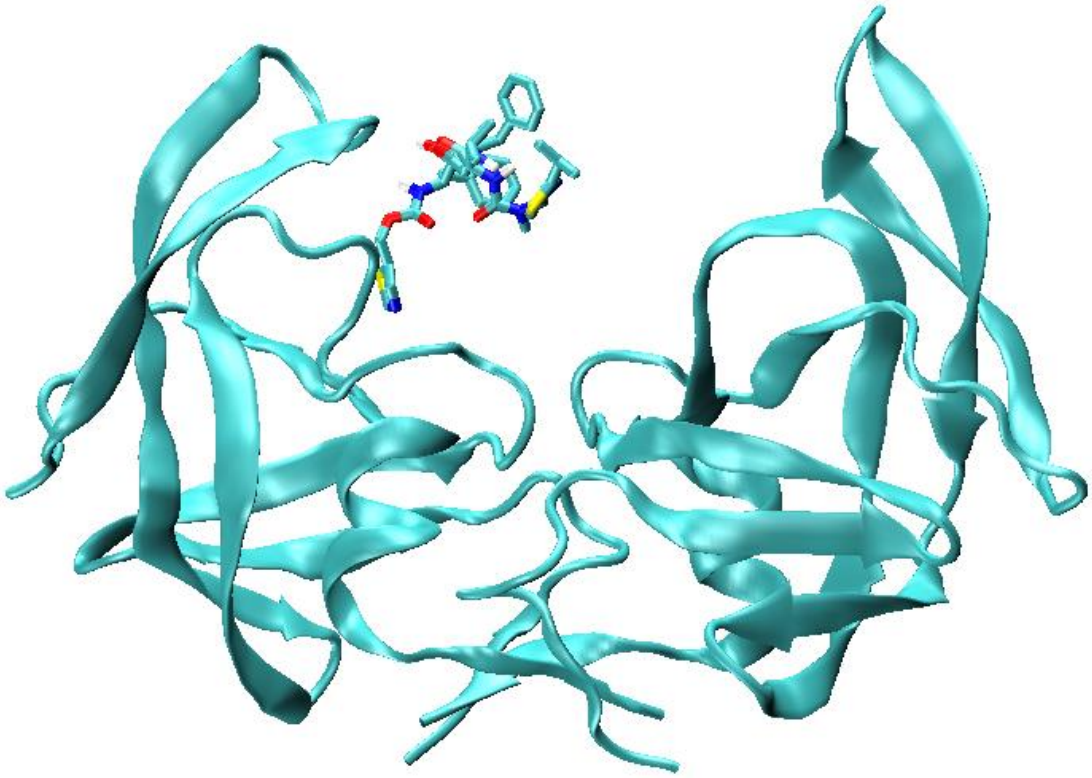


Figure 3.12 Molecular interactions of ritonavir and HIV protease at energy barrier E under pathway B.

As ritonavir moved along flap region, it encountered Lys 142 and formed transient hydrogen bond, giving a small energy well F (Figure 3.13).

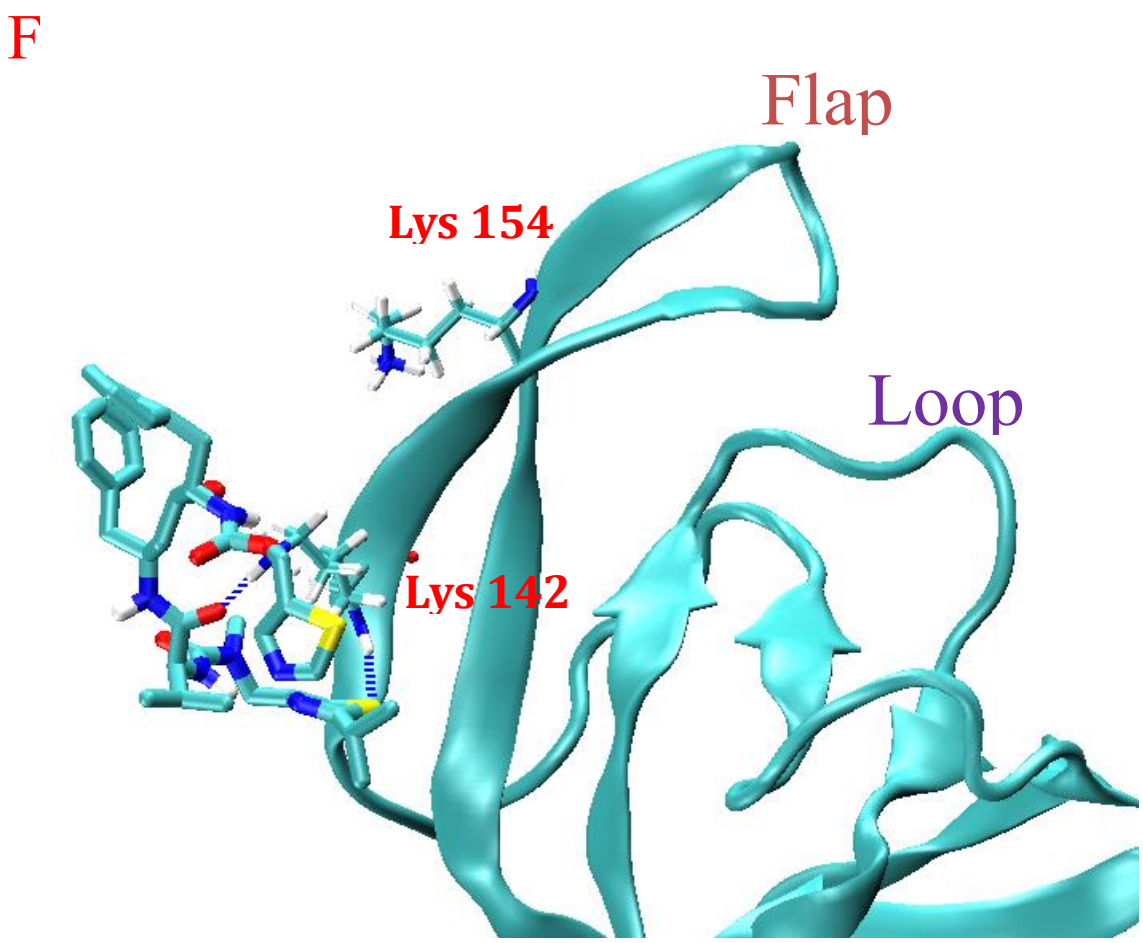


Figure 3.13 Molecular interactions of ritonavir and HIV protease at energy minimum F under pathway B.

3.4.3 Free energy landscape under pathway C

We further examine the free energy landscape of ritonavir unbinding from HIV protease under pathway C. BKiT correctly capture the co-crystal bound state of ritonavir-HIV protease and when they broke the interaction at energy barrier B (Figure 3.14).\

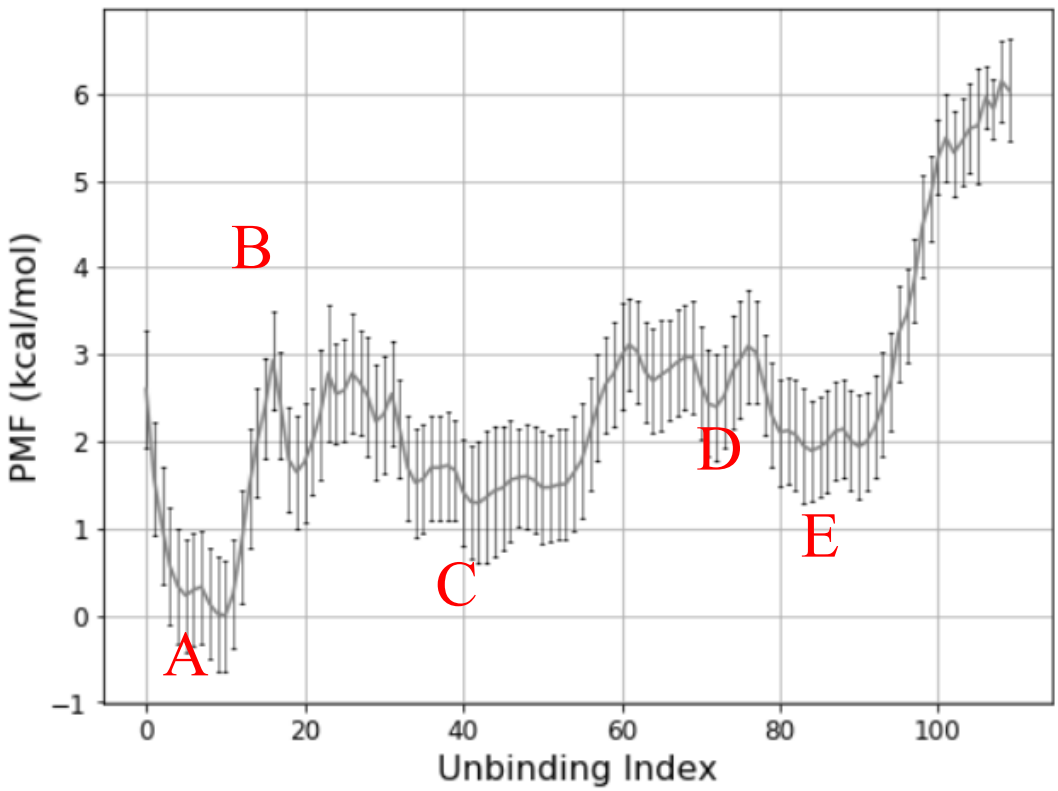


Figure 3.14 Free energy landscape of ritonavir unbinding from HIV protease under pathway C. Major energy barriers/wells are label.

At local energy minimum C, ritonavir was mostly attached to flap region with hydrogen bond to Ile 50, Ile 149, and Gly 151 (Figure 3.15).

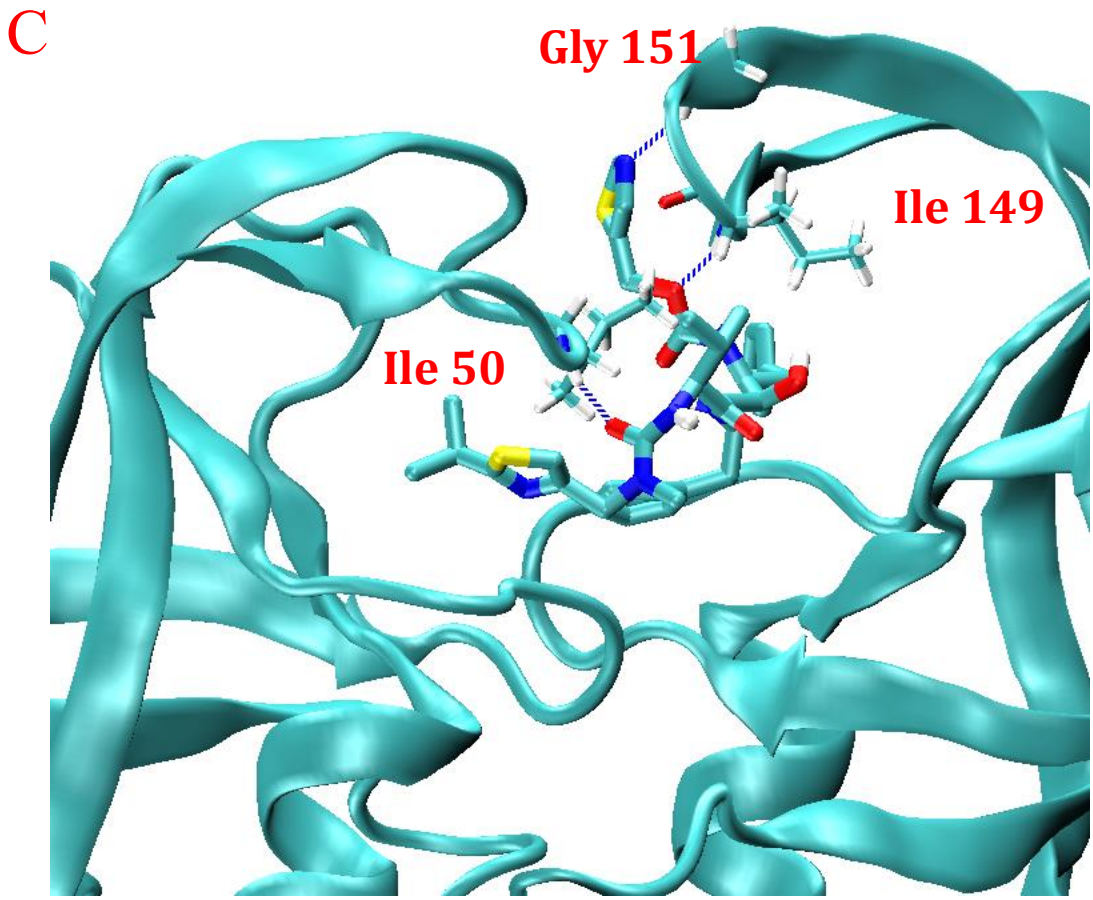


Figure 3.15 Molecular interactions of ritonavir and HIV protease at energy minimum C under pathway C.

More interactions between ritonavir and interface region of HIV protease were observed at energy well D (Figure 3.16). Ritonavir left flap region and started contacting with Arg 8 with two hydrogen bonds, plus hydrogen bond with Asp 129.

D

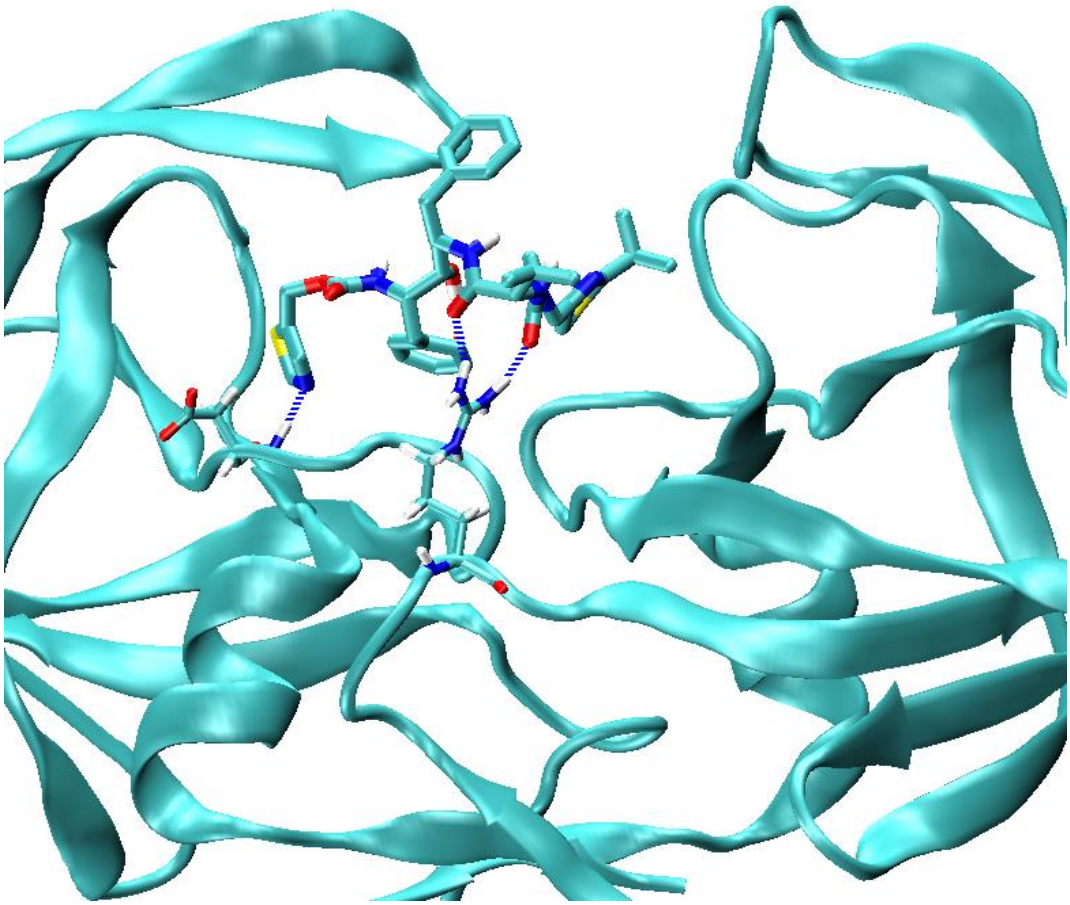


Figure 3.16 Molecular interactions of ritonavir and HIV protease at energy minimum D under pathway C.

As ritonavir kept moving towards interface region, attraction between ritonavir and Arg 8 grew strong. Three hydrogen bonds were observed between ritonavir and Arg 8 at energy well E (Figure 3.17).

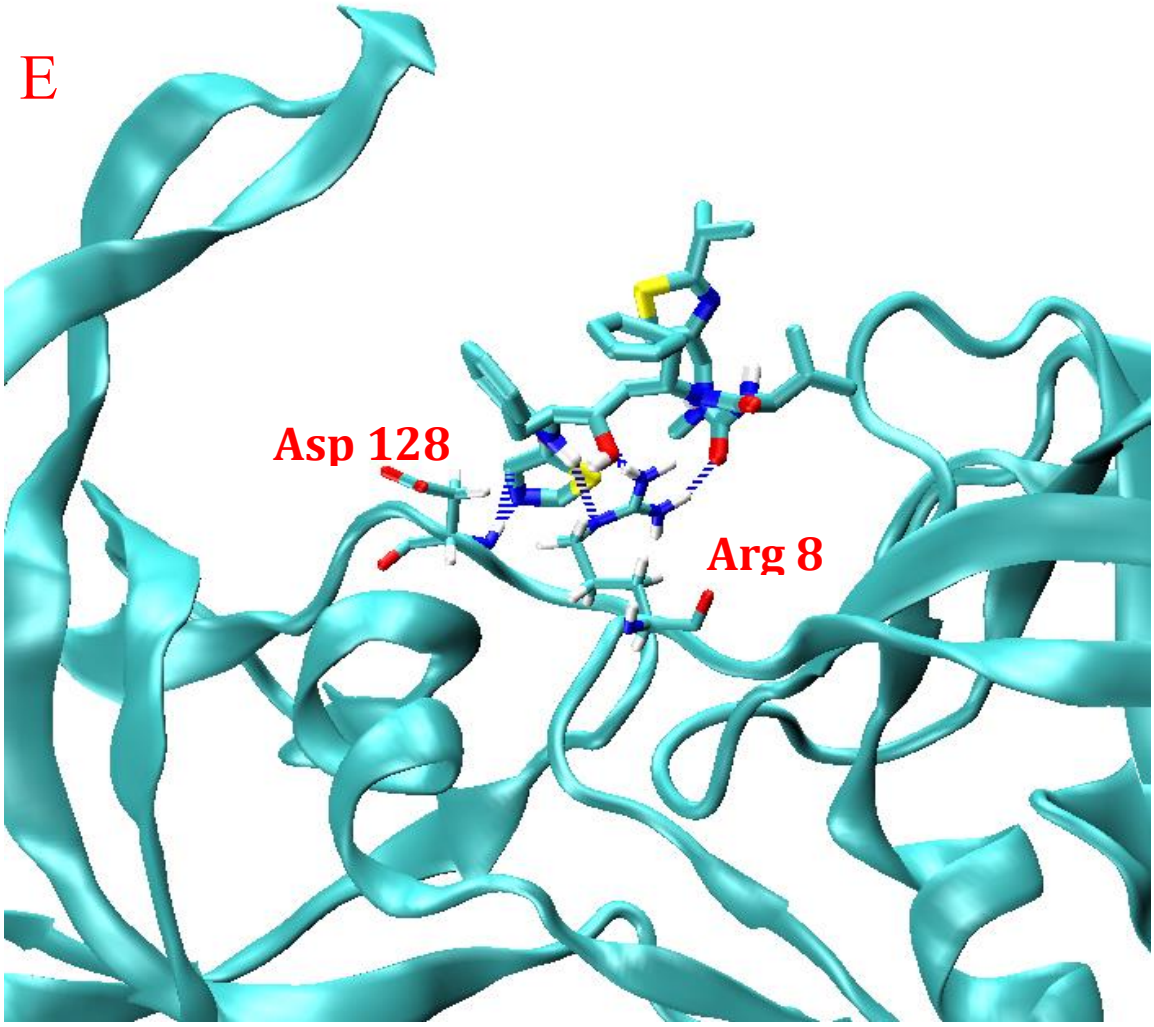


Figure 3.17 Molecular interactions of ritonavir and HIV protease at energy minimum E under pathway C.

3.3.4 Improving unbinding free energy landscape and residence time approximation using ligand RMSD

Using ritonavir dissociation trajectories under different pathways and BKiT, we constructed unbinding free energy landscape and identified key residues that contact the ligand under each pathway. Our computed unbinding free energy and residence time are

slightly off the experimental value with -13.7 kcal/mol for free energy and 463 s for residence time. Improving the estimated result is discussed below.

Principal component analysis captures the major motions of ligand-protein system when we use the cartesian coordinates of alpha-carbon from protein and heavy atoms from ligand. Each PC only covers a percentage of all motions. For example, the combined coverage of PC1 + PC2 for ritonavir unbinding from HIV protease under pathway A was merely 56%, meaning that 44% of all motions were lost in the dimension reduction procedure.

To use a simpler reaction coordinate and make sure that ligand dissociation is fully captured, we assigned unbinding indexes based on ligand RMSD with 0.1 angstrom interval. Then we computed the unbinding free energy landscape and residence time. The newly computed unbinding free energy for ritonavir dissociation from HIV protease under pathway A, B, and C are -14.1 kcal/mol, -15.4 kcal/mol, and -14.9 kcal/mol, respectively (Figure 3.18-20). These free energies are much closer to the experimental data, -13.7 kcal/mol, when compared to the computed data using PC as reaction coordinates. The newly computed drug residence time for ritonavir dissociation from HIV protease under pathway A, B, and C are 2.5 ns, 41.6 ns, and 29.6 ns, respectively. These residence time also showed great improvement from previous results.

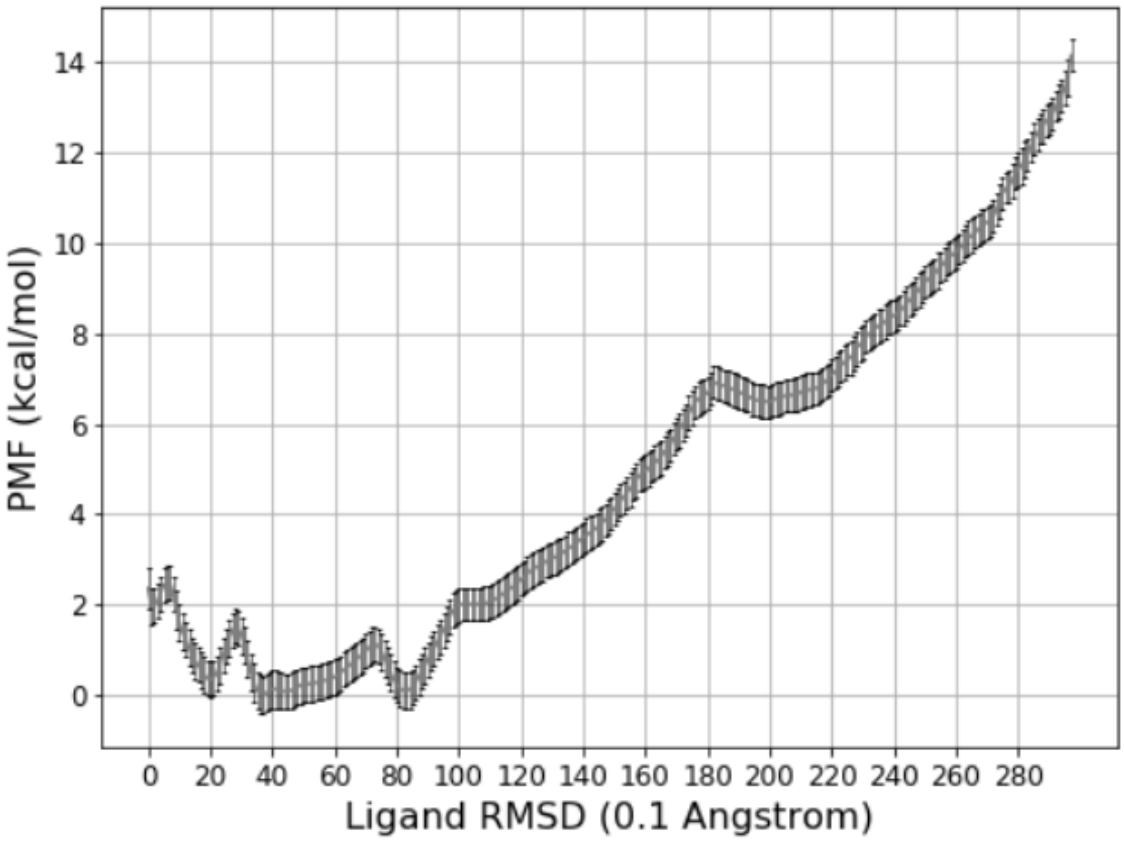


Figure 3.18 Free energy landscape of ritonavir unbinding from HIV protease under pathway A using ligand RMSD as reaction coordinates.

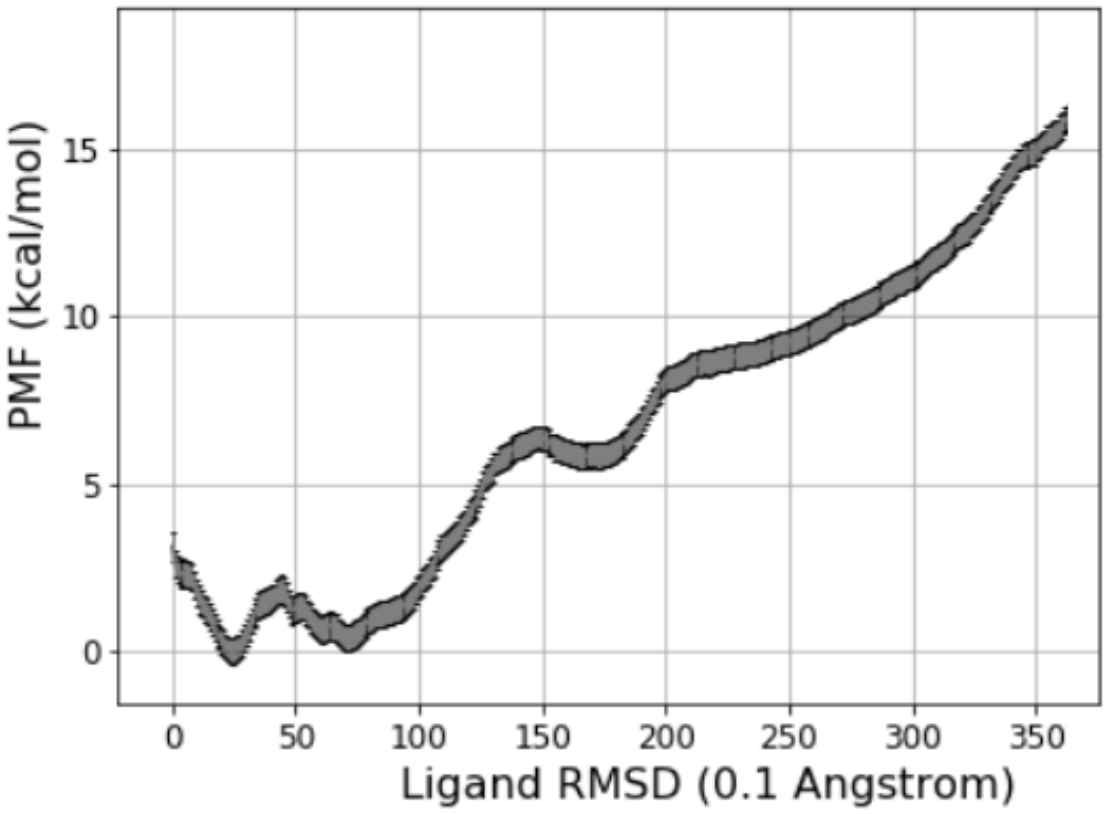


Figure 3.19 Free energy landscape of ritonavir unbinding from HIV protease under pathway B using ligand RMSD as reaction coordinates.

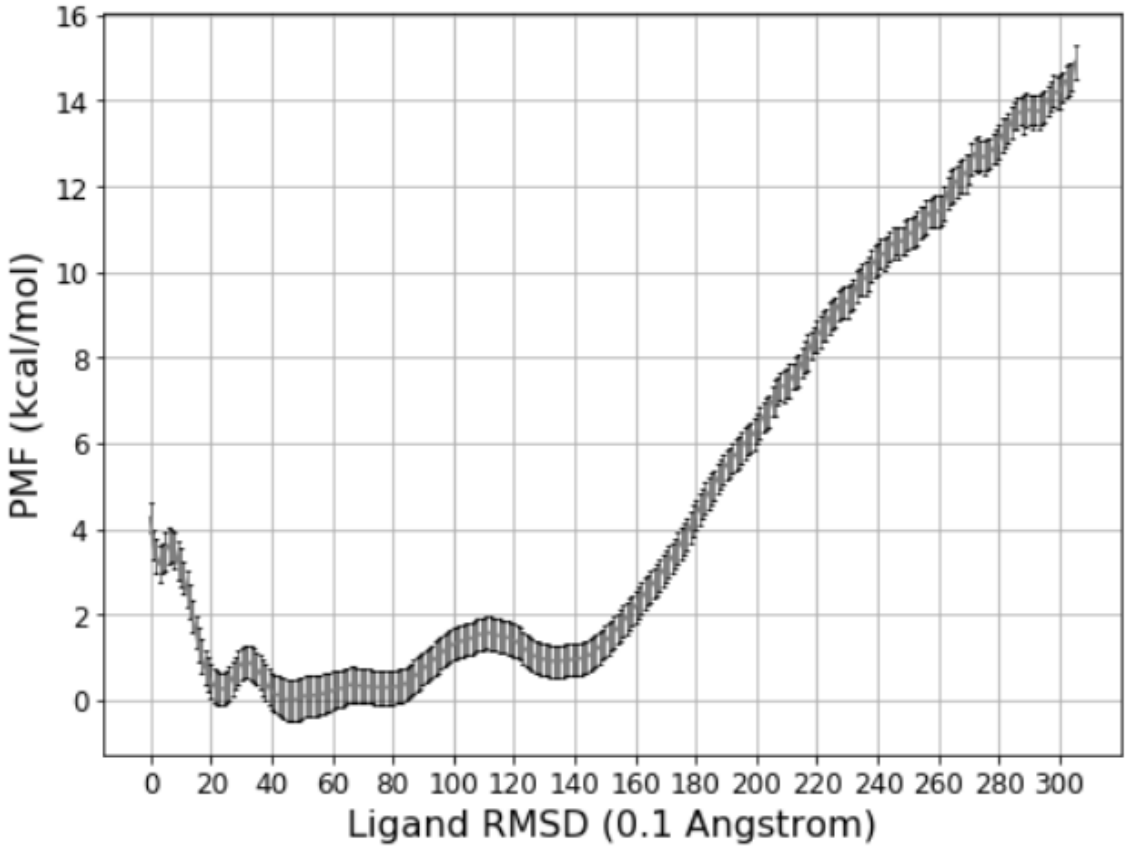


Figure 3.20 Free energy landscape of ritonavir unbinding from HIV protease under pathway C using ligand RMSD as reaction coordinates.

3.5 Conclusion

Our study shows that BKiT is great at capturing detailed ligand-protein interactions during ligand unbinding using PC as reaction coordinates, however, it may give free energy and residence time approximation at moderate accuracy. When we use ligand RMSD as reaction coordinates, we obtained close estimation of unbinding free energy and drug residence time, however, the free energy landscape is not rugged enough to tell us ligand-protein interactions. Combining these two approaches, out BKiT is useful in

understanding molecular recognition during ligand binding/unbinding and computing binding free energy.

3.6 REFERENCES

1. Kohl, N.E.; Emini, E.A.; Schleif, W.A.; Davis, L.J.; Heimbach, J.C.; Dixon, R.A.F.; Scolnick, E.M.; Sigal, I.S. Active Human Immunodeficiency Virus Protease Is Required for Viral Infectivity. Proc. Nat. Acad. Sci. USA 1988, 85, 4686–4690. https://doi.org/10.1073/pnas.85.13.4686.

2. Tomasselli, A.G.; Heinrikson, R.L. Targeting the HIV-Protease in AIDS Therapy: A Current Clinical Perspective. Biochim. Biophys. Acta-Protein Struct. Mol. Enzymol. 2000, 1477, 189–214. https://doi.org/10.1016/s0167-4838(99)00273-3.

3. Weber, I.T.; Wang, Y.F.; Harrison, R.W. HIV Protease: Historical Perspective and Current Research. Viruses 2021, 13. 839. https://doi.org/10.3390/v13050839.

4. Zuo, X.; Huo, Z.; Kang, D.; Wu, G.; Zhou, Z.; Liu, X.; Zhan, P. Current Insights into Anti-HIV Drug Discovery and Development: A Review of Recent Patent Literature (2014–2017). Expert Opin. Ther. Pat. 2018, 28, 299–316. https://doi.org/10.1080/13543776.2018.1438410.

5. Voshavar, C. Protease Inhibitors for the Treatment of HIV/AIDS: Recent Advances and Future Challenges. Curr. Top. Med. Chem. 2019, 19, 1571–1598. https://doi.org/10.2174/1568026619666190619115243.

6. Pawar, S.D.; Freas, C.; Weber, I.T.; Harrison, R.W. Analysis of Drug Resistance in HIV Protease. Bmc Bioinform. 2018, 19. 1–6. https://doi.org/10.1186/s12859-018-2331-y.

7. Whitfield, T.W.; Ragland, D.A.; Zeldovich, K.B.; Schiffer, C.A. Characterizing Protein-Ligand Binding Using Atomistic Simulation and Machine Learning: Application to Drug Resistance in HIV-1 Protease. J. Chem. Theory Comput. 2020, 16, 1284–1299. https://doi.org/10.1021/acs.jctc.9b00781.

8. Nascimento, A.; Fernandes, R.P.; Quijia, C.; Araujo, V.H.S.; Pereira, J.; Garcia, J.S.; Trevisan, M.G.; Chorilli, M. Pharmacokinetic Parameters of HIV-1 Protease Inhibitors. Chemmedchem 2020, 15, 1018–1029. https://doi.org/10.1002/cmdc.202000101.

9. Trylska, J.; Tozzini, V.; Chang, C.-E.A.; McCammon, J.A. HIV-1 Protease Substrate Binding and Product Release Pathways Explored with Coarse-Grained Molecular Dynamics. Biophys. J. 2007, 92, 4179–4187. https://doi.org/.

10. Lexa, K.W.; Carlson, H.A. Binding to the Open Conformation of HIV-1 Protease. Proteins-Struct. Funct. Bioinform. 2011, 79, 2282–2290. https://doi.org/10.1002/prot.23054.

11. Brik, A.; Wong, C.H. HIV-1 Protease: Mechanism and Drug Discovery. Org. Biomol. Chem. 2003, 1, 5–14. https://doi.org/10.1039/b208248a.

12. Copeland, R.A. The Drug-Target Residence Time Model: A 10-Year Retrospective. Nat. Rev. Drug Discov. 2016, 15, 87–95. https://doi.org/10.1038/nrd.2015.18.

13. Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. Annu. Rev. Phys. Chem. 2019, 70, 143–171. https://doi.org/10.1146/annurev-physchem-042018-052340.

14. Lu, H.; Iuliano, J.N.; Tonge, P.J. Structure-Kinetic Relationships That Control the Residence Time of Drug-Target Complexes: Insights from Molecular Structure and Dynamics. Curr. Opin. Chem. Biol. 2018, 44, 101–109. https://doi.org/10.1016/j.cbpa.2018.06.002.

15. Ribeiro, J.M.L.; Tsai, S.T.; Pramanik, D.; Wang, Y.H.; Tiwary, P. Kinetics of Ligand-Protein Dissociation from All-Atom Simulations: Are We There Yet? Biochemistry 2019, 58, 156–165. https://doi.org/10.1021/acs.biochem.8b00977.

16. Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. Int. J. Mol. Sci. 2020, 21, 6339. https://doi.org/10.3390/ijms21176339.

17. Huang, Y.M.M.; Kang, M.; Chang, C.E.A. Switches of Hydrogen Bonds during Ligand-Protein Association Processes Determine Binding Kinetics. J. Mol. Recognit. 2014, 27, 537–548. https://doi.org/10.1002/jmr.2377.

18. Yu, Y.X.; Liu, W.T.; Li, H.Y.; Wang, W.; Sun, H.B.; Zhang, L.L.; Wu, S.L. Decoding Molecular Mechanism Underlying Binding of Drugs to HIV-1 Protease with Molecular Dynamics Simulations and MM-GBSA Calculations. Sar Qsar Environ. Res. 2021, 32, 889–915. https://doi.org/10.1080/1062936x.2021.1979647.

19. Peng, C.; Wang, J.A.; Xu, Z.J.; Cai, T.T.; Zhu, W.L. Accurate Prediction of Relative Binding Affinities of a Series of HIV-1 Protease Inhibitors Using Semi-Empirical Quantum Mechanical Charge. J. Comput. Chem. 2020, 41, 1773–1780. https://doi.org/10.1002/jcc.26218.

20. Gupta, S.; Senapati, S. Mechanism of Inhibition of Drug-Resistant HIV-1 Protease Clinical Isolates by TMC310911: A Molecular Dynamics Study. J. Mol. Struct. 2019, 1198. 126893. https://doi.org/10.1016/j.molstruc.2019.126893.

21. Li, D.C.; Ji, B.H.; Hwang, K.C.; Huang, Y.G. Strength of Hydrogen Bond Network Takes Crucial Roles in the Dissociation Process of Inhibitors from the HIV-1 Protease Binding Pocket. PLoS ONE 2011, 6. e19268. https://doi.org/10.1371/journal.pone.0019268.

22. Huang, S.H.; Zhang, D.; Mei, H.; Kevin, M.; Qu, S.J.; Pan, X.C.; Lu, L.C. SMD-Based Interaction-Energy Fingerprints Can Predict Accurately the Dissociation Rate Constants of HIV-1 Protease Inhibitors. J. Chem. Inf. Model. 2019, 59, 159–169. https://doi.org/10.1021/acs.jcim.8b00567.

23. Ngo, S.T.; Nguyen, M.T. Determination of the Absolute Binding Free Energies of HIV-1 Protease Inhibitors Using Non-Equilibrium Molecular Dynamics Simulations. Chem. Phys. Lett. 2017, 676, 12–17. https://doi.org/10.1016/j.cplett.2017.03.034.

24. Bhattarai, A.; Miao, Y.L. Gaussian Accelerated Molecular Dynamics for Elucidation of Drug Pathways. Expert Opin. Drug Discov. 2018, 13, 1055–1065. https://doi.org/10.1080/17460441.2018.1538207.

25. Furfine, E.S.; D'Souza, E.; Ingold, K.J.; Leban, J.J.; Spector, T.; Porter, D.J. Two-Step Binding Mechanism for HIV Protease Inhibitors. 1992, 31, 7886–7891. https://doi.org/10.1021/bi00149a020.

26. Katoh, E.; Louis, J.; Yamazaki, T.; Gronenborn, A.; Torchia, D.; Ishima, R. A Solution NMR Study of the Binding Kinetics and the Internal Dynamics of an HIV-1 Protease-Substrate Complex. Biol. Magn. Reson. Data Bank 2012.

27. Huang, Y.M.M.; Raymundo, M.A.V.; Chen, W.; Chang, C.E.A. Mechanism of the Association Pathways for a Pair of Fast and Slow Binding Ligands of HIV-1 Protease. Biochemistry 2017, 56, 1311–1323. https://doi.org/10.1021/acs.biochem.6b01112.

28. Qu, S.J.; Huang, S.H.; Pang, X.C.; Yang, L.; Mei, H. Constructing Interconsistent, Reasonable, and Predictive Models for Both the Kinetic and Thermodynamic Properties of HIV-1 Protease Inhibitors. J. Chem. Inf. Model. 2016, 56, 2061–2068. https://doi.org/10.1021/acs.jcim.6b00326.

29. Bruce NJ, Ganotra GK, Kokh DB, Sadiq SK, Wade RC. New approaches for computing ligand–receptor binding kinetics. Current Opinion in Structural Biology. p. 1–10 (2018).

30. Decherchi S, Cavalli A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. Chem Rev. 120:12788–12833 (2020).

31. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks [Internet]. Dill KA, editor. Annual Review of Biophysics, Vol 46. p. 531–558 (2017). Available from: <Go to ISI>://WOS:000402908700024

32. Abel R, Wang LL, Harder ED, Berne BJ, Friesner RA. Advancing Drug Discovery through Enhanced Free Energy Calculations. Accounts of Chemical Research. p. 1625–1632 (2017).

33. Song LF, Lee TS, Zhu C, York DM, Merz KM. Using AMBER18 for Relative Free Energy Calculations. Journal of Chemical Information and Modeling. p. 3128–3135 (2019).

34. Re SY, Oshima H, Kasahara K, Kamiya M, Sugita Y. Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. Proceedings of the National Academy of Sciences of the United States of America. p. 18404–18409 (2019).

35. Ribeiro JML, Tsai ST, Pramanik D, Wang YH, Tiwary P. Kinetics of Ligand-Protein Dissociation from All-Atom Simulations: Are We There Yet? Biochemistry. p. 156–165 (2019).

36. Tang ZY, Chen SH, Chang CEA. Transient States and Barriers from Molecular Simulations and the Milestoning Theory: Kinetics in Ligand-Protein Recognition and Compound Design. Journal of Chemical Theory and Computation. p. 1882–1895 (2020).

37. Tonge PJ. Drug-Target Kinetics in Drug Discovery. Acs Chemical Neuroscience. p. 29–39 (2018).

38. Huang Y-MM. Multiscale computational study of ligand binding pathways: Case of p38 MAP kinase and its inhibitors. Biophys J. 120:3881–3892 (2021). PMCID: PMC8511166

39. Ganesan A, Coote ML, Barakat K. Molecular “time-machines” to unravel key biological events for drug design [Internet]. Wiley Interdisciplinary Reviews-Computational Molecular Science. (2017). Available from: <Go to ISI>://WOS:000403439500003

40. Pan AC, Xu HF, Palpant T, Shaw DE. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. Journal of Chemical Theory and Computation. p. 3372–3377 (2017).

41. Lotz SD, Dickson A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. J Am Chem Soc. 140:618–628 (2018). PMID: 29303257

42. Dixon T, Uyar A, Ferguson-Miller S, Dickson A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. Biophys J. 120:158–167 (2021). PMCID: PMC7820730

43. Scafuri N, Soler MA, Spitaleri A, Rocchia W. Enhanced Molecular Dynamics Method to Efficiently Increase the Discrimination Capability of Computational Protein-Protein Docking. Journal of Chemical Theory and Computation. p. 7271–7280 (2021).

44. Miao YL, McCammon JA. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. Molecular Simulation. p. 1046–1055 (2016).

45. Kokh DB, Amaral M, Bomke J, Gradler U, Musil D, Buchstaller HP, Dreyer MK, Frech M, Lowinski M, Vallee F, Bianciotto M, Rak A, Wade RC. Estimation of Drug-Target Residence Times by tau-Random Acceleration Molecular Dynamics Simulations. Journal of Chemical Theory and Computation. p. 3859–3869 (2018).

46. Schuetz DA, Bernetti M, Bertazzo M, Musil D, Eggenweiler HM, Recanatini M, Masetti M, Ecker GF, Cavalli A. Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics. Journal of Chemical Information and Modeling. p. 535–549 (2019).

47. Barducci A, Bonomi M, Parrinello M. Metadynamics. Wiley Interdisciplinary Reviews-Computational Molecular Science. p. 826–843 (2011).

48. Doshi U, Hamelberg D. Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. Biochimica Et Biophysica Acta-General Subjects. p. 878–888 (2015).

49. Grazioli G, Andricioaei I. Advances in milestoning. I. Enhanced sampling via wind-assisted reweighted milestoning (WARM) [Internet]. Journal of Chemical Physics. (2018). Available from: <Go to ISI>://WOS:000444035800008

50. Zuckerman DM, Chong LT. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. Annual Review of Biophysics, Vol 46. p. 43–57 (2017).

51. Tran DP, Takemura K, Kuwata K, Kitao A. Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics. Journal of Chemical Theory and Computation. p. 404–417 (2018).

52. Wong CF. Steered molecular dynamics simulations for uncovering the molecular mechanisms of drug dissociation and for drug screening: A test on the focal adhesion kinase. Journal of Computational Chemistry. p. 1307–1318 (2018).

53. Tang ZY, Roberts CC, Chang CEA. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. Frontiers in Bioscience-Landmark. p. 960–981 (2017).

54. Ahmad K, Rizzi A, Capelli R, Mandelli D, Lyu W, Carloni P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. Front Mol Biosci. 9:899805 (2022).

55. Kastner J. Umbrella sampling. Wiley Interdisciplinary Reviews-Computational Molecular Science. p. 932–942 (2011).

56. Awasthi S, Kapil V, Nair NN. Sampling Free Energy Surfaces as Slices by Combining Umbrella Sampling and Metadynamics. Journal of Computational Chemistry. p. 1413–1424 (2016).

57. You WL, Tang ZY, Chang CEA. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? Journal of Chemical Theory and Computation. p. 2433–2443 (2019).

58. Elber R. A new paradigm for atomically detailed simulations of kinetics in biophysical systems [Internet]. Quarterly Reviews of Biophysics. (2017). Available from: <Go to ISI>://WOS:000402400900002

59. Maragliano L, Vanden-Eijnden E, Roux B. Free Energy and Kinetics of Conformational Transitions from Voronoi Tessellated Milestoning with Restraining Potentials. Journal of Chemical Theory and Computation. p. 2589–2594 (2009).

60. Votapka LW, Stokely AM, Ojha AA, Amaro RE. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. J Chem Inf Model. 62:3253–3262 (2022).

61. Narayan B, Fathizadeh A, Templeton C, He P, Arasteh S, Elber R, Buchete N-V, Levy RM. The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning. Biochim Biophys Acta Gen Subj. 1864:129508 (2020). PMCID: PMC7012767

62. Chodera JD, Noe F. Markov state models of biomolecular conformational dynamics. Current Opinion in Structural Biology. p. 135–144 (2014).

63. Berezhkovskii AM, Szabo A. Committors, first-passage times, fluxes, Markov states, milestones, and all that [Internet]. Journal of Chemical Physics. (2019). Available from: <Go to ISI>://WOS:000458109300007

64. Husic BE, Pande VS. Markov State Models: From an Art to a Science. Journal of the American Chemical Society. p. 2386–2396 (2018).

65. David CC, Jacobs DJ. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. Protein Dynamics: Methods and Protocols. p. 193–226 (2014).

66. Ahmad M, Helms V, Kalinina OV, Lengauer T. Relative Principal Components Analysis: Application to Analyzing Biomolecular Conformational Changes. J Chem Theory Comput. 15:2166–2178 (2019).

67. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. Journal of Physical Chemistry. p. 2567–2572 (1996).

68. Allison JR. Computational methods for exploring protein conformations. Biochemical Society Transactions. 48:1707–1724 (2020).

69. Noé F, Clementi C. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. Current Opinion in Structural Biology. 41:141–147 (2017).

70. Sidky H, Chen W, Ferguson AL. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. Molecular Physics. 118:e1737742 (2020).

71. Tao P, Sodt AJ, Shao YH, Konig G, Brooks BR. Computing the Free Energy along a Reaction Coordinate Using Rigid Body Dynamics. Journal of Chemical Theory and Computation. p. 4198–4207 (2014).

72. Bai F, Xu Y, Chen J, Liu Q, Gu J, Wang X, Ma J, Li H, Onuchic JN, Jiang H. Free energy landscape for the binding process of Huperzine A to acetylcholinesterase. Proceedings of the National Academy of Sciences of the United States of America. p. 4273–4278 (2013).

73. Amiri S, Amiri S. Cyclodextrins : Properties and Industrial Applications. (2018).

46. Suarez D, Diaz N. Affinity Calculations of Cyclodextrin Host-Guest Complexes: Assessment of Strengths and Weaknesses of End-Point Free Energy Methods. Journal of Chemical Information and Modeling. p. 421–440 (2019).

74. Philip S, Kumarasiri M, Teo T, Yu M, Wang S. Cyclin-Dependent Kinase 8: A New Hope in Targeted Cancer Therapy? Journal of Medicinal Chemistry. p. 5073−5092 (2018).

75. Knuesel MT, Meyer KD, Bernecky C, Taatjes DJ. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. 23:439–451 (2009). PMCID: PMC2648653

76. Schneider EV, Bottcher J, Huber R, Maskos K, Neumann L. Structure-kinetic relationship study of CDK8/CycC specific compounds. Proceedings of the National Academy of Sciences of the United States of America. p. 8081–8086 (2013).

77. Nguyen H, Case DA, Rose AS. NGLview-interactive molecular graphics for Jupyter notebooks. Bioinformatics. 34:1241–1242 (2018). PMCID: PMC6031024

78. Roe DR, Cheatham TE. Parallelization of CPPTRAJ Enables Large Scale Analysis of Molecular Dynamics Trajectory Data. Journal of Computational Chemistry. p. 2110–2117 (2018).

79. Bello-Rivas JM, Elber R. Exact milestoning. The Journal of Chemical Physics. 142:094102 (2015).

80. Tang ZY, Chang CEA. Binding Thermodynamics and Kinetics Calculations Using Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. Journal of Chemical Theory and Computation. p. 303–318 (2018).

Chapter 4 Future works

4.1 Block analysis of dissociation trajectory

On PC space, the trajectory shows trend on PC1/PC2 directions at different time periods, we are working on cutting the whole trajectory into blocks and ensure the system motion within one block is one directional. Then for each block, we use ligand RMSD as reaction coordinates to compute the unbinding free energy and drug residence time within such block. We hope to keep improving the accuracy of free energy approximation of our BKiT package and adding more functionality into BKiT

4.2 Machine Learning using short MDs

Short MDs are essential in construct transition matrix in BKiT and milestoning theory. We would like to utilize the abundance of short MDs and test how these short MDs help with building machine learning models or extracting precious information of ligand-protein molecular recognition.