

Evolution of drug resistance drives destabilization of flap region dynamics in HIV-1 protease

Madhusudan Rajendran,¹ Maureen C. Ferran,¹ Leora Mouli,¹ Gregory A. Babbitt,^{1,*} and Miranda L. Lynch^{2,*}

¹Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York and ²Hauptman-Woodward Medical Research Institute, Buffalo, New York

ABSTRACT The HIV-1 protease is one of several common key targets of combination drug therapies for human immunodeficiency virus infection and acquired immunodeficiency syndrome. During the progression of the disease, some individual patients acquire drug resistance due to mutational hotspots on the viral proteins targeted by combination drug therapies. It has recently been discovered that drug-resistant mutations accumulate on the “flap region” of the HIV-1 protease, which is a critical dynamic region involved in nonspecific polypeptide binding during invasion and infection of the host cell. In this study, we utilize machine learning-assisted comparative molecular dynamics, conducted at single amino acid site resolution, to investigate the dynamic changes that occur during functional dimerization and drug binding of wild-type and common drug-resistant versions of the main protease. We also use a multiagent machine learning model to identify conserved dynamics of the HIV-1 main protease that are preserved across simian and feline protease orthologs. We find that a key conserved functional site in the flap region, a solvent-exposed isoleucine (Ile50) that controls flap dynamics is functionally targeted by drug resistance mutations, leading to amplified molecular dynamics affecting the functional ability of the flap region to hold the drugs. We conclude that better long-term patient outcomes may be achieved by designing drugs that target protease regions that are less dependent upon single sites with large functional binding effects.

WHY IT MATTERS Around the world, HIV/AIDS continues to be a major health problem. Several treatment regimens have dramatically reduced HIV-related morbidity and mortality. HIV-1 protease is one of the critical targets of these regimens. However, given that most protease inhibitors are active site binders, this drug target is prone to development of treatment resistance. Using comparative computer simulations of HIV-1 viral protein motions, we observe that a key residue of the protease flap region targeted by HIV drug therapy became malfunctional in its molecular motions, as has been observed by others. Our analyses contribute to understanding the mechanism of drug resistance and how HIV evolves to evade different drugs. As a result, our study can help inform on the design of future HIV therapeutics that are less prone to the rapid emergence of viral resistance.

INTRODUCTION

In the early stages of the global spread of acquired immune deficiency syndrome (AIDS), infection with human immunodeficiency virus (HIV), the causative agent of AIDS, was essentially a fatal diagnosis, as there were no treatment options to combat progression to AIDS. Since then, however, an arsenal of therapeutics that target the virus has been developed. Drugs are available that target every stage of the HIV replication cycle. Recent information from the US

Department of Health and Human Services HIV/AIDS medical practice guidelines list 23 individual FDA-approved HIV drugs, with an additional 23 unique combination regimens for treating HIV infection (1). These drugs target individual events in viral replication, from viral fusion and entry, to reverse transcription of the HIV RNA genome and integration into the host genome, to viral assembly, budding, and maturation into new infectious particles. For many infected patients now, therapeutic intervention has rendered the disease a chronic, managed condition with reasonable expectation of near-normal life span.

The number of people living with HIV continues to grow, as does the number of deaths due to AIDS. An estimated 38.4 million people were living with HIV in

Submitted February 19, 2023, and accepted for publication August 7, 2023.

*Correspondence: gabsbi@rit.edu or mlynch@hwi.buffalo.edu

Editor: Sarah Rauscher.

<https://doi.org/10.1016/j.bpr.2023.100121>

© 2023 The Authors.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



2021, with a total estimated 650,000 deaths from AIDS in 2021 (2). The widespread use of highly active antiretroviral therapy (HAART) has dramatically reduced HIV-related morbidity and mortality. For example, AIDS-related death has reduced by 68% since its peak in 2004 and by 52% since 2010 (2). The introduction of HIV-1 protease inhibitors and their use in HAART has significantly decreased AIDS-related deaths. HIV-1 protease inhibitors usually bind to the protease's active site and block the cleavage of viral poly-protein precursors resulting in the formation of immature protein precursors, thus forming noninfectious viral particles (3). Yet despite the resounding success of pharmaceutical development for HIV, there remains the widely recognized problem of drug resistance. Drug resistance refers to the failure of a previously successful therapeutic intervention to maintain viral suppression within a patient. Development of resistance mutations in the viral genome undermines the efficacy of modern HIV combination therapy, and resistance monitoring is an important component of a treatment regimen (4,5). As HIV therapeutics span a range of targets in the HIV replication cycle, so too is there a potentially wide range of mechanisms by which therapeutic resistance evolves in a treated patient. There has been tremendous effort to map the viral genomic changes that underlie the evolution of viral escape from therapeutic selective pressures and to comprehend the role of viral transmission in propagating emergent drug-resistant genetic variants. These genetic mutations have been extensively charted and are made available to the research community in a dedicated curated database (6). The functional outcomes of mutations are often altered structures of viral proteins that can imply a malfunctioning biophysical component to the protein's ability to develop a therapeutic escape. Given the biophysical nature of the evolution of drug resistance, a comparative perspective on the mechanistic dynamic changes could be highly informative to future drug design.

Many structural studies of HIV drug resistance have focused on comparative analyses of the drug target structures to locate how drug binding pockets are affected (7–9). An additional consideration for probing HIV drug resistance mechanisms involves looking at not just how protein shape is altered but also how protein dynamics are affected. Viral protein actions are key to how the virus exploits the host cell machinery and also to how drugs bind to and change protein targets (8). Because protein structure is often more conserved than protein sequence, protein dynamics must also be subject to evolutionary conservation (10). In this work, we employ a novel comparative analysis method, applied to molecular dynamics simulations, in order to probe the structural dynamic func-

tioning of both normal wild-type (WT) and emergent drug-resistant viral protein targets. Our work uses machine learning-based comparative analysis tools to uncover how different forms of drug resistance impact the overall dynamics of HIV proteins in consistent ways. We use the DROIDS analysis tool, developed in previous work by our research group, to carry out proper statistical comparisons of many replicates of protein molecular motions modeled via repeated short-term molecular dynamics (MD) simulations, focused on the class of HIV-1 protease inhibitors (PIs) (11,12). Unlike more well-known methods designed for long-term sampling of transitional states in MD (i.e., Markov-state modeling), this method employs site-specific sampling of short-term MD simulations representing two well-defined functional states of proteins (e.g., bound versus unbound or WT versus mutant), so they can be statistically compared at single-site resolution. PIs were one of the first classes of HIV antivirals to benefit from concerted structure-based drug design (13); however analyses of protein dynamics that shed light on structural aspects of viral proteins that contribute to drug resistance have received less attention. Here we are able to demonstrate the role of adaptive evolution to alter the MDs of the flap region of the proteolytic binding pocket of HIV-1 protease in response to the selective pressure of competitive inhibitor drug therapies. These analyses provide an important contribution to our understanding of the mechanisms of HIV drug resistance and could lead to the design of future HIV therapeutics less prone to rapid emergence of viral resistance in patients.

MATERIALS AND METHODS

PDB structure and model preparation

Protein structures of WT and mutant HIV-1 protease bound and unbound to drug inhibitors were obtained from the Protein Data Bank (PDB). Summary of the different PDB structures used for the MD simulation runs are listed in Table 1. After downloading the structures from the PDB database, any crystallographic reflections, ions, and other solvents used in the crystallization process were removed. Any missing loop structures in the protein structures were inferred using the SWISS-MODEL homology modeling server (14,15). Using pdb4amber (AmberTools20), hydrogen atoms were added, and crystallographic water molecules were removed (16).

Molecular dynamic simulation protocols

For each MD comparison (monomer versus dimer, WT versus mutant protease; protease bound and unbound to drug), large replicate sets of accelerated MD simulations were performed. MD simulation protocol was followed as previously described, with slight modifications (11,12,17–19). In brief, for each MD comparison, large replicate sets of accelerated MD simulation were prepared and then conducted using the particle mesh Ewald method implemented on

TABLE 1 List of query and reference protein structures used in this molecular dynamics simulation

Effect of ... on MD	Query protein (PDB ID)	Reference protein (PDB ID)	Figure
Dimerization	WT HIV-1 protease dimer (6DGX)	WT HIV-1 protease monomer (6DGX)	2
Drug binding	WT HIV-1 protease with darunavir (6DGX)	WT apo HIV-1 protease (2PC0)	3, 5A
Drug binding	darunavir-resistant mutant HIV-1 protease with darunavir (6OPV)	WT apo HIV-1 protease (2PC0)	4, 5B
Drug binding	MDR-769 mutant HIV-1 protease with lopinavir (4L1A)	WT apo HIV-1 protease (2PC0)	6
Mutation	MDR-769 with lopinavir (4L1A) darunavir-resistant mutant HIV-1 protease with darunavir (6OPV)	unbound MDR-769 (4L1A) WT HIV-1 protease with darunavir (6DGX)	7
Drug binding	WT HIV-1 protease with pepstatin (5HVP)	WT apo HIV-1 protease (2PC0)	S4
Conserved dynamics	simian SIV protease (1YTG, 1YTH, 1YTI, 1YTJ)	HIV-1 protease monomer (6DGX)	S2, S3
Conserved dynamics	feline FIV protease (1B11)	HIV-1 protease (6DGX)	S2

A100 and V100 NVIDIA graphical processor units by pmemd.cuda running Amber20 (16,20–23). The MD simulations were on a high-performance computing cluster hosted by Rochester Institute of Technology (Rochester, NY) (24). All comparative MD analysis via our Detecting Relative Outlier Impacts in Dynamic Simulation 4.0 (DROIDS 4.0) was based upon 100 replicated sets of 1-ns accelerated MD runs (i.e., 100×1 ns MD run in each comparative state, e.g., monomer versus dimer, WT versus mutant, protease bound to drug versus protease unbound to drug). Explicitly solvated protein systems were first prepared using pdb4amber to add hydrogens and remove crystallographic waters and then solvated and charge-neutralized using tLeap (AmberTools 20), using ff14SSB protein force field, in conjunction with modified GAFF2 small molecule force field (25,26). Solvation was generated using the Tip3p water model in a 12-nm octahedral water box. Charge neutralization was performed using Na^+ and Cl^- ions using the Ambertools20 tLeap program. We note that charge neutralization can be controlled more directly by the user in the MD simulation, via alternate methods allowing for dynamically fluctuating salt pairs (27), and opted to work in neutralized salt environments for our simulation studies. Force field modifications for the small molecule ligands were generated using scaled quantum mechanical optimization via the sqm version 17 program in antechamber/Amber20 (28). For each MD comparison, an energy minimization was first performed and then heated to 300 K for 300 ps, followed by 100 ns of equilibration, a 100 replicate set of random time spacing runs between 0 and 0.5 ns, and then finally a replicate set of 100 MD production runs was created for each comparative state. Each MD production run was simulated for 1 ns of time. All simulations were regulated using the Anderson thermostat at 300 K and 1 atm (29). Root mean-square atom fluctuations and atom correlations were conducted in CPPTRAJ using atomicfluct and atomiccorr commands (30).

Comparative protein dynamic and statistical analyses with DROIDS 4.0

An overview of the DROIDS 4.0 pipeline is available in our software notes (11,12,19). In brief, comparative signatures between the query and reference protein were presented as site-wise divergence in atom fluctuation. Site-wise divergences were calculated using signed symmetric Kullback-Leibler (KL) divergence calculation in DROIDS 4.0. Significance tests and p-values for the site-wise differences were calculated using a two-sample Kolmogorov-Smirnov (KS) test with the less conservative Benjamin-Hochberg multiple test correction. The mathematical and statistical details of DROIDS 4.0 site-wise comparative protein dynamic analysis were published previously by our group (11,12,18,19). Furthermore, the code for our

DROIDS 4.0 pipeline is available at our GitHub web landing: <https://gbabbitt.github.io/DROIDS-4.0-comparativeprotein-dynamics/>, which is also available at our GitHub repository <https://github.com/gbabbitt/DROIDS-4.0-comparative-protein-dynamics>. For readers wishing for tutorial introductions to the software, we specifically highlight reference (19), as well as the tutorials available on the GitHub landing site.

Identifying regions of conserved dynamics with maxDemon 2.0

An overview of the maxDemon 2.0 pipeline using the HIV-1 protease dimer as the query protein and the HIV-1 protease monomer as the reference protein is shown in Fig. S1. In summary, this method trains a multiagent learning model comprised of seven learning methods (adaboost, random forest, decision tree, support vector machines with linear and radial basis function kernels, linear discriminant analysis, quadratic discriminant analysis, naïve Bayes, and K nearest neighbors) on the monomer and dimer dynamic states in human HIV-1 protease. The learner is then deployed upon simulations of evolutionary orthologs in unclassified functional states of dimerization (i.e., the learner attempts to classify the site-wise ortholog dynamics as either human monomer or dimer). The feature vector for the multiagent learner is composed of the atom fluctuations at a given site as well as atom correlations taken at 1, 3, 5, and 9 sites downstream on the structure. The multiagent learner is a stacked model comprised of K nearest neighbors, naïve Bayes, linear discriminant analysis, quadratic discriminant analysis, support vector machine, random forest, and adaptive gradient boosting algorithms employed in R language packages (base, MASS, kernlab, adaboost, and randomForest). This classification is attempted at every 50-frame time slice of a 10-ns simulation, and the frequency of correct classification at each site is calculated. A learning profile curve is generated across all sites of the protein, where values of 0.5 indicate learning is random, and the dynamics of a particular site are indistinguishable as to whether the protease is in a monomerized or dimerized functional state. Values of 0 or 1 on the learning profile indicate 100% successful machine learning classification of the functional states (dimer = 0, monomer = 1). A canonical correlation analysis of the seven learning profiles for each method generated on the human and ortholog dynamics and hypothesis test with Wilk's lambda ($\alpha < 0.01$) was used to identify regions with significantly conserved dynamics. The orthologs we used for human HIV-1 protease (PDB: 6DGX) included several simian SIV proteases and a single feline FIV protease (PDB: 1YTG, 1YTH, 1YTI, 1YTJ, and 1B11) (31,32). For more detailed information, see our previous software publications and protocols (11,12,19).

RESULTS

HIV-1 protease and its inhibitors

The HIV-1 protease is a homodimeric aspartyl protease enzyme, with each monomeric subunit comprising 99 amino acid residues, and with the catalytic aspartic acid (D) residue at position 25 as part of the common triad Asp-Thr-Gly. Fig. 1 shows a ribbon diagram of HIV-1 protease, with main domains indicated. The single catalytic site occurs across the twofold symmetry axis between the two monomers, so only the dimeric form of the protein is active. An important flap domain near the top of the dimer (subunit residues 45–54) undergoes pronounced translocation to create substrate access to the active site. The protease is essential for the HIV-1 replication cycle, as it processes the GAG and GAG-POL polyproteins into the functional components needed for maturation of the intact HIV virion. Structural characterizations of the protein have highlighted key motions of the protease in its catalytic cycle (33,34). The importance of dynamics in HIV-1 protease function led us to examine further using comparative dynamics tools to assess the role of protein motion in the overall evolution of protease inhibitor drug resistance.

HIV-1 protease dimerization causes dampening of atom fluctuation in the flap region

HIV-1 protease exists in its active form as a homodimer. To investigate dimerization's importance, we conducted comparative dynamics between the HIV-1 protease monomer and the HIV-1 protease dimer bound to a peptide ligand. In this particular MD simulation, the monomer was the reference state, and the dimer was the query state (Fig. 2 A). We first looked at the average flux of the dimer and monomer as a function of the amino acid position (Fig. 2 C). The average flux profiles for the monomer and the dimer are comparatively different, as expected. The process of dimerization of the HIV-1 protease has drawn considerable attention, as only the dimeric enzyme is functional, and both experimental and computational studies have attempted to understand the mechanistic steps of dimer formation (35,36). Indeed, prevention of dimerization is a potential therapeutic target (37), and the PI drug darunavir has antidimerization actions in addition to being an active-site binder blocking catalytic function (38). Our study contributes further to this work in its examination of comparative dynamics.

Furthermore, the average atom fluctuation value for the dimer is lower than the monomer, mainly due to

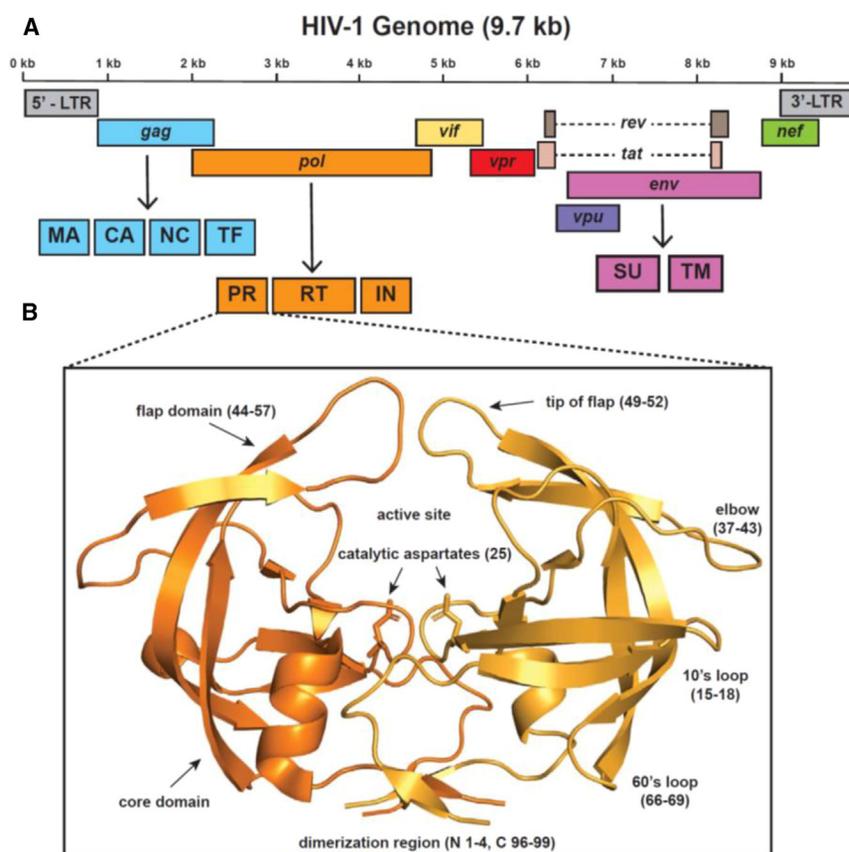


FIGURE 1 Schematic diagram of (A) HIV-1 genome and the (B) protease dimer. (A) The total size of the HIV-1 genome is approximately 9.7 kb. The HIV-1 viral genes are drawn based on the RNA genome's relative orientation. Arrow points to the cleaved protein products. The dashed line represents RNA splicing. LTR, long-term repeat; GAG, group-specific antigen; MA, matrix protein; CA, capsid domain; NC, nucleocapsid; TF, trans-frame protein; Pol, polymerases; PR, protease; RT, reverse transcriptase; IN, integrase; Env, envelope protein; SU, surface membrane protein; TM, transmembrane protein; Vif, viral infectivity factor; Vpr, viral protein R; Vpu, viral protein U; Nef, negative regulatory factor; Rev, regulator of expression of viral proteins; Tat, trans-activator of transcription. (B) The cartoon diagram of HIV-1 protease (PDB: 6DGX) shows the monomers in light orange and dark orange. Key regions of the protease are labeled, and the relevant residue numbers are denoted in parenthesis.

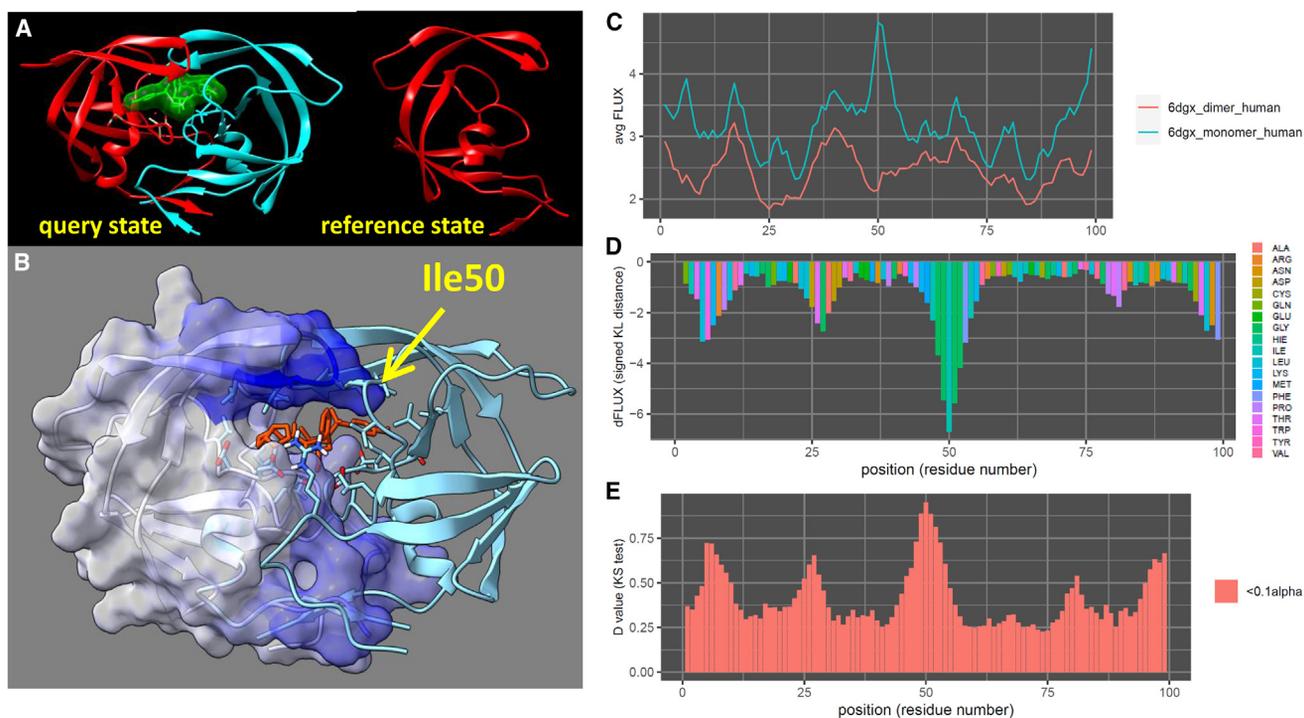


FIGURE 2 Site-wise effects of functional dimerization in HIV-1 protease on molecular dynamics. (A) The molecular dynamics of the HIV-1 protease dimerization function were compared by querying the peptide-bound dimer against the unbound monomer (from PDB: 6dgx). (B) A key site driving functional dimerization is isoleucine at position 50, which is double flanked by glycine (i.e., GGIGG) at the tip of the main protease “flap.” This highly solvent-exposed nonpolar hydrophobic site is highly dampened during dimerization (indicated in blue). This is also represented in the site-wise plots of (C) atom fluctuations in monomer and dimer, (D) signed KL divergence metric, and (E) the two-sample KS hypothesis test as a function of amino acid position. ILE50 and catalytic site are shown in the (D) KL divergence plot.

the structure being stabilized by dimerization and existing with its native polypeptide ligand. To investigate further and tease out any amino acid-specific importance, we looked at the site-wise divergence between the monomer and dimer, which was calculated using signed symmetric KL divergence (Fig. 2 D). We observed negative signed KL divergences at all amino acid positions, indicating a universal dampening of atomic fluctuation at all sites in the dimer state. Furthermore, we observe a more substantial dampening of atomic fluctuation at the catalytic site (Asp25, Thr26, and Gly27) and Ile50 (Fig. 2 B and D). As expected, forming a proper homodimer is essential for an active catalytic site in HIV-1 protease. Other studies have also shown the importance of Ile50, with the position only tolerating a handful of amino acid substitutions (39). Ile50 residue is a part of the flap region and plays a crucial role in stabilizing the protease’s open confirmation. Lastly, we also observe a dampening of atomic fluctuation in the N-terminal residues 1–5 (Fig. 2 D). The N-terminal residues 1–4 contribute to dimer stability in the protease (40). When the dampening of the atomic fluctuations was color mapped to the HIV-1 protease, we observed that the strongest dampening happened across the flap regions of the dimer (Fig. 2 B). To further confirm

that the dampening of atomic fluctuations is significant, we calculated p-values across each amino acid residue using a two-sample KS test with the less conservative Benjamin-Hochberg multiple test correction. We see that the atomic fluctuation differences are significant from the N-terminus to the C-terminus of the protease (Fig. 2 E).

Importance of functionally conserved dynamics in the HIV-1 protease dimerization

Functionally conserved dynamics are defined as repeated, sequence-dependent dynamics discovered after training machine learners on the functional state ensembles derived from our DROIDS pipeline. To detect functionally conserved dynamics after training and validation, an additional new MD run matching the functional reference state is simulated. The learning performance of this run is compared with the MD validation run using a canonical correlation analysis conducted using all selected learners across both space and time (12). Any sequence-dependent or functionally conserved dynamics can be recognized through a significant canonical correlation in the profile of the overall learning performance along the amino acid positions for the two similar

state runs (11,12). See Fig. S1 for a schematic overview of this analysis.

To identify sites of functionally conserved dynamics, the maxDemon multiagent classifier was deployed on each site. From this, a learning performance profile across all sites on the protein is generated. We first compared the multiagent learning methods across HIV-1 protease dimer, HIV-1 protease monomer, feline immunodeficiency virus (FIV) protease monomer, and simian immunodeficiency virus (SIV) protease monomer (Fig. S2). In Fig. S2 A–D, an average classification of 0 or 1 indicated perfect learning classification at a given site, whereas an average learning classification of 0.5 indicates no learning was achieved at that respective site. Robust learning is observed at Ile50, a key residue of the flap region, across all dimers and monomers. As one might anticipate, we also observed strong learning in the catalytic site of the proteases (Asp25, Thr26, and Gly27). We also see near-perfect learning at the N-terminus (1–5), residues that are known to contribute to dimer stability.

To further investigate the importance of Ile50, we identified sites where functionally conserved dynamics are observed using a learning profile. Using the multiagent classifier, we compared the classification of HIV-1 protease and SIV protease (Fig. S3). We observed a highly correlated multiagent learning classification for the dimerized HIV-1 and SIV protease (Fig. S3 A). Notably, a strong majority of amino acid residues (~90%) between the HIV-1 protease and SIV protease dimer show regions of conserved dynamics with high R values from the canonical correlational analysis (Fig. S3 B and C). These analyses suggest that the functional sites involved in dimerization, catalysis, and stabilization in HIV-1 protease have been maintained over viral evolution as different mammalian strains have arisen.

Protease inhibitors cause dampening of atom motions in the flap region of the wild-type HIV-1 protease

The US FDA has approved nine PIs to treat HIV infection. However, some of the inhibitors are no longer used due to their high dosage requirements and side effects. We employed our comparative dynamics framework to investigate binding and behavior of two PIs in current clinical use: darunavir, which was approved in 2006, and lopinavir, approved in 2000. Both of these regimens require boosting with a low dose of the first-generation PI zidovudine, which functions to alter metabolism of the PIs, rendering them more bioavailable.

Darunavir is one such US FDA-approved protease inhibitor with high binding affinity and can be effective against strains where resistance to other inhibitors has developed (41,42). To understand the binding inter-

action of darunavir with HIV-1 protease, we did MD simulations with HIV-1 protease bound and unbound to darunavir. To compare atomic fluctuation between HIV-1 protease bound and unbound to darunavir, we used site-wise KL divergence along with multiple test-corrected two-sample KS tests. The more negative the KL divergence value of a specific amino acid residue is, the stronger is the dampening of atomic fluctuations due to darunavir's interactions with protease. In contrast, the more positive the KL divergence value of a specific amino acid residue is, the stronger is the amplification of the atomic fluctuations due to darunavir's interaction with the protease.

Comparative analyses of protein dynamic simulations of darunavir-bound WT HIV-1 protease (PDB: 6DGX) and an unbound apo form of WT HIV-1 protease (PDB: 2PC0) show strong dampening of the key functional site Ile50 on both chains A and B of the darunavir-bound dimer (Fig. 3) with more pronounced general dampening of atom fluctuation on chain A. We suspect this asymmetry in our result may be due to asymmetry in the protease structure caused by the asymmetry of the darunavir structure itself. We color-mapped the fluctuation of the atom motion. We observed dampening of the atom motion in the tightly closed flap region of the protease when darunavir is bound to HIV-1 protease, indicating that the flap region plays a role in the binding of darunavir (Fig. 3 A–C). Other studies have also found that binding of darunavir alters the flap region, with some studies finding that darunavir binding exhibits a unique curling conformation of the flap region, whereas others observing extended flap conformations (43,44). We also calculated significance tests and p-values for these site-wise differences. We observed significant atomic fluctuation motion in the key functional region near Ile50 in both chains A and B as well as near the catalytic site on chain B (Fig. 3 D and E).

To investigate the consistency of this functional response of the WT HIV-1 protease to other drugs, we conducted the same analysis comparing a pepstatin-bound WT HIV-1 dimer (PDB: 5HVP) to the apo form (PDB: 2PC0) (Fig. S4). Here we observe a similar signature of significant dampening of atom fluctuation at the same key functional site of the flap region. However, the asymmetry in this case shows stronger effect in chain A rather than chain B. Further investigation into the consistency of our results with respect to chain asymmetry in the reference comparison structure is presented in Fig. S5.

Drug-resistant mutations amplify the atom motions in the flap region of the HIV-1 protease

The selection pressure of darunavir has induced a genetic variant that resists the inhibitory effects of the

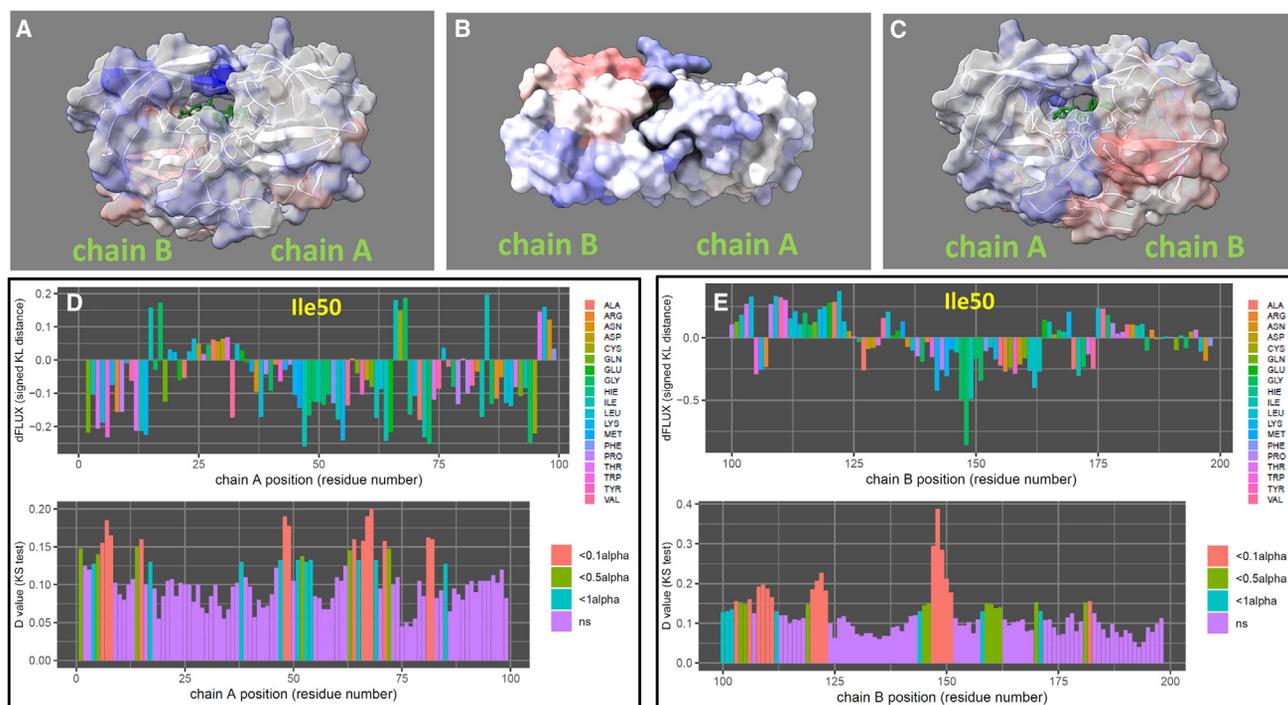


FIGURE 3 Analysis of flap region stabilization during darunavir binding to WT HIV-1 protease. (A–C) The signed Kullback-Leibler (KL) divergence or distance between the distribution in atom fluctuation due to darunavir binding in the main protease dimer is compared in the molecular dynamics (MD) simulations of darunavir-bound wild-type HIV-1 protease dimer (PDB: 6DGX) and wild-type apo HIV-1 protease dimer (PDB: 2PC0) and color mapped to PDB: 6DGX. Three orientations are shown: (A) front, (B) top, and (C) back. The top view is shown without transparency in full ambient occlusion so as to highlight the structural opening/closing of the flap region. Blue denotes a KL divergence representative dampened atom motions in the MD comparison, whereas red denotes amplified motions (range is -0.75 to 0.75). (D and E) Site-wise plots of KL divergence (*top*) and d-values from two-sample Kolmogorov-Smirnov tests corrected for false discovery rate (*bottom*) on (D) chain A and (E) chain B are given for more fine detail. The alpha level for the test is 0.01. ns denotes nonsignificant changes in dynamics at the given site. The key functional site (Ile50) on the protease flap region (see Fig. 2) is also labeled. Also note Y axes are auto-scaled.

drug. The structure of this variant bound to darunavir is databased as PDB: 6OPV. Comparative analyses of protein dynamic simulations of darunavir-bound drug-resistant mutant HIV-1 protease (PDB: 6OPV) and an unbound apo form of WT HIV-1 protease (PDB: 2PC0) show strong significant and symmetric amplification of atom motion at the key functional site Ile50 on both chains A and B of the darunavir-bound dimer (Fig. 4). Close-up comparison of the WT and mutant catalytic site where darunavir binds is shown in Fig. 5.

Like darunavir, lopinavir is another HIV-1 protease inhibitor with high specificity for the protease. We wanted to investigate the importance of the flag region further. MDR769 is an HIV strain that has accumulated multiple drug resistance mutations in the protease, which has resulted in the decreased potency of PIs against HIV. MDR769 consists of 10 amino acid substitutions in the protease: L10I, M36V, M46L, I54V, I62V, L63P, A71V, V82A, I84V, and L90M (45). Therefore, we performed MD simulation using multidrug-resistant 769 (MDR769) strain bound and unbound to lopinavir. Unlike our previous simulation with daruna-

vir, MDR769 is resistant to lopinavir. Compared with WT HIV-1 protease, MDR769 has about 4.3-fold drug resistance against lopinavir (46). Comparative MD simulation of lopinavir-bound MDR769 bound to the apo form of the WT dimer (PDB: 2PC0) also shows strong significant signatures of amplification of atomic motion in both chains of the protease (Fig. 6). This pattern is asymmetric with stronger affect in chain A. The mutation of MDR769 in the flap region results in a "wide-open" structure representing an opening that is 8 Å wider than the "open" structure of the WT protease (47). This is clearly evident in the ambient occlusion image (Fig. 6 B). This prevents clashing of amino acid side chains in the flap region. MDR769 is also associated with a decrease in the volume of amino acid side chains with the active site cavity (48). This provides additional space for the amino acid side chains of the flap region to destabilize.

Lastly, we conducted a comparative MD analysis to isolate the dynamic effects of the key darunavir resistance mutations in PDB: 6DGX from the dynamic effects of binding by comparing the site-wise divergence in atom fluctuations in the darunavir-bound

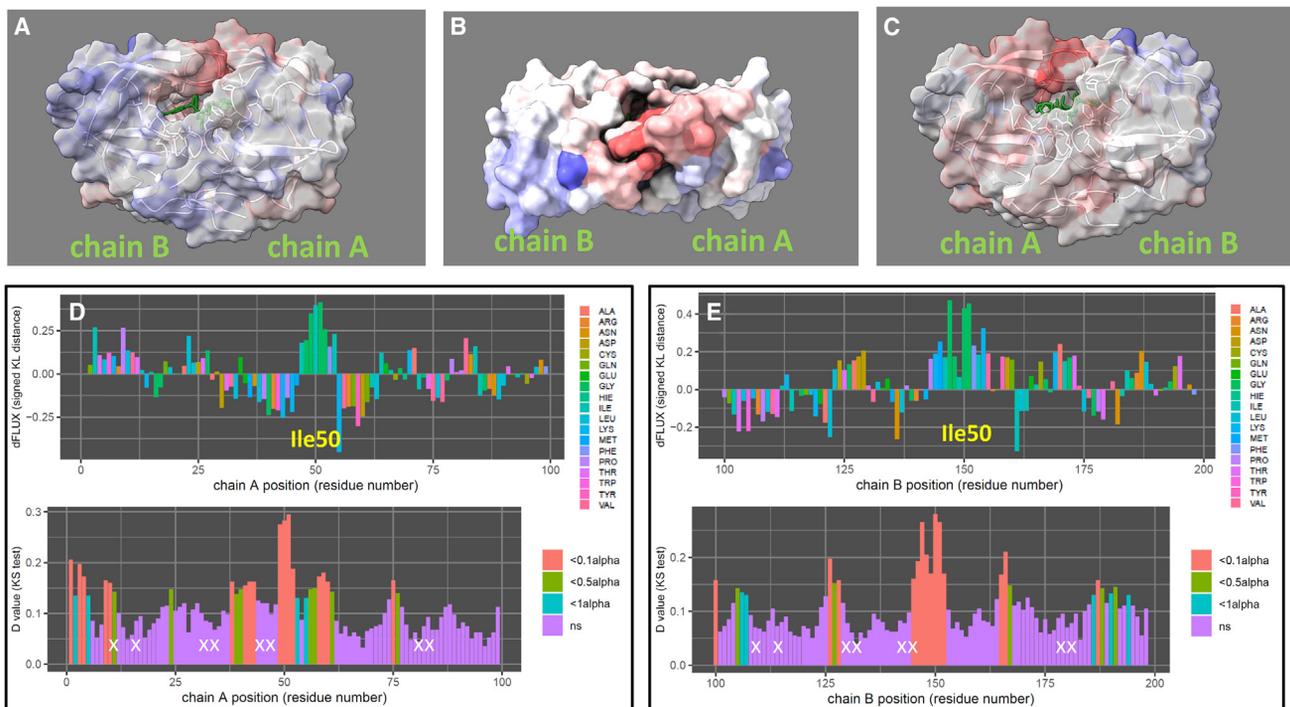


FIGURE 4 Analysis of flap region destabilization during darunavir binding to drug-resistant HIV-1 protease. (A–C) The signed Kullback-Leibler (KL) divergence or distance between the distribution in atom fluctuation due to darunavir binding in the main protease dimer are compared in the molecular dynamics (MD) simulations of darunavir-bound mutant HIV-1 protease dimer (PDB: 60PV) and wild-type apo HIV-1 protease dimer (PDB: 2PC0) and color mapped to PDB: 60PV. Three orientations are shown: (A) front, (B) top, and (C) back. The top view is shown without transparency in full ambient occlusion so as to highlight the structural opening/closing of the flap region. Blue denotes a KL divergence representative dampened atom motions in the MD comparison, whereas red denotes amplified motions (range is -0.75 to 0.75). (D and E) Site-wise plots of KL divergence (*top*) and d-values from two-sample Kolmogorov-Smirnov tests corrected for false discovery rate (*bottom*) on (D) chain A and (E) chain B are given for more fine detail. The alpha level for the test is 0.01. ns denotes nonsignificant changes in dynamics at the given site. The key functional site (Ile50) on the protease flap region (see Fig. 2) is also labeled. The positions of the key drug-resistant mutations (I13V, G16E, V32I, L33F, K45I, M46I, V82F, I84V) are approximated with white “X.” Note Y axes are auto-scaled.

mutant (PDB: 60PV) to the darunavir-bound WT HIV-1 protease (PDB: 6DGX) (Fig. 7). Thus the only difference in the structures and their dynamics was induced by

the substitutions at sites I13V, G16E, V32I, L33F, K45I, M46I, V82F, and I84V. Here, we find a clear, strong, and significant signature of amplified atom

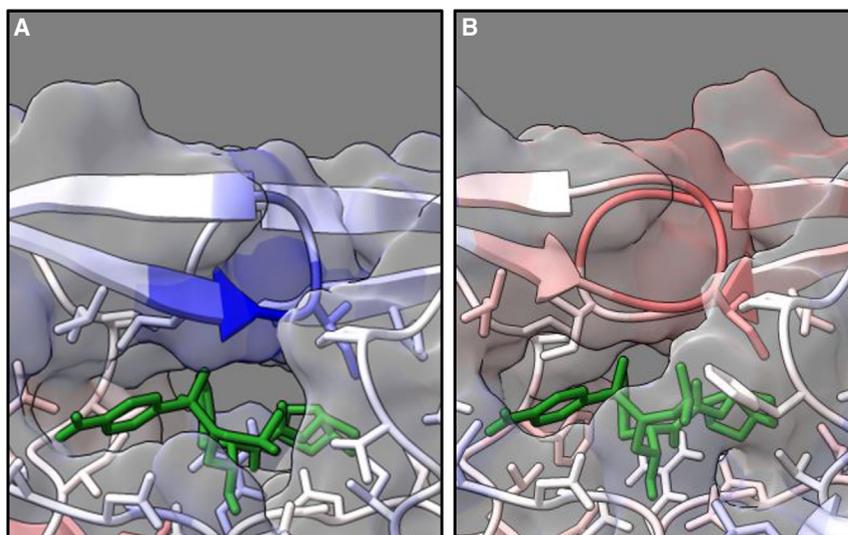


FIGURE 5 Close-up positioning of darunavir in the (A) wild-type HIV-1 protease dimer PDB: 6DGX and (B) drug-resistant HIV-1 protease dimer PDB: 60PV. In the color-mapping, blue denotes a signed KL divergence representative of (A) dampened atom motions in the MD comparisons to apo HIV-1 protease dimer (PDB: 2PC0), whereas (B) red denotes amplified motions in the same MD comparison to PDB: 2PC0 (range is -0.75 to 0.75).

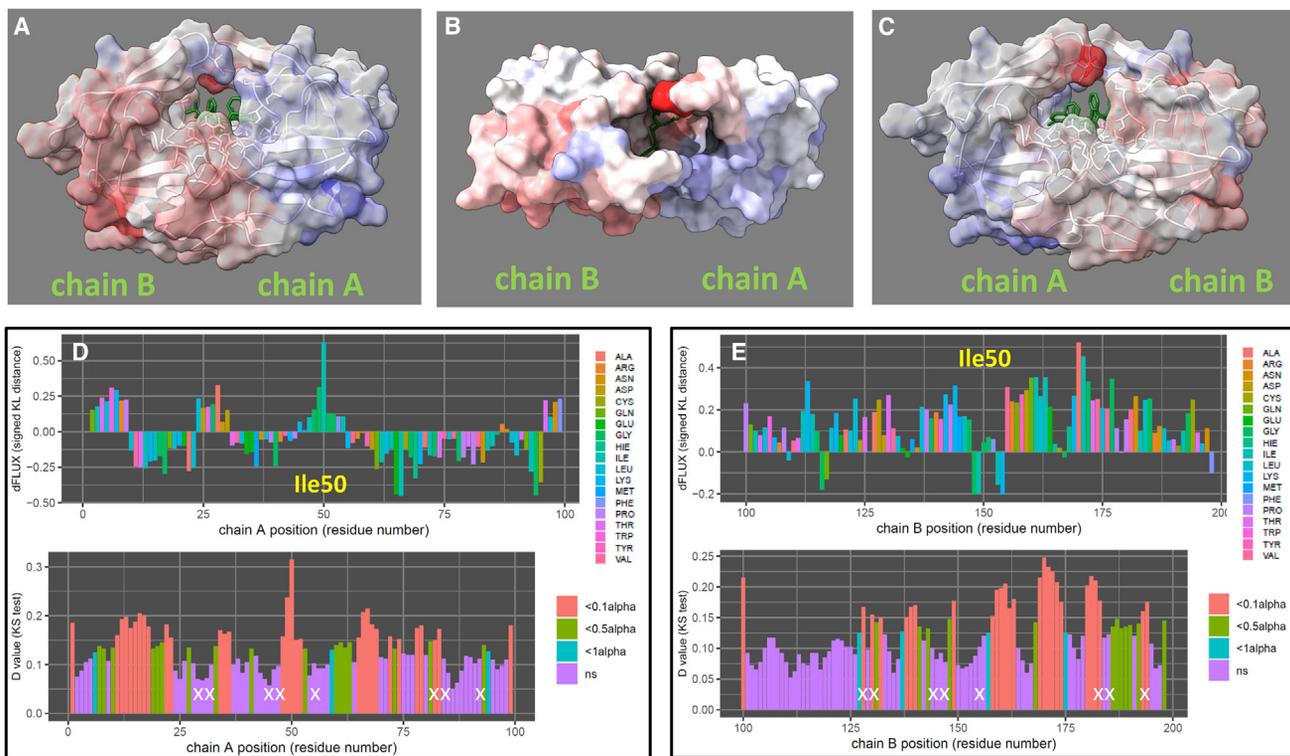


FIGURE 6 Analysis of flap region destabilization during lopinavir binding to multidrug resistant HIV-1 protease MDR-769. (A–C) The signed Kullback-Leibler (KL) divergence or distance between the distribution in atom fluctuation due to lopinavir binding in the main protease dimer is compared in the molecular dynamics (MD) simulations of lopinavir-bound mutant HIV-1 protease dimer (PDB: 4L1A) and wild-type apo HIV-1 protease dimer (PDB: 2PC0) and color mapped to PDB: 4L1A. Three orientations are shown: (A) front, (B) top, and (C) back. The top view is shown without transparency in full ambient occlusion so as to highlight the structural opening/closing of the flap region. Blue denotes a KL divergence representative dampened atom motions in the MD comparison, whereas red denotes amplified motions (range is -0.75 to 0.75). (D and E) Site-wise plots of KL divergence (*top*) and d-values from two-sample Kolmogorov-Smirnov tests corrected for false discovery rate (*bottom*) on (D) chain A and (E) chain B are given for more fine detail. The alpha level for the test is 0.01. ns denotes nonsignificant changes in dynamics at the given site. The key functional site (ile50) on the protease flap region (see Fig. 2) is also labeled. The positions of the key drug-resistant mutations (32I, L33F, 46I, 47A, I54V, V82A, I84V, and L90M) are approximated with white “X.” Note Y axes are auto-scaled.

motion in the flap region that is independent of drug binding, further supporting the functional implication of these amino acid substitutions in combining to alter the flap dynamics in the vicinity of Ile50. Most of the evolutionary changes observed over time in these drug-resistant variants are nearly neutral and do not much alter the amino acid type. The only exceptions to this are the two charge-altering mutations in the darunavir-resistant form (i.e., G16E and K45I). The lopinavir-resistant variant exhibits no mutations that alter amino acid type. This suggests that the combined effect on dynamics is not always easily interpretable from the perspective of amino acid composition and protein sequence alone.

DISCUSSION

Substantial effort has gone into elucidating the important role of protein structure and motion in the function and resistance development of HIV-1 protease. Our comparative dynamics study identifies and under-

scores some of the key molecular motions that play a role in protease inhibitor drug resistance. Our DROIDS 5.0 analytical tool is capable of locating critical dynamics at a residue-level of specificity using site-wise divergences computed from ensembles of MD runs comparing a query and reference protein structures along with statistical tools for evaluation. The inclusion of the stacked machine learner maxDemon 2.0 provides additional information on which protein regions possess functionally conserved dynamics, maintained over evolutionary time since the divergence of simian and feline lineages from that of our own. Our maxDemon analysis probes the orthologs to the viral protease from a human-infecting HIV-1 that exist in viruses infecting these related mammalian backgrounds. The idea is that the SIV and FIV viruses will have evolved proteases within their respective hosts, and that comparison of the orthologs will inform on which functional viral protein motions are retained in the diverse mammalian lineages. We investigated dimerization in HIV-1 protease,

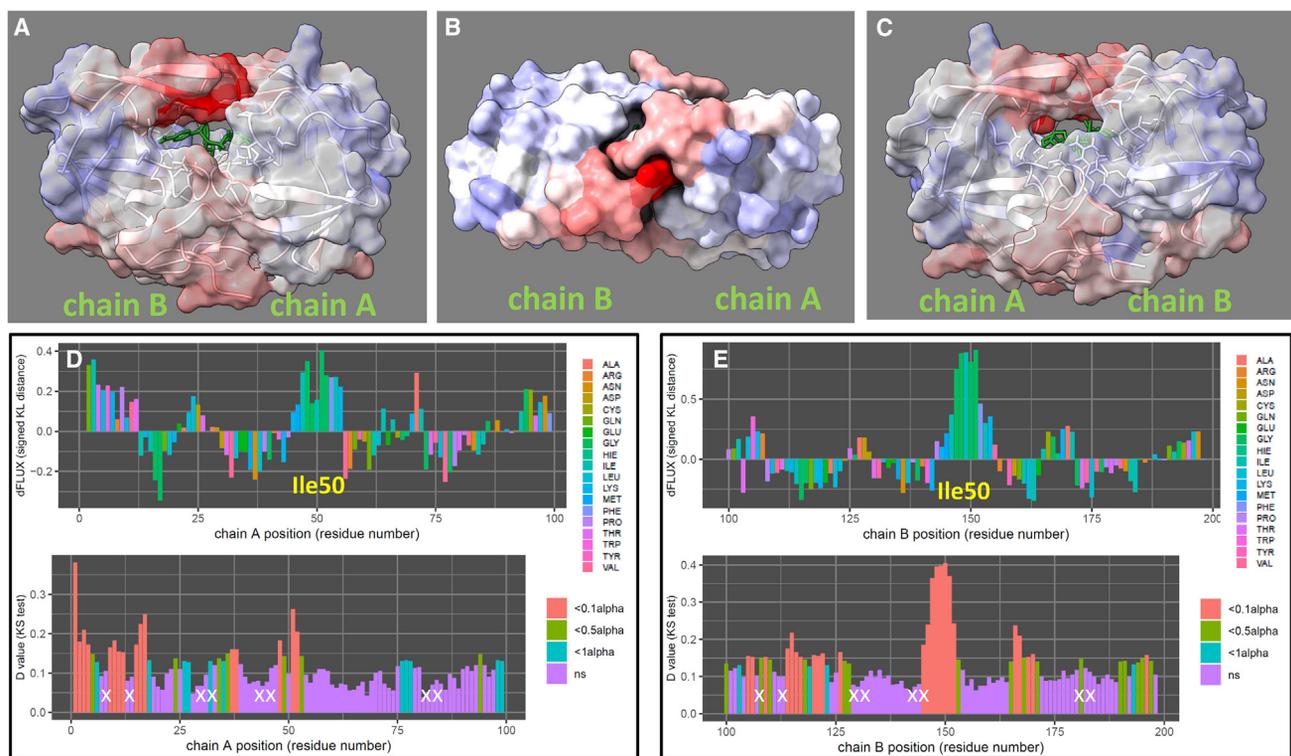


FIGURE 7 Analysis of flap region destabilization due to mutations present in drug-resistant HIV-1 protease. (A–C) The signed Kullback-Leibler (KL) divergence or distance between the distribution in atom fluctuation due to darunavir resistant mutations in the main protease dimer is compared in the molecular dynamics (MD) simulations of darunavir-bound mutant HIV-1 protease dimer (PDB: 6OPV) and wild-type darunavir-bound wild-type HIV-1 protease dimer (PDB: 6DGX) and color mapped to PDB: 6DGX. Three orientations are shown: (A) front, (B) top, and (C) back. The top view is shown without transparency in full ambient occlusion so as to highlight the structural opening/closing of the flap region. Blue denotes a KL divergence representative dampened atom motions in the MD comparison, whereas red denotes amplified motions (range is -0.75 to 0.75). (D and E) Site-wise plots of KL divergence (*top*) and d-values from two-sample Kolmogorov-Smirnov tests corrected for false discovery rate (*bottom*) on (D) chain A and (E) chain B are given for more fine detail. The alpha level for the test is 0.01. ns denotes nonsignificant changes in dynamics at the given site. The key functional site (Ile50) on the protease flap region (see Fig. 2) is also labeled. The positions of the key drug-resistant mutations (I13V, G16E, V32I, L33F, K45I, M46I, V82F, I84V) are approximated with white “X.” Note Y axes are auto-scaled.

a process with strong experimental evidence for structure stabilization upon dimerization with concomitant motion dampening, as a confirmatory check on our methods. Our study detects the global stabilization of the dimeric form, seen in signatures of damped motion of negative KL divergence in atom fluctuation across the entire sequence for the protease monomer compared with the dimer, with functionally important sites like Ile50 showing stronger relative compaction.

Analyses of drug-bound compared with unbound proteases revealed several important trends. First of all, our results often captured asymmetric movements between both chains of the protease homodimer, often reflecting the structural asymmetry caused by the ability of the protease to dimerize on a variety of asymmetric small molecule substrates. Secondly, previous research has focused on expansion of the active site as a mechanistic driver of resistance development (48,49) and has discovered the role of the substrate envelope in influencing drug resistance (50). Our ana-

lyses indicate differential behavior between the individual chains of the homodimeric protease bound to both PIs studied, an experimental result seen previously for darunavir (51). In addition, we were able to observe in the case of both darunavir and lopinavir binding to drug-resistant proteases, a significant malfunctioning amplification of atom motion dynamics in the flap region of the drug/ligand binding pocket, consistent with the expansion of the active site expected for these two multidrug-resistant examples. This mutational shift in dynamics is clearly indicative of a protease that can much more easily release the inhibitory drug and perhaps continue to function normally in the interest of the virus. Very few studies have been able to tease out the more nuanced picture of resistance development that extends beyond the structurally implied “expanded pocket” model. However, our comparative dynamics-based method adds much more additional perspective indicating that the evolution of drug resistance is also generally

accompanied by significant changes in the soft matter biophysics of this region that may not be entirely related to expansion of the shape of the binding pocket but that still might create disruptive changes in response to competitive inhibition induced by current small molecule drugs.

Viral-host interactions are well known to drive adaptive evolution at the protein level (52,53). This has been well documented by decades of work in comparative genomics. However, whereas comparative sequence analyses can easily determine local signals of natural selection acting on proteins (i.e., dN/dS type approaches), they have always been hard pressed to determine what the functional drivers of this evolution are (54,55). We developed our methods largely in response to this “black box” problem in molecular evolutionary studies. Function in molecular biology is ultimately defined by both the structure and soft matter dynamics of proteins. Therefore, we have long conjectured that a proper comparative approach to MD simulations conducted on well-defined functional states (e.g., bound versus unbound or WT versus mutant) can help provide this missing perspective to the functional evolution of many protein systems. In this work, we have clearly demonstrated that adaptive evolution in the HIV-1 main protease in response to the selective pressures of drug therapy has subsequently driven a series of directional mutations that combine to alter the stability of the flap region of the proteolytic binding pocket, thereby altering its ability to function. Our work also demonstrates that whereas this evolution is directed, it can achieve similar destabilization of the biophysics in this region regardless of the specific mutations and drug therapies involved. In summary, our overall findings suggest a common functional evolutionary route leading to most HIV drug resistance to competitive PIs. Future drug therapies that can avoid this rapid viral evolution may have more lasting benefit during the lifespan of the patient.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpr.2023.100121>.

AUTHOR CONTRIBUTIONS

M.L.L., M.C.F., M.R., and G.A.B. designed the study and conducted preliminary model selection and preparation, G.A.B. wrote the software pipeline, and G.A.B., M.R., and L.M. implemented the computational aspects of the research.

ACKNOWLEDGMENTS

No funding supported this work.

DECLARATION OF INTERESTS

All of the authors on this manuscript declare no conflicts of interest.

REFERENCES

1. HIVinfo | Information on HIV/AIDS Treatment, Prevention and Research | NIH. <https://hivinfo.nih.gov/home-page>.
2. Global HIV & AIDS statistics – Fact sheet. <https://www.unaids.org/en/resources/fact-sheet>.
3. Debouck, C. 1992. The HIV-1 Protease as a Therapeutic Target for AIDS. *AIDS Res. Hum. Retrovir.* 8:153–164.
4. Günthard, H. F., V. Calvez, ..., D. D. Richman. 2019. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society–USA Panel. *Clin. Infect. Dis.* 68:177–187.
5. Günthard, H. F., M. S. Saag, ..., P. A. Volberding. 2016. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2016 Recommendations of the International Antiviral Society–USA Panel. *JAMA.* 316:191–210.
6. HIV Drug Resistance Database. <https://hivdb.stanford.edu/>.
7. Tu, X., K. Das, ..., E. Arnold. 2010. Structural basis of HIV-1 resistance to AZT by excision. *Nat. Struct. Mol. Biol.* 17:1202–1209.
8. Brown, K., L. Stewart, ..., E. Lathouwers. 2018. Prevalence of Darunavir Resistance in the United States from 2010 to 2017. *AIDS Res. Hum. Retrovir.* 34:1036–1043.
9. Ren, J., and D. K. Stammers. 2008. Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus Res.* 134:157–170.
10. Gilson, A. I., A. Marshall-Christensen, ..., E. I. Shakhnovich. 2017. The Role of Evolutionary Selection in the Dynamics of Protein Structure Evolution. *Biophys. J.* 112:1350–1365.
11. Babbitt, G. A., J. S. Mortensen, ..., J. K. Liao. 2018. DROIDS 1.20: A GUI-Based Pipeline for GPU-Accelerated Comparative Protein Dynamics. *Biophys. J.* 114:1009–1017.
12. Babbitt, G. A., E. P. Fokoue, ..., L. E. Adams. 2020. DROIDS 3.0—Detecting Genetic and Drug Class Variant Impact on Conserved Protein Binding Dynamics. *Biophys. J.* 118:541–551.
13. Arts, E. J., and D. J. Hazuda. 2012. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb. Perspect. Med.* 2:a007161.
14. Bienert, S., A. Waterhouse, ..., T. Schwede. 2017. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 45:D313–D319.
15. Waterhouse, A., M. Bertoni, ..., T. Schwede. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:296–303.
16. Case, D. A., T. E. Cheatham, ..., R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
17. Rajendran, M., and G. A. Babbitt. 2022. Persistent cross-species SARS-CoV-2 variant infectivity predicted via comparative molecular dynamics simulation [Internet]. *Bioinformatics.* <https://doi.org/10.1101/2022.04.18.488629>.
18. Rajendran, M., M. C. Ferran, and G. A. Babbitt. 2022. Identifying vaccine escape sites via statistical comparisons of short-term molecular dynamics. *Biophys. Rep.* 2, 100056.
19. Babbitt, G. A., E. P. Fokoue, ..., M. Rajendran. 2022. Statistical machine learning for comparative protein dynamics with the DROIDS/maxDemon software pipeline. *STAR Protoc.* 3, 101194.
20. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
21. Ewald, P. P. 1921. Die Berechnung optischer und elektrostatistischer Gitterpotentiale. *Ann. Phys.* 369:253–287.

22. Pierce, L. C. T., R. Salomon-Ferrer, ..., R. C. Walker. 2012. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theor. Comput.* 8:2997–3002.
23. Salomon-Ferrer, R., A. W. Götz, ..., R. C. Walker. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theor. Comput.* 9:3878–3888.
24. Rochester Institute of Technology. 2022. Research Computing Services. <https://www.rit.edu/researchcomputing/>.
25. Wang, J., R. M. Wolf, ..., D. A. Case. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–1174.
26. Maier, J. A., C. Martinez, ..., C. Simmerling. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theor. Comput.* 11:3696–3713.
27. Ross, G. A., A. S. Rustenburg, ..., J. D. Chodera. 2018. Biomolecular Simulations under Realistic Macroscopic Salt Conditions. *J. Phys. Chem. B.* 122:5466–5486.
28. Walker, R. C., M. F. Crowley, and D. A. Case. 2008. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.* 29:1019–1031.
29. Andersen, H. C. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72:2384–2393.
30. Roe, D. R., and T. E. Cheatham. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theor. Comput.* 9:3084–3095.
31. Rose, R. B., C. S. Craik, ..., R. M. Stroud. 1996. Three-Dimensional Structures of HIV-1 and SIV Protease Product Complexes. *Biochemistry.* 35:12933–12944.
32. Li, M., G. M. Morris, ..., A. Gustchina. 2000. Structural studies of FIV and HIV-1 proteases complexed with an efficient inhibitor of FIV protease. *Proteins.* 38:29–40.
33. Heaslet, H., R. Rosenfeld, ..., C. D. Stout. 2007. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr. D Biol. Crystallogr.* 63:866–875.
34. Ragland, D. A., E. A. Nalivaika, ..., C. A. Schiffer. 2014. Drug Resistance Conferred by Mutations Outside the Active Site through Alterations in the Dynamic and Structural Ensemble of HIV-1 Protease. *J. Am. Chem. Soc.* 136:11956–11963.
35. Hayashi, H., N. Takamune, ..., H. Mitsuya. 2014. Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir. *Proc. Natl. Acad. Sci. USA.* 111:12234–12239.
36. Levy, Y., A. Caflich, ..., P. G. Wolynes. 2004. The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations. *J. Mol. Biol.* 340:67–79.
37. Pietrucci, F., A. V. Vargiu, and A. Kranjc. 2015. HIV-1 Protease Dimerization Dynamics Reveals a Transient Druggable Binding Pocket at the Interface. *Sci. Rep.* 5, 18555.
38. Huang, D., and A. Caflich. 2012. How Does Darunavir Prevent HIV-1 Protease Dimerization? *J. Chem. Theor. Comput.* 8:1786–1794.
39. Shao, W., L. Everitt, ..., R. Swanstrom. 1997. Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants. *Proc. Natl. Acad. Sci. USA.* 94:2243–2248.
40. Louis, J. M., R. Ishima, ..., A. M. Gronenborn. 2003. Revisiting Monomeric HIV-1 Protease. *J. Biol. Chem.* 278:6085–6092.
41. De Meyer, S., H. Azijn, ..., M. P. de Béthune. 2005. TMC114, a Novel Human Immunodeficiency Virus Type 1 Protease Inhibitor Active against Protease Inhibitor-Resistant Viruses, Including a Broad Range of Clinical Isolates. *Antimicrob. Agents Chemother.* 49:2314–2321.
42. King, N. M., M. Prabu-Jeyabalan, ..., C. A. Schiffer. 2004. Structural and Thermodynamic Basis for the Binding of TMC114, a Next-Generation Human Immunodeficiency Virus Type 1 Protease Inhibitor. *J. Virol.* 78:12012–12021.
43. Nakashima, M., H. Ode, ..., Y. Iwatani. 2016. Unique Flap Conformation in an HIV-1 Protease with High-Level Darunavir Resistance. *Front Microbiol.* <https://doi.org/10.3389/fmicb.2016.00061/abstract>.
44. Zhang, Y., Y. C. E. Chang, ..., I. T. Weber. 2014. Structures of darunavir-resistant HIV-1 protease mutant reveal atypical binding of darunavir to wide open flaps. *ACS Chem. Biol.* 9:1351–1358.
45. Yedidi, R. S., G. Proteasa, ..., L. C. Kovari. 2014. A multi-drug resistant HIV-1 protease is resistant to the dimerization inhibitory activity of TLF-PafF. *J. Mol. Graph. Model.* 53:105–111.
46. Liu, Z., R. S. Yedidi, ..., L. C. Kovari. 2013. Crystallographic study of multi-drug resistant HIV-1 protease lopinavir complex: Mechanism of drug recognition and resistance. *Biochem. Biophys. Res. Commun.* 437:199–204.
47. Martin, P., J. F. Vickrey, ..., L. C. Kovari. 2005. Wide-Open" 1.3 Å Structure of a Multidrug-Resistant HIV-1 Protease as a Drug Target. *Structure.* 13:1887–1895.
48. Logsdon, B. C., J. F. Vickrey, ..., L. C. Kovari. 2004. Crystal Structures of a Multidrug-Resistant Human Immunodeficiency Virus Type 1 Protease Reveal an Expanded Active-Site Cavity. *J. Virol.* 78:3123–3132.
49. Sheik Amamuddy, O., N. T. Bishop, and Ö. Tastan Bishop. 2018. Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics. *Sci. Rep.* 8, 17938.
50. Kurt Yilmaz, N., R. Swanstrom, and C. A. Schiffer. 2016. Improving Viral Protease Inhibitors to Counter Drug Resistance. *Trends Microbiol.* 24:547–557.
51. Lockbaum, G. J., F. Leidner, ..., C. A. Schiffer. 2019. Structural Adaptation of Darunavir Analogues against Primary Mutations in HIV-1 Protease. *ACS Infect. Dis.* 5:316–325.
52. Wang, W., H. Zhao, and G. Z. Han. 2020. Host-Virus Arms Races Drive Elevated Adaptive Evolution in Viral Receptors. *Parrish CR. J. Virol.* 94:e00684–20.
53. Enard, D., L. Cai, ..., D. A. Petrov. 2016. Viruses are a dominant driver of protein adaptation in mammals. *Elife.* 5, e12469.
54. Suzuki, Y. 2010. Statistical methods for detecting natural selection from genomic data. *Genes Genet. Syst.* 85:359–376.
55. Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47:97–120.