



OPEN

FMO-guided design of darunavir analogs as HIV-1 protease inhibitors

Hathaichanok Chuntakaruk^{1,2}, Kowit Hengphasatporn³✉, Yasuteru Shigeta³, Chanat Aonbangkhen⁴, Vannajan Sanghiran Lee⁵, Tanatorn Khotavivattana⁴✉, Thanyada Rungrotmongkol^{1,2}✉ & Supot Hannongbua⁶✉

The prevalence of HIV-1 infection continues to pose a significant global public health issue, highlighting the need for antiretroviral drugs that target viral proteins to reduce viral replication. One such target is HIV-1 protease (PR), responsible for cleaving viral polyproteins, leading to the maturation of viral proteins. While darunavir (DRV) is a potent HIV-1 PR inhibitor, drug resistance can arise due to mutations in HIV-1 PR. To address this issue, we developed a novel approach using the fragment molecular orbital (FMO) method and structure-based drug design to create DRV analogs. Using combinatorial programming, we generated novel analogs freely accessible via an on-the-cloud mode implemented in Google Colab, Combined Analog generator Tool (CAT). The designed analogs underwent cascade screening through molecular docking with HIV-1 PR wild-type and major mutations at the active site. Molecular dynamics (MD) simulations confirmed the assess ligand binding and susceptibility of screened designed analogs. Our findings indicate that the three designed analogs guided by FMO, 19–0–14–3, 19–8–10–0, and 19–8–14–3, are superior to DRV and have the potential to serve as efficient PR inhibitors. These findings demonstrate the effectiveness of our approach and its potential to be used in further studies for developing new antiretroviral drugs.

Keywords HIV-1 protease, Darunavir analogs, Structure-based drug design, Fragment molecular orbital (FMO), Combined analog generator tool

The human immunodeficiency virus type 1 (HIV-1) is a retrovirus that progressively destroys essential white blood cells of the immune system called CD4+ helper T cells. According to the World Health Organization (WHO), the number of people living with HIV has risen to nearly 38 million worldwide by 2022¹. Even though most infected people have no symptoms, they need to receive HIV treatment and antiretroviral therapy (ART). The guideline for ART emphasizes interrupting different vital stages in the viral life cycle as an essential key for HIV drug development, including non-nucleoside reverse transcriptase inhibitors (NNRTIs), nucleoside reverse transcriptase inhibitors (NRTIs), integrase strand transfer inhibitors (INSTIs), and protease inhibitors (PIs)². HIV-1 protease (HIV-1 PR) is a homodimeric aspartic PR consisting of ninety-nine amino acids per monomer with a catalytic D25, and two flexible β -hairpin flaps that cover the active site¹ (Fig. S1 in Supporting Information). The enzyme cleaves viral polyproteins at specific sites, contributing to the mature structure and function of HIV-1 proteins^{3,4}. Blocking the PR active site by PIs is crucial to prevent the mature virion assembly, viral replication, and increased viral load. Today, ten PIs have been approved by the Food and Drug Administration (FDA)^{5–7}. One of the most potential PIs currently recommended as second or third-line antiretroviral therapy is darunavir (DRV)¹. It was developed from amprenavir and fosamprenavir by modifying 3-methyl tetrahydrofuran, since several studies suggest the prevalence of HIV-1 drug resistance on PIs⁸. Moreover, the increased dissociative half-life of DRV results in prolonged release of the drug in the body, indicated by its binding affinity

¹Program in Bioinformatics and Computational Biology, Graduate School, Chulalongkorn University, Bangkok 10330, Thailand. ²Center of Excellence in Structural and Computational Biology, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ³Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan. ⁴Center of Excellence in Natural Products Chemistry, Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ⁵Chemistry Department, Faculty of Science, University Malaya, Kuala Lumpur 50603, Malaysia. ⁶Center of Excellence in Computational Chemistry (CECC), Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ✉email: kowith@ccs.tsukuba.ac.jp; tanatorn.k@chula.ac.th; thanyada.r@chula.ac.th; supot.h@chula.ac.th

higher than 100-fold compared to other PIs⁹. Although DRV is highly effective in most patients, the current ART is still inadequate as HIV-1 can develop drug resistance through an accumulation of PR mutations, including V82A, I84V, and L90M^{10–12}. Major PR mutations including D30N, V32I, M46L, G48V, I50V, I54M/V, L76V, and N88S were also found in the Weber et al.¹³ and, Stanford University¹⁴ study according to their databases on HIV treatment resistance. Furthermore, many studies have reported that several HIV patients show DRV drug resistance against PRs¹⁵.

Quantum mechanical (QM) calculations have made structure-based drug design more effective by accurately predicting molecular and electronic properties, including effects missing in molecular mechanics (MM), such as electronic polarization, charge transfer, halogen bonding, and covalent bond construction, which are superior to MM force fields. The fragment molecular orbital (FMO) method offers faster computational speed than conventional QM calculations. It is suitable for calculating macromolecular systems by partitioning them into several fragments and calculating them in parallel^{16,17}. The pair interaction energy (PIE) and decomposed analysis of PIE (PIEDA) provide the contributed interaction energy and key chemical groups of known drugs that could be used to search for novel potent inhibitors or be a clue for drug design^{18–21}.

The need for the commercially licensed software and standalone personal computers for substructure combination in structure-based drug design can pose significant challenges for researchers in terms of cost and accessibility^{22–29}. This can limit the scope of research and hinder progress in developing effective antiretroviral drugs for diseases like HIV-1. Our study aims to develop a new approach that uses the FMO calculation to introduce the chemical substructure for cascade screening. The programming for combinatorial chemistry in drug design, Combined Analog generator Tool (CAT), was developed to build novel drug analogs with all possible combinations of substructures without requiring a license or expensive software. Initially, the FMO method was employed to investigate the ligand–protein interactions of the DRV/HIV-1 PR crystal structure complex. Based on the FMO results, substructure modification was carried out in four fragments (F1, F1', F2, and F2'), as shown in Fig. 1A. These chemical substructures were considered as single-position modified DRV analogs and then screened using molecular docking. The potent chemical substructures were combined using CAT to create four-position combined DRV analogs. These were evaluated for ADME (absorption, distribution, metabolism, and excretion) prediction and beyond rule of five (bRo5) properties to identify potential drug-like

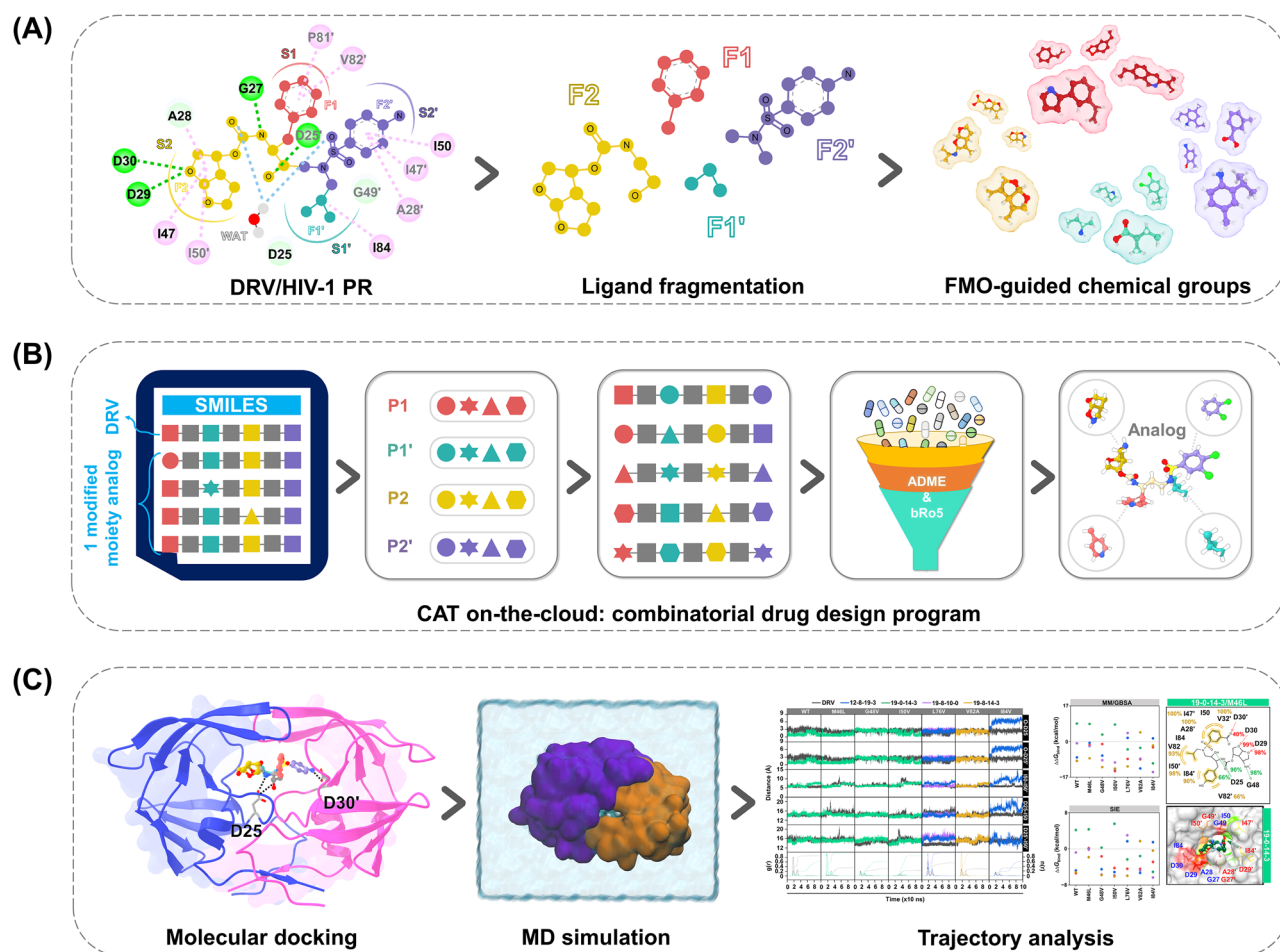


Figure 1. Workflow of the present study. **(A)** FMO-guided drug design in four fragments F1, F1', F2, and F2' of darunavir (DRV). **(B)** Combined Analog generator Tool (CAT)-generated analogs with favorable drug-like properties. **(C)** Molecular docking and molecular dynamics (MD) simulations of designed analogs.

analogues (Fig. 1B). After that, a set of four-position combined DRV analogues were docked into wild-type (WT) and major mutants (D30N, V32I, M46L, G48V, I50V, I54M, I54V, L76V, V82A, I84V, N88S, and L90M) HIV-1 PR, and the effective hits in most PRs were further analyzed using molecular dynamics (MD) simulations. The designed compounds' structural dynamics, ligand–protein interactions, and predicted binding free energies were compared to those of DRV (Fig. 1C). The potential candidate analogues will be synthesized and tested for their biological activity in future studies.

Materials and methods

FMO calculation on DRV/HIV-1 PR complex

To analyze the binding energy of DRV/HIV-1 PR complex in terms of QM³⁰, FMO calculation was performed using the DRV/HIV-1 PR crystal structure (PDB ID: 4LL3³¹), with a focus on amino acid residues within 7 Å of DRV. The General Atomic and Molecular Electronic Structure System (GAMESS)³² software was utilized for this purpose. In FMO calculation, the protein was typically fragmented into monomers by bond detachment atom (BDA) techniques³³, which involved fragmenting the protein at the bond between Ca–C of the protein backbone. Based on this approach, the DRV molecular structure was divided into four fragments: F1, F1', F2, and F2' (Fig. 3A). PIEDA was computed using the second-order Møller–Plesset perturbation theory (MP2) with the resolution of identity (RI) approximation and polarizable continuum model (PCM) solvation effect (FMO-RIMP2/PCM) at the B3LYP/6-31G* basis set³⁴. The crucial interacting residue for DRV binding was identified by computing the pair interaction of monomers (I and J) using the following equation.

$$PIE = \Delta E_{IJ}^{ES} + \Delta E_{IJ}^{CT+mix} + \Delta E_{IJ}^{DI} + \Delta E_{IJ}^{EX} + \Delta G_{Sol}^{PCM}$$

The ligand binding interaction is denoted as the PIE term, and it comprises of the electrostatic interaction (ΔE_{IJ}^{ES}), the charge transfer with higher-order mixed terms energies (ΔE_{IJ}^{CT+mix}), dispersion (ΔE_{IJ}^{DI}), exchange-repulsion (ΔE_{IJ}^{EX}), and the polarizable continuum model (PCM) solvation effect (ΔG_{Sol}^{PCM})^{35–37}.

Molecular docking study of designed analogs

In this study, cascade screening using molecular docking were performed with two sets of ligands: (i) single-position modified DRV analogs and (ii) four-position combined DRV analogs. In the first screening, a set of 20 chemical groups was designed for each substructure, namely P1, P1', P2, and P2', corresponding to the fragments (F1, F1', F2, and F2') computed through FMO calculations. A total of 80 single-position modified DRV analogs were created for the first molecular docking with WT HIV-1 PR. The crystal structure of WT HIV-1 PR in complex with DRV (PDB ID: 4LL3³¹) was prepared as the protein receptor for the molecular docking study. The DRV ligand was removed, and hydrogen atoms were added. The protonation state was assigned using PDB2PQR web server³⁸, except for D25 and D25', which were assigned according to the previous studies³⁹. Molecular docking study was conducted using the GOLD program with genetic algorithm (GA)⁴⁰. The docking site was defined as a 10 Å sphere centered at DRV, located in the active site of HIV-1 PR. To validate the system, redocking of DRV was performed and compared to the crystallized ligand, resulting in a root mean square deviation (RMSD) of 0.65 Å (as depicted in Fig. S3 in the Supporting Information).

After the first round of molecular docking, the single-position modified DRV analogs with higher binding affinity scores compared to DRV were chosen to construct the four-position combined DRV analogs (see detail in section "Combined analog generator tool" Combined analog generator tool of "Materials and methods"). These analogs were then subjected to the second-round molecular docking that focused on both WT and mutant systems. A total of 12 major HIV-1 PR mutations, including D30N (PDB ID: 7DOZ⁴¹), V32I (PDB ID: 2HS1⁴²), M46L (PDB ID: 2HS2⁴²), G48V (PDB ID: 3CYW⁴³), I50V (PDB ID: 6DH6¹⁰), I54M (PDB ID: 3D1Z⁴³), I54V (PDB ID: 3D20⁴³), L76V (PDB ID: 3PWM⁴⁴), V82A (PDB ID: 2IDW⁴⁵), I84V (PDB ID: 2IEO⁴⁵), N88S (PDB ID: 3LZU⁴⁶), and L90M (PDB ID: 6OOS⁴⁶), were prepared as the receptor according to the previously described method. The interactions between the ligand and protein were visualized using the UCSF Chimera package⁴⁷ and Accelrys Discovery Studio 3.0 (Accelrys Inc.)⁴⁸.

Combined analog generator tool

The four-position combined DRV analogs were generated using a Python-based combinatorial algorithm called the CAT⁴⁹, implemented in Google Colab⁵⁰. The process involved three steps: (i) installing dependencies and preparing DRV, (ii) preparing single-position modified DRV analogs, and (iii) generating combined analogs and performing post-analysis (Fig. 2). Initially, DRV and potent single-position modified DRV analogs were identified for core structures and substructures (P1, P1', P2, and P2') using Open Babel version 3.1.1.16⁵¹. The list of selected potent single-position modified DRV analogs for each position (P1, P1', P2, and P2') was uploaded to Google Colab. These analogs were then combined in all possible combinations based on the core structure of DRV and converted to 3D structures using the RDKit package⁵², employing the ETKDG method⁵³. Ligand conformers were created, aligned, and minimized using the Universal Force Field (UFF)⁵⁴. Additionally, the physicochemical features, including ADME and bRo5 properties of the analogs, were also predicted by the RDKit package⁵².

Molecular dynamics (MD) simulations

To prepare the system for all-atom MD simulations, the AMBER ff19SB force field was applied to the protein, while the ligands for this step were optimized by the HF/6–31d level of theory using the Gaussian 09 program^{55–57} and applied for the General AMBER Force Field 2 (GAFF2)⁵⁸, as the standard protocol^{56,57,59}. To add any missing hydrogen atoms, the tLEaP module was employed, followed by energy minimization using steepest descent (SD) and conjugated gradient (CG) methods with 1,500 steps to eliminate any improper contacts. Counterions

(i) Install dependencies and DRV preparation

DRV in SMILES format was classified to substructures and core structures.

```
CC(C)CN(CC(C(CC1=CC=CC=C1)NC(=O)OC2COC3C2CCO3)O)S(=O)(=O)C4=CC=C(C=C4)N
```

```
CC(C)Cucc(C(CC1=CC=CC=C1)NC(=O)C2COC3C2CCO3)OS(=O)=OC4=CC=C(C=C4)N
```

(ii) Input analogs: sdf or SMILES format

Substructures specified by user in sdf or SMILES format were collected in the list.

Input substructures

Input: sdf or SMILES formats

>> Check example files in the local system in Google Colab. <<

> Select a molecule file

Upload_sdf_file:

Upload_SMILES_file:

> Enter a list of SMILES

Enter_SMILES:

Select: moiety you want to modify

P1_moiety:

P1p_moiety:

P2_moiety:

P2p_moiety:

Show code

Modified analog(s) at the P1 moiety

Please enter the number of compound(s):

(iii) Generate combined analogs and post-analysis

All analogs were predicted for ADME and bRo5 properties.

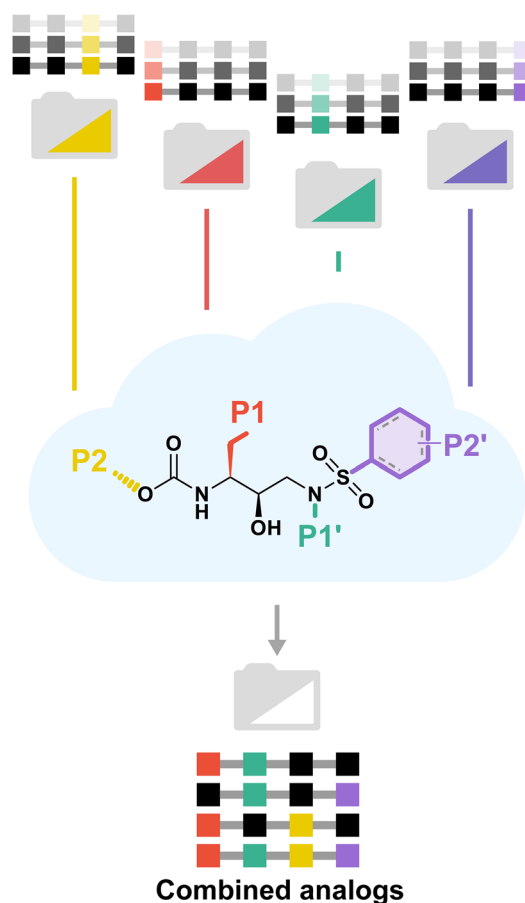


Figure 2. Workflow of Combined Analog generator Tool (CAT) (Fig. S2 in Supporting Information).

were then introduced to neutralize the system, and the TIP3P water model was used to immerse the system in an octahedral box that extended at least 10 Å from the protein surface⁶⁰. Energy minimization was performed on the explicit water and neutralized ion molecules using SD and CG methods with the same iterations. During the MD simulations, nonbonded interactions were cutoff at 10 Å for short-range interactions, and the Particle Mesh Ewald (PME) summation approach was used for long-range electrostatic interactions^{61,62}. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm, and temperature and pressure were controlled. The simulated models were heated from 10 to 310 K over 100 ps and held at this temperature for 100 ns under periodic boundary conditions with an isobaric-isothermal (NPT) ensemble, implemented in the AMBER 20 package program⁵⁸. The last 20 ns-MD trajectories were used to analyze in terms of the structural dynamics of the complex using the CPPTRAJ module⁶³. The binding affinity between the ligand and protein was then calculated using two methods: MM/GBSA^{64,65} and SIE⁶⁶.

Results and discussion

FMO-guided design of drug analogs

The structural-based drug design of DRV analogs was guided by an insight into DRV/HIV-1 PR interactions (Fig. 3A) resulting from the FMO-RIMP2/PCM calculation. The DRV molecular structure was divided into four fragments, F1, F1', F2, and F2', by covalent bond fragmentation at sp³ carbon atoms, and named according to their interacting PR subsites: S1, S1', S2, and S2'. The drug/protein interactions were described in terms of PIE and PIEDA in Fig. 3 and the supporting Table S1. FMO-RIMP2/PCM detected 42 residues and the bridging water involved in the DRV/HIV-1 PR complex (Fig. 3B). The result of the PIE suggested that the primary interaction of DRV resulted mainly from the fragments F1, F2, and F2', while the smaller fragment F1' showed a lower contribution to the overall binding energy due to the size of fragment. Some residues provided stabilization of more than one fragment, such as catalytic D25 (−19.21 kcal/mol for F1', F2, and F2') and I84 (−2.28 kcal/mol for F1' and F2), which are an important residues³⁹. Previous studies have reported that the V82T/I84V mutation prevents the closed conformation of flaps and causes DRV resistance^{10,12,67,68} (Table S1 in Supporting Information).

The PIEDA for all four fragments was plotted separately to reveal the energy components of interaction (Fig. 3C) and used to guide for the chemical substructures in each moiety (Fig. 4). Hydrogen bond (H-bond) interaction and salt bridges refer to the electrostatic (ΔE_{IJ}^{ES}) and charge transfer with higher-order mixed terms energies (ΔE_{IJ}^{CT+mix}), while hydrophobic interaction can be suggested by dispersion (ΔE_{IJ}^{DI}). The steric repulsion

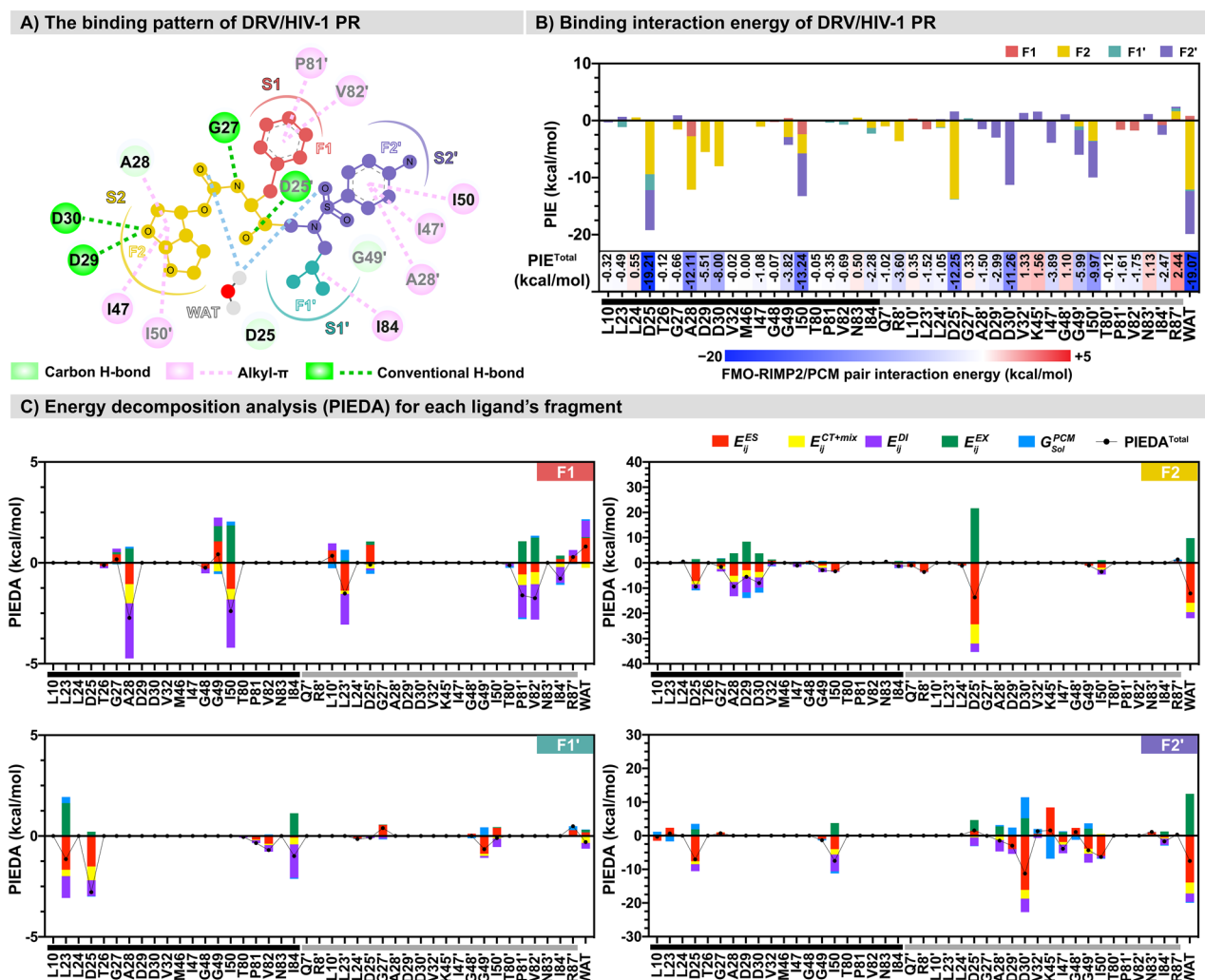


Figure 3. (A) The 2D intermolecular interactions of DRV binding to the active site of HIV-1 PR were obtained from FMO-RIMP2/PCM calculations. (B) The pair interaction energy (PIE) of the four fragments of DRV interacting with four subsites, S1, S1', S2, and S2', of HIV-1 PR was plotted as a stacked bar. (C) The decomposition free energy analysis (PIEDA) of each fragment, F1, F1', F2, and F2', was independently represented as a bar graph. Residues on chains A and B were shown in black and gray, respectively.

between atoms was described by the exchange-repulsion energy (ΔE_{IJ}^{EX})^{21,69}. Our selection of substructures used at various substitution sites (P1, P2, P1', P2') was elucidated based on guidelines from previous studies. The aim in designing each substitution was not to replicate similar substructures from previous studies; instead, we endeavored to modify their substructures to achieve better interactions with HIV-1 PR. Considering that F1 interacts with the S1 subsite, a relatively good PIEDA showed that A28, I50, L23', P81', and V82' mainly interacted with F1 through hydrophobic interaction (ΔE_{IJ}^{DI}). However, the high ΔE_{IJ}^{EX} found in catalytic I50 could affect the system. To increase the diverse interactions of DRV's P1, Gordon's research indicated that the DRV analog (GS-8374), which incorporates a phosphonate group with an aromatic core, enhances van der Waals (vdW) interactions with PR⁷⁰. In our study, we modified substructures by adding hydroxyl (substructures 10, 18, and 19), carboxylate (substructures 1, 7, 8, 17 and 20), and amino groups (substructures 2, 13, 14, 15 and 16) to the aromatic core to achieve more robust hydrogen bonding with the hydrophobic residue. Additionally, the replacement of the P1-phenyl group with two fluorine atoms to create bis-fluoro-benzyl has been shown to introduce halogen bonds with PR and increase lipophilicity against PR mutant⁷¹. Therefore, the halogen group in the substructure was changed to chlorine (substructure 11) to introduce halogen bonds more effectively. The ΔE_{IJ}^{DI} -associated hydrophobic forces were essential for the interaction of the F1' sec-butyl moiety of DRV with L23, D25, and I84. Most of the substructures in the P1' moiety were modified to larger hydrophobic groups (such as a hydrocarbon chain) in comparison to DRV. This modification targeted both WT PR and variants with main mutations^{10,68,72}, aiming to enhance potency in cellular assays compared to DRV^{10,70,72}. Hence, a larger hydrocarbon (substructures 4, 6, 9, 11, and 13), along with the incorporation of halogen (substructures 2, 5, and 14), carboxylate (substructures 13 and 18) and amino groups (substructures 16 and 17) was also considered in designing the substructures in this moiety to raise electrostatic attraction. Moreover, the substitution with aromatic rings, such as substructures

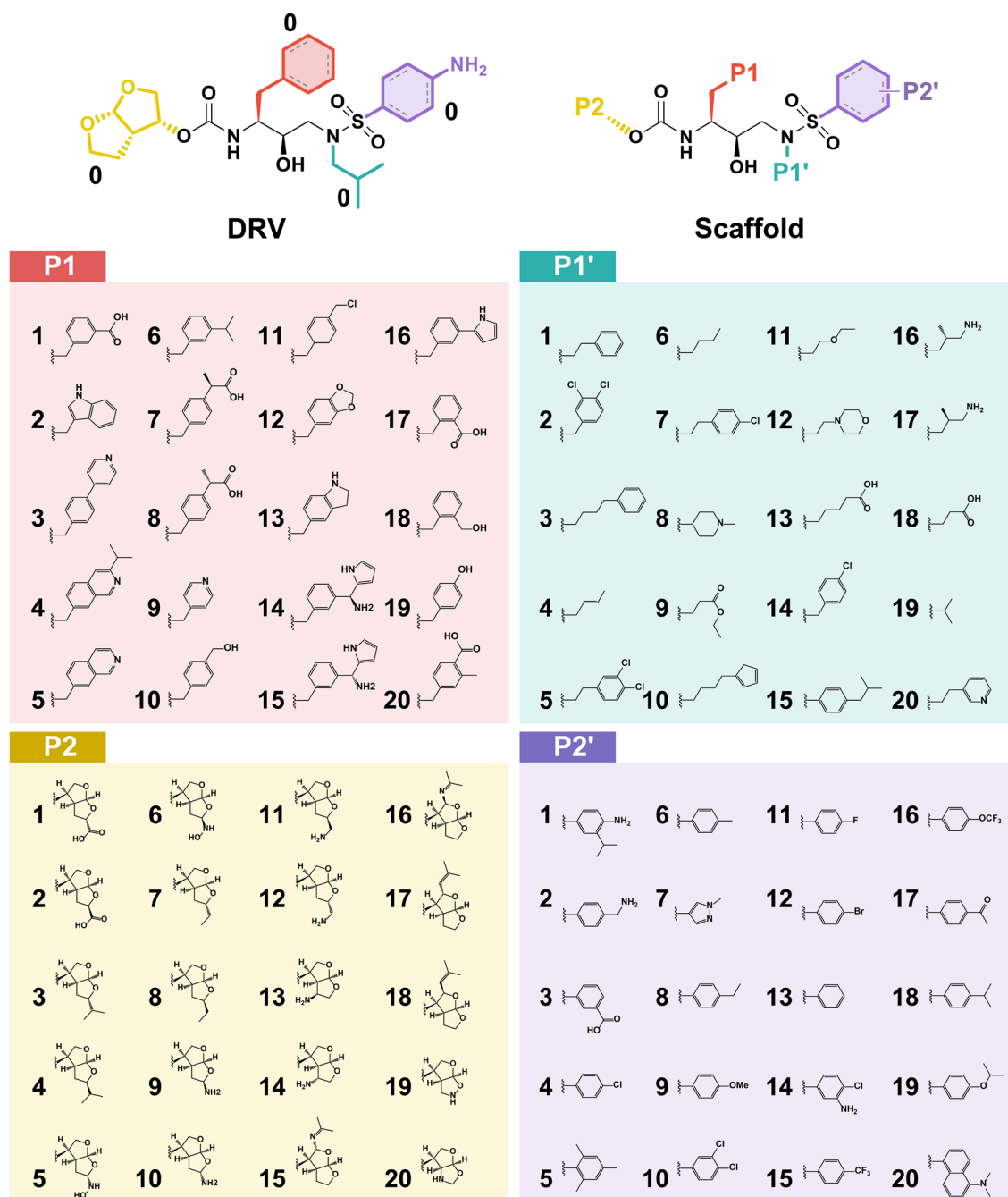


Figure 4. 2D structures of DRV and the scaffold of DRV analogs with different substitutions on the P1, P1', P2, and P2' positions, where the chemical structures of modified functional groups (1–20) at each position are shown below.

1, 5, 8, and 20 at P1', had not been previously investigated. Intriguingly, in our study, both an aromatic ring and a hydrocarbon chain were designed for this moiety to increase hydrophobic interactions.

In addition to the H-bond formation with D25' in the core structure of DRV's F2, the *bis*-tetrahydrofuran (THF) urethane was likely stabilized by A28, D29, D30, I47, and I50'. Designing additional groups, such as *tris*-THF urethane, methoxy, and *gem*-difluoro at P2 moiety to introduce H-bonds with the conserved main chain atoms of the PR (D29, D30, G27, and G48), which cannot be easily altered by mutations, was supported by previous findings^{73,74}. Consequently, our designed substructures were modified by adding hydroxyl (substructures 5 and 6), carboxylate (substructures 1 and 2), and amino groups (substructures 9 to 14) to the *bis*-THF moiety to increase interactions with additional residues. The *bis*-THF urethane was retained for P2 substructures, while the hydroxyl, carboxylate, and amino groups were introduced to strengthen H-bonds with D25, D29, D30, G48, D25', G27', and G49'. The PIEDA of F2' revealed that CH3- π and electrostatic interactions were important for the 4-methylaniline moiety. Antiviral DRV analogs (GRL-04410 and GRL-0519) have a methoxy substitution

at P2', as previously described by Weber and co-workers⁷³. Modifications with methyl, halogen and isopropyl groups at this position improved the antiviral activity of the analogs^{74,75}. In our study, to expand hydrogen bonding, hydrophobicity, and electrostatic attraction, we added halogen (such as meta and para-dichloro, para-fluoro in substructures 10, 11, and 12), isopropyl (substructure 18), isopropyl ether (substructure 19), and amino (substructures 1, 2, and 14) groups to the original DRV aromatic ring. The DRV analogs from an earlier study revealed that the modification at P2' using a carboxylate group maintained polar contact and exhibited exceptional potency in an enzymatic assay⁷⁶. Thus, the introduction of carboxylate and amino groups was designed in this moiety to enhance H-bond interactions. Adding a halogen atom or amino group to this fragment could increase the hydrophobic interaction with A28' and I47' and electrostatic attraction with D25, I50, D29', D30', I47', G49', and I50' through ΔE_{ij}^{ES} and ΔE_{ij}^{CT+mix} . The introduction of carboxylate and amino groups enhances the H-bond interaction with residues D29', D30', and G48'.

Cascade screening

In this study, we introduce an approach utilizing the FMO method to guide the design of chemical substructures with the aim of developing analogs of DRV. The cascade screening strategy employed molecular docking to evaluate two sets of DRV analogs: (i) single-position modified analogs and (ii) four-position combined analogs. Specifically, we designed a comprehensive library of 80 single-position modified DRV analogs, with 20 analogs per position suggested by the FMO calculation. Subsequently, we conducted molecular docking simulations to predict the binding energies of these analogs with the WT HIV-1 PR protein, comparing them to the original DRV as a baseline during the initial screening phase. Notably, our results revealed significant improvements in binding energies for several chemical substructures compared to the original DRV compound. A molecular docking study revealed that positions P1, P1', P2, and P2' contained 5, 4, 8, and 3 potent substructures, respectively, which demonstrated favorable binding properties. As a result, these promising substructures were selected for further investigation. To explore their therapeutic potential, we expanded upon these potent chemical substructures by constructing an extensive library of four-position combined DRV analogs, comprising a total of 1,080 analogs (Fig. S5 in Supporting Information). Subsequently, we performed molecular docking experiments to assess the binding energies of these analogs against both the WT HIV-1 PR protein and 12 mutants.

To create the four-position combined DRV analogs, we developed a user-friendly and efficient combinatorial program using a Python script that was implemented in Google Colab, CAT. These substructures were then combined in all possible positions using the combinatorial programming algorithm (Fig. 2). The pharmacological properties were investigated for drug-likeness in terms of molecular weight (MW), the number of H-bond donors (HBD) and acceptors (HBA), polar surface area (PSA), and lipophilicity (LogP) using the RDKit package⁷⁷ implemented in Google Colab. This is one of the features included in our proposed program in this study. The results suggested that all designed analogs and DRV were well within the bRo5 criteria, including (i) $MW \leq 1000$ Da, (ii) $HBD \leq 6$ and $HBA \leq 15$, (iii) $PSA \leq 250$ Å, and (iv) $LogP \leq 10^{78}$ (File2 in Supporting Information).

To investigate the binding mechanism of DRV and generated analogs towards the binding site of WT HIV-1 PR and 12 major mutated PR strains (D30N, V32I, M46L, G48V, I50V, I54M, I54V, L76V, V82A, I84V, N88S, and L90M). These compounds were docked into the binding site of WT HIV-1 PR and 12 mutants using the GOLD program. All analogs in each system were then calculated to determine which analogs provided high fitness scores frequently in overall WT and mutants compared to DRV. Afterwards, the strains with high frequency found in a group with high fitness scores were selected for further investigation. The grid map in Fig. 5A shows five potent DRV analogs, including **10-8-19-3**, **12-8-19-3**, **19-0-14-3**, **19-8-10-0**, and **19-8-14-3** with a higher fitness score (blue) than DRV in WT and most mutations. The fitness scores of these five analogs are shown in Table S2 in the Supporting Information.

By observing the common structural features of P1 and P1' from the five potent analogs, it was found that the original DRV, namely 4-hydroxybenzyl for P1 (**19-x-x-x**) and 1-methylpiperidin-4-yl for P1' (**x-8-x-x**), were the most effective. In addition, some other variations, such as introducing benzyl and methyl groups as 4-(hydroxymethyl)benzyl and benzo[d][1,3]dioxol-5-ylmethyl for P1 (**10-x-x-x**) and (**12-x-x-x**), respectively, or introducing a phenyl group for P1' (**x-0-x-x**), could lead to favorable interactions with HIV-1 PR. Furthermore, for P2, all of the five analogs benefited from introducing the (4R)-amino group as (3S,3aR,4R,6aS)-4-aminohexahydrofuro[2,3-b]furan-3-yl (**x-x-14-x**) and hydroxyl hydrogen group as (3aR,4S,6aS)-hexahydrofuro[3,2-d]isoxazol-4-yl (**x-x-19-x**), suggesting a critical interaction in this position. Lastly, introducing a hydrophobic carboxyl group as 3-carboxyphenyl (**x-x-x-3**) seems to be the most common feature in the five analogs for P2' modification.

Due to the preference for aromatic residues, strong hydrophobic interactions were observed at positions P1 and P1' with I50, I84, and I84' for P1 and L23, V82, and I50' for P1' (Fig. 5B). These positions correspond to the modeled peptide position in positions P1' and P1 of the PR-peptide complex⁷⁹ inserted into S1' and S1 subsites. Previous studies have supported the design of cyclic, acyclic, and aromatic non-polar N-alkoxy moieties at the P1' position^{76,80}, similar to the P1 position suggested by FMO calculation in our study. Novelty, this result suggests that the preferable chemical moiety for S1' and S1 subsites should exhibit hydrophobic properties, while the S2 and S2' subsites (half P2 and P2') can accommodate both polar and hydrophobic effects⁸¹⁻⁸⁵.

To evaluate the binding interactions of the most promising DRV analogs, GOLD docking provided information on several interactions, including H-bonds, electrostatic interactions, and hydrophobic interactions. By comparing them to DRV, we found new interactions among our designed analogs in residues (i) G48 and I84 on both chains, (ii) R8, V32, G49, T80, P81 on chain A, and (iii) L23', D29', K45', I84' on chain B. The results verified that the FMO-guided design could be advantageous for introducing plausible chemical groups to enhance binding interactions²¹. However, the effective five analogs were selected for MD simulations to evaluate the physical movements of atoms and molecules of complexes.

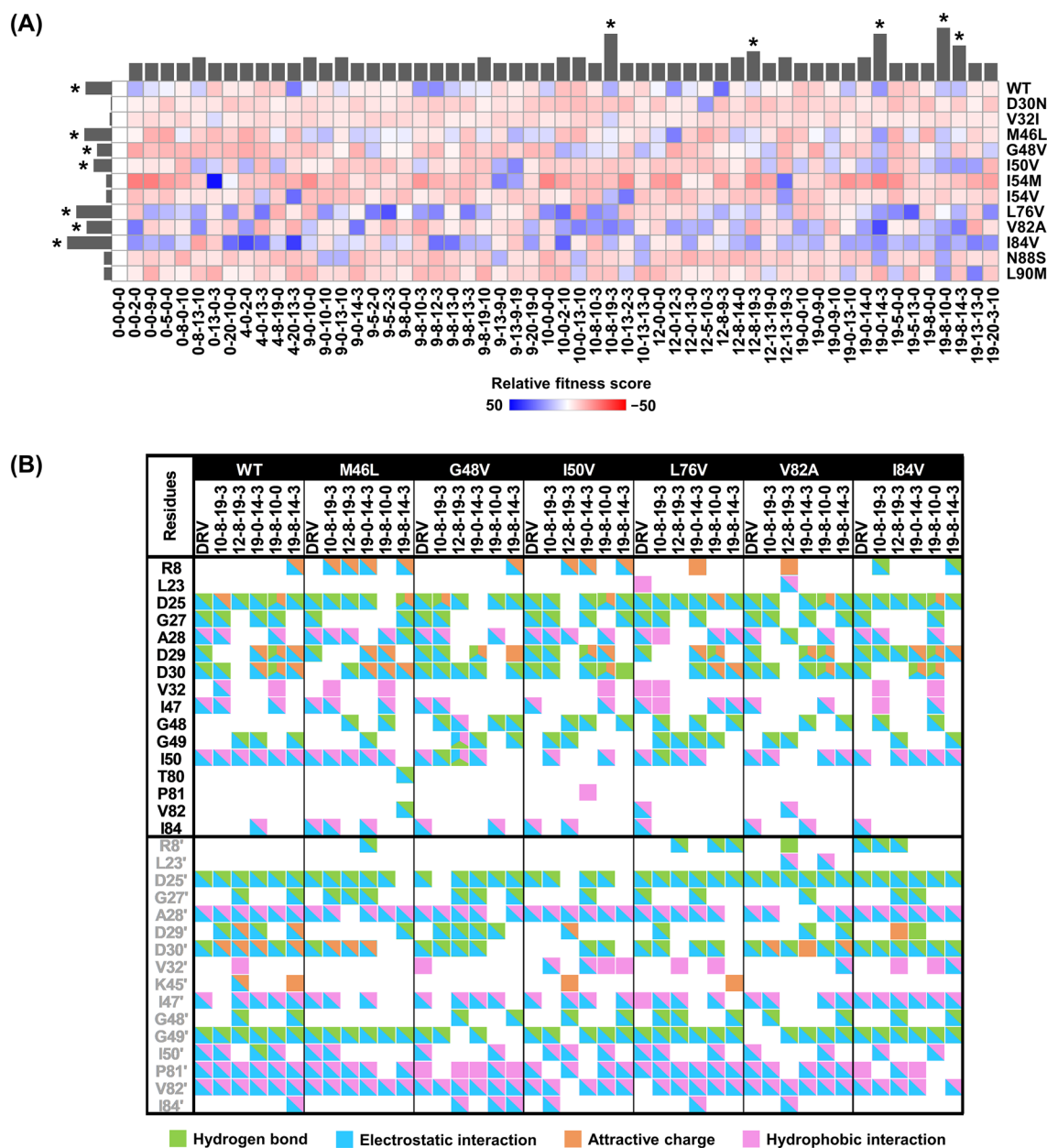


Figure 5. (A) the relative fitness scores of DRV (0–0–0–0) and selected analogs compared to the fitness score of DRV, as demonstrated by the grid map and collective data bar plot in the above and right panels. The potent analogs, ranked by fitness scores and marked with an asterisk (*), are also presented. (B) the interactions between the potent analogs and HIV-1 PR are depicted in the grid map using various colors to indicate different types of contact residues. For example, green, blue, orange, and pink colors represent H-bond, electrostatic interaction, attractive charge, and hydrophobic interaction, respectively.

MD-based screening via binding free energy

To evaluate the efficacy of various analogs in both WT and mutant systems, the dynamic and solvation effects were taken into consideration. To accomplish this, a total of 42 complexes, consisting of six screened compounds and seven proteins, were subjected to MD simulations. The binding affinity of DRV and its analogs to HIV-1 PRs was assessed using the SIE and MM/GBSA methods on 100 snapshots taken from last 20 ns-MD trajectory. The last 20 ns of the simulation was selected for binding affinity calculation because every system reached an equilibrium state during this period, making these trajectories suitable for analysis. The binding energy (ΔG_{bind}) of screened analogs was compared to DRV by considering the change in binding free energy values ($\Delta\Delta G_{bind} = \Delta G_{bind}^{DRV} - \Delta G_{bind}^{analog}$) in WT and mutated systems (Fig. 6A). If $\Delta\Delta G_{bind}$ has a positive value, it indicates that the analog binds to PRs better than the DRV system. Conversely, a negative value of $\Delta\Delta G_{bind}$ suggests that DRV binds to PRs better than the analog system. When $\Delta\Delta G_{bind}$ is zero, it indicates that the analog binds to PRs comparably with the DRV system. Furthermore, as part of the initial screening process to identify potent DRV analogs against both WT and mutated HIV-1 PR, 100-ns MD simulations were conducted in a single run. The

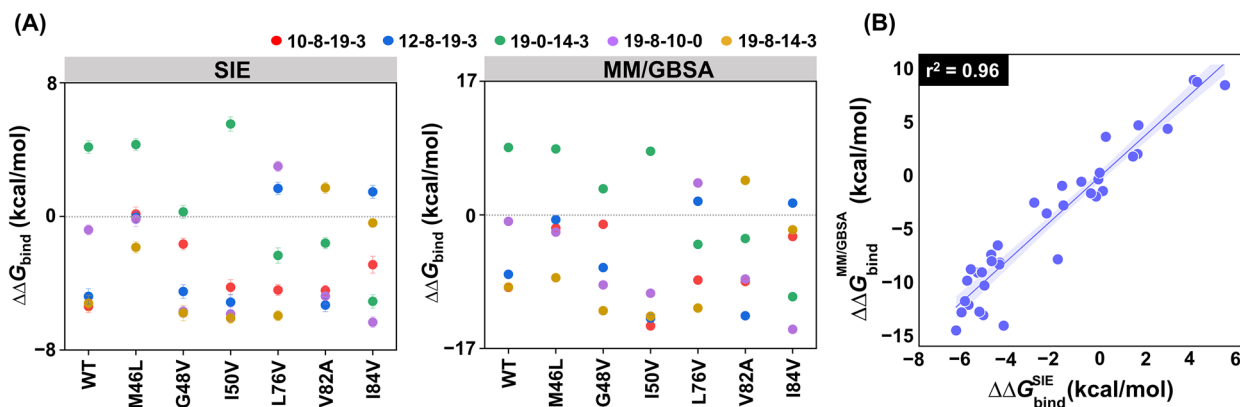


Figure 6. (A) The results of $\Delta\Delta G_{bind}$ (kcal/mol) based on the SIE and MM/GBSA methods are shown as the mean \pm standard error of the mean (SEM) of simulations (B) Pearson correlation coefficient (r^2) analyses between SIE and MM/GBSA are plotted. $\Delta\Delta G_{bind}^{DRV}$ illustrated by the dash line as a standard $\Delta\Delta G_{bind}$.

reference DRV was also subjected to 100-ns MD simulations in triplicates, and the binding affinity was calculated to assess the consistency of the results, as illustrated in Figure S6 and Table S3 in the Supporting Information. The result suggested that four analogs (12-8-19-3, 19-0-14-3, 19-8-10-0, and 19-8-14-3) in seven systems (WT, M46L, G48V, I50V, L76V, V82A, and I84V) could be the most promising compounds for PIs. Notably, the $\Delta\Delta G_{bind}$ values obtained using both methods exhibited a high correlation with $r^2 = 0.96$ based on Pearson correlation⁸⁶ (Fig. 6B).

According to $\Delta\Delta G_{bind}$, the 19-0-14-3 could be the common potent analog that effective against WT and the other three vital mutations: M46L, G48V, and I50V. While 12-8-19-3 could bind to the L76V and I84V systems. The analogues 19-8-10-0 exhibited higher binding affinity and inhibitory efficiency than DRV against only the L76V mutation, while 19-8-14-3 demonstrated good binding free energy to V82A mutations. The relationship between the five potent analogs was associated with combined critical substructures in two patterns: (i) sharing the chemical moiety at P1, P2, and P2', such as 19-0-14-3 and 19-8-14-3, and (ii) sharing the chemical moiety in P1', P2, and P2', such as 10-8-19-3 and 12-8-19-3. Notably, chemical substructure 3 of P2' moiety was shared in both analogs, thus indicating that this optimal chemical group may be the most important for developing DRV analogs. Consequently, distance analysis and radial distribution function (RDF) were performed to estimate the potency of designed substructures, including analogs.

Ligand/protein stability evaluation

The ligand-binding stability of DRV and the five screened analogs in complex to WT HIV-1 PR and mutants were investigated from MD trajectory, as shown in Fig. 7. Previous studies have indicated that the OH of DRV forms an H-bond with catalytic residues D25 and D25' of PR, maintaining a similar binding model^{76,87,88}. Thus, the distance between the core structure of PRs (O) and D25 or D25' (OD1, OD2) of DRV and analogs remained stable throughout the simulation, except in the 12-8-19-3/I84V system. The O-D25 distance of PRs and DRV ranged from 2.0 to 2.8 Å, while the distance between PRs and the four analogs in six systems was from 1.9 to 2.9 Å, similar to the distance of O-D25' between PRs and DRV or analogs systems. The results illustrated that the distance between PRs and D25 or D25' of six PRs/analog systems (19-0-14-3/WT, 19-0-14-3/M46L, 19-0-14-3/G48V, 19-0-14-3/I50V, 12-8-19-3/L76V, 19-8-10-0/L76V, and 19-8-14-3/V82A) remained stable during the MD simulation compared to PRs/DRV systems. Therefore, the good interaction of analogs (Fig. 5 and 6) was estimated to be caused by the effectiveness of the designed substructures, not the core structure.

Comprehending the challenges associated with regulating the mobility of the HIV-PR flap can offer valuable insights into developing novel therapeutic inhibitors capable of disrupting flap opening and stabilizing them in the closed conformation. Thus, several prior computational studies have aimed to understand flap opening dynamics⁸⁹⁻⁹². Previously, the interatomic distance between Ile50/50', which is located at the flap tips of each monomer, was used to define the various conformations (closed, semi-open, and fully open conformations). Except for the system of 12-8-19-3/I84V, the distance between the I50-I50' residues of both the PRs/DRV and analog systems remained constant throughout the simulation, at approximately 6.2 Å. This finding is consistent with a previous study, which detected a closed state at approximately 5.8 Å in the presence of the inhibitor^{93,94}. In the absence of the inhibitor, the HIV-PR structure undergoes a conformational change to a more flexible and semi-open state, allowing for proper flap reversal. Typically, PR mutations weaken the hydrophobic interactions related to I50 and I50'¹⁵; however, the results of the I50-I50' distance reveal that potent analogs in the six systems play an essential role in maintaining the closed conformation not only of the WT but also of mutated PRs. Nonetheless, the I50-I50' distance can be affected by both flap asymmetry and curling of the flap tip; therefore, it would be more reasonable to measure the distance between D25-I50 of chain A and D25'-I50' of chain B in PRs^{15,94-96}.

Note that, the D25-I50 and D25'-I50' distances showed a similar pattern to the previous distance measurement. All six DRV/analog systems tested against PRs maintained a closed conformation (~14 Å) for both chains A and B. In contrast, the system of 12-8-19-3/I84V exhibited an open conformation, characterized by higher fluctuations than the other systems (~18 Å). The ability of this system to stabilize a more open conformation

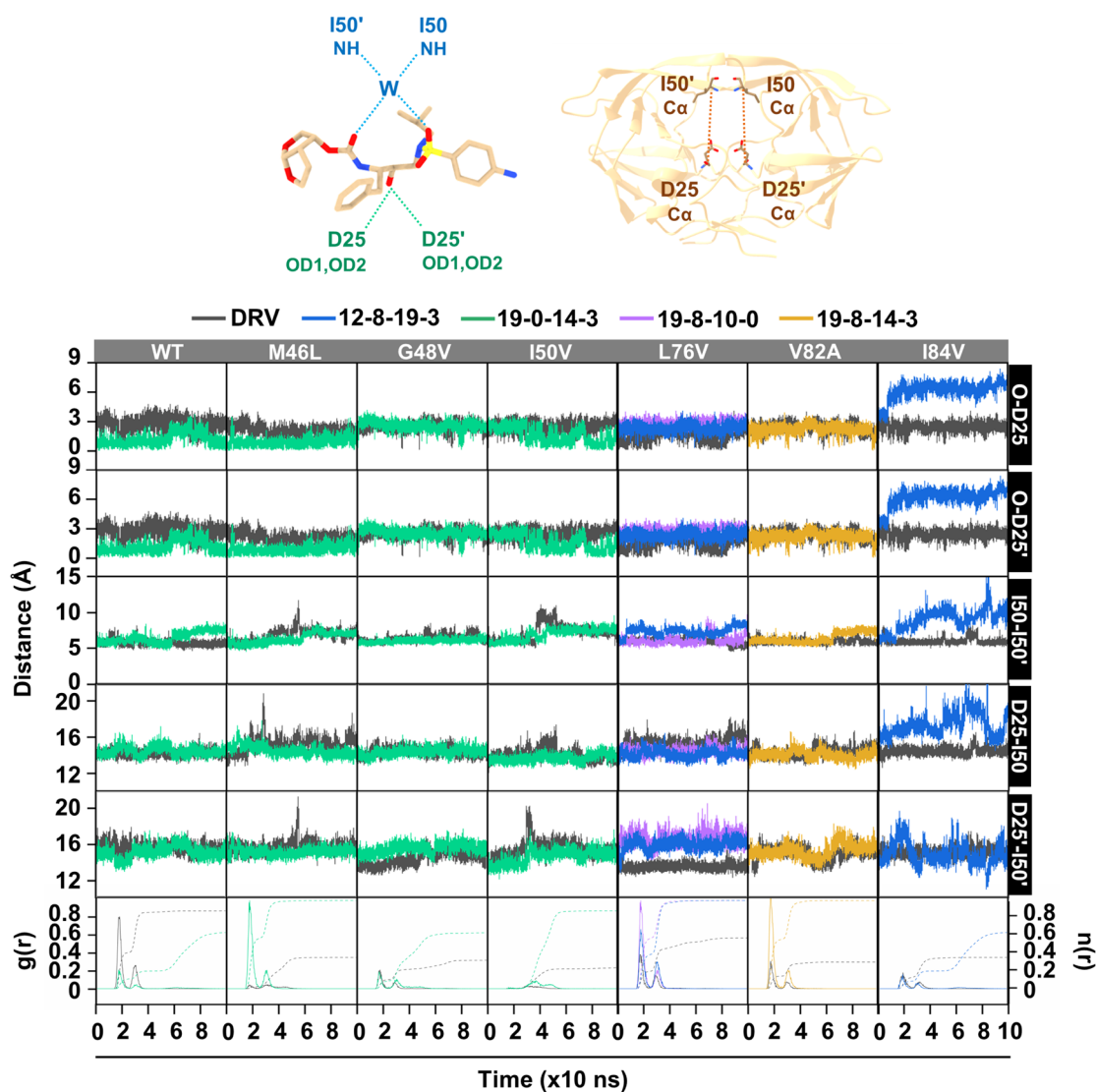


Figure 7. The three-dimensional structures of DRV and PR are represented in brown sticks, with dotted lines indicating distance. Water molecules are represented by W, while NH refers to the main chain amide group, C α to the main chain carbon alpha, and OD1 and OD2 to the main chain oxygen group. The time series of O-D25, O-D25', I50-I50', D25-I50, and D25'-I50' distances of the DRV/PRs and analogs/PRs systems are shown. The radial distribution function (RDF) of water oxygens around the NH nitrogen of DRV and analogs is calculated from the last 20-ns simulations of PRs solvated by TIP3P waters.

would cause an increase in the volume of the active site cavity, which could have a vital impact on the binding of less favorable PIs^{15,94}. Interestingly, water-mediated H-bonds were observed, forming attractive networks connecting a sulfonyl oxygen and hydroxyl group of DRV or analogs with the amino group on PRs^{76,88}. The RDF results showed that around 0.6 to 0.9 water molecules were stabilized by the NH of analogs/PRs systems, whereas 0.3 to 0.8 water molecules of DRV/PRs systems were found. Therefore, all systems had water-mediated interactions with PR in the WT and mutated strains. Based on the distance analysis and RDF, the stable systems, except 12-8-19-3/I84V, were further analyzed in depth for the interactions of overall compounds and each moiety.

Binding pattern and interaction of DRV and the designed analogs

Several key interactions visualized by LigandScout 4.4.8⁹⁷, consist of steric and electronic features shared by numerous active substances with similar biological targets. These features include positive and negatively charged groups, hydrophobic and aromatic areas, as well as HBD and HBA capabilities⁹⁸⁻¹⁰⁰. In this study, the MD trajectories of DRV and potent in complex with analogs HIV-1 PR were analyzed for pharmacophore, which is specialized in detecting ligand-protein interactions¹⁰¹. Figure 8 illustrates the 2D representative pharmacophore models (RPMS) of DRV and analogs (12-8-19-3, 19-0-14-3, 19-8-10-0, and 19-8-14-3) extracted from the last 20 ns of the MD trajectories. Hydrogen bonding, HBD, or HBA, as well as hydrophobic interactions, were important chemical pharmacophore features for DRV and all analogs binding to HIV-1 PR through the core and substructures. The formation of H-bonds was investigated by defining criteria: the distance between the HBA

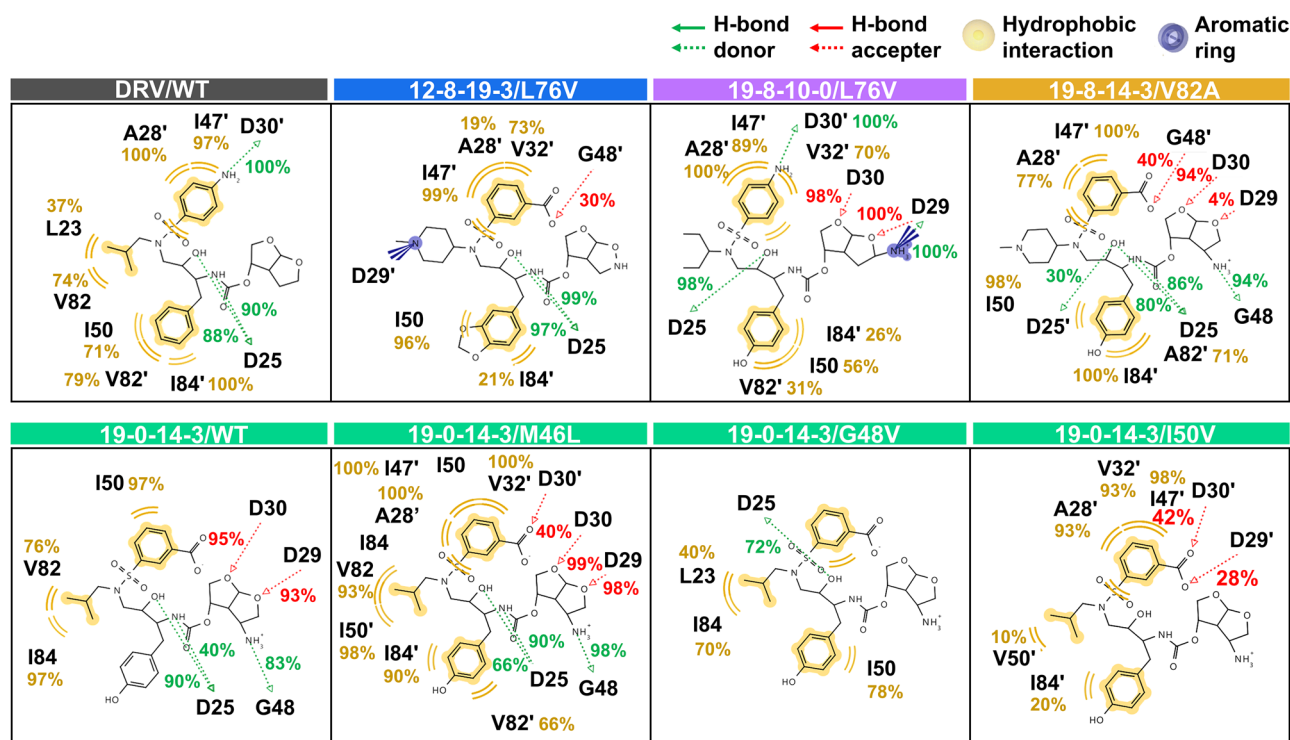


Figure 8. The vital interactions of DRV and potent analogs in complex with HIV-1 PR shown in the 2D pharmacophore models, which is analyzed from the last 20 ns-MD simulations. The green arrow, red arrow, yellow circle, and purple circle represent the pharmacophore features of hydrogen bond donor (HBD) and acceptor (HBA), hydrophobic interaction properties, and aromatic ring, respectively.

and the HBD was ≤ 3.5 Å, and the angle of HBD–H...HBA was $\geq 120^\circ$. The percentage occupation of H-bonding of the DRV/analog complexes is shown in Fig. 8. In the case of the reference drug, DRV showed hydrophobic interactions with L23, I50, V82, A28', I47', V82', and I84', similar to most analogs. However, there were some additional residues, V32' for 12-8-19-3/L76V and 19-8-10-0/L76V systems, I84 for 19-0-14-3/M46L and 19-0-14-3/G48V systems, and V32', I50' for 19-0-14-3/M46L system. By focusing on H-bond, 19-0-14-3/M46L, 19-8-10-0/L76V, and 19-8-14-3/V82A showed a high occurrence of hydrogen bonding with the active binding residues D29 (99%), D29 (100%), and D30 (94%), respectively, and hydrophobic interaction with A28' (77–100%), V32' (70–100%), and I47' (89–100%) in chain B. Moreover, an H-bond with G48 had been identified, while the 19-0-14-3/M46L and 19-8-10-0/L76V systems had a H-bond with D30'. Only the 19-8-14-3/V82A complex was able to form a H-bond with G48' (40%) operating as the active binding residue. Based on pharmacophore analysis, 19-0-14-3/M46L, 19-8-10-0/L76V, and 19-8-14-3/V82A exhibit the highest number and percentage of pharmacophores. Therefore, these systems were selected for further exploration of the key binding residues compared to DRV in complex with M46L, L76V, and V82A.

Key binding residue evaluation

Figure 9 illustrates critical binding residue evaluation. The per-residue decomposition free energy ($\Delta G_{bind}^{residue}$) calculation for 19-0-14-3/M46L, 19-8-10-0/L76V, and 19-8-14-3/V82A compared to DRV in complex with M46L, L76V, and V82A by considering each modified position indicated that the aromatic rings at the P1 position of DRV and analogs were inserted into the S1 pocket of HIV-1 PR. These residues, such as I50 and I84 formed hydrophobic contacts and electrostatic interactions with DRV and three potent analogs, while G27 and G49 residues interacted with these analogs through H-bond. Only G27 was found to interact with DRV. These residues are consistent with previous studies that have identified significant binding residues such as D25, G27, A28, D29, and G49^{76,102,103}.

For the P1' position, R8 residue bound to the amino groups of 19-0-14-3 analog with attractive charge and electrostatic interaction in the docking process, while 19-8-14-3 analog represented vdW interaction with residues R8 and L23 in MD simulation. Both DRV and 19-8-10-0 interacted with I50' in the L76V system via hydrophobic and electrostatic interactions. However, $\Delta G_{bind}^{residue}$ values showed that 19-8-10-0 had a more effective interaction than DRV. The heterocyclic groups at the P2 position of DRV were inserted into the S2 pocket to form both vdW and electrostatic interactions with several residues. Interestingly, three potent analogs interacted with D29 and D30 through attractive charge, H-bond, and electrostatic interactions, resulting in more potent inhibitory effects against PRs than DRV. However, their contacts with residue A28 showed only vdW and electrostatic interactions.

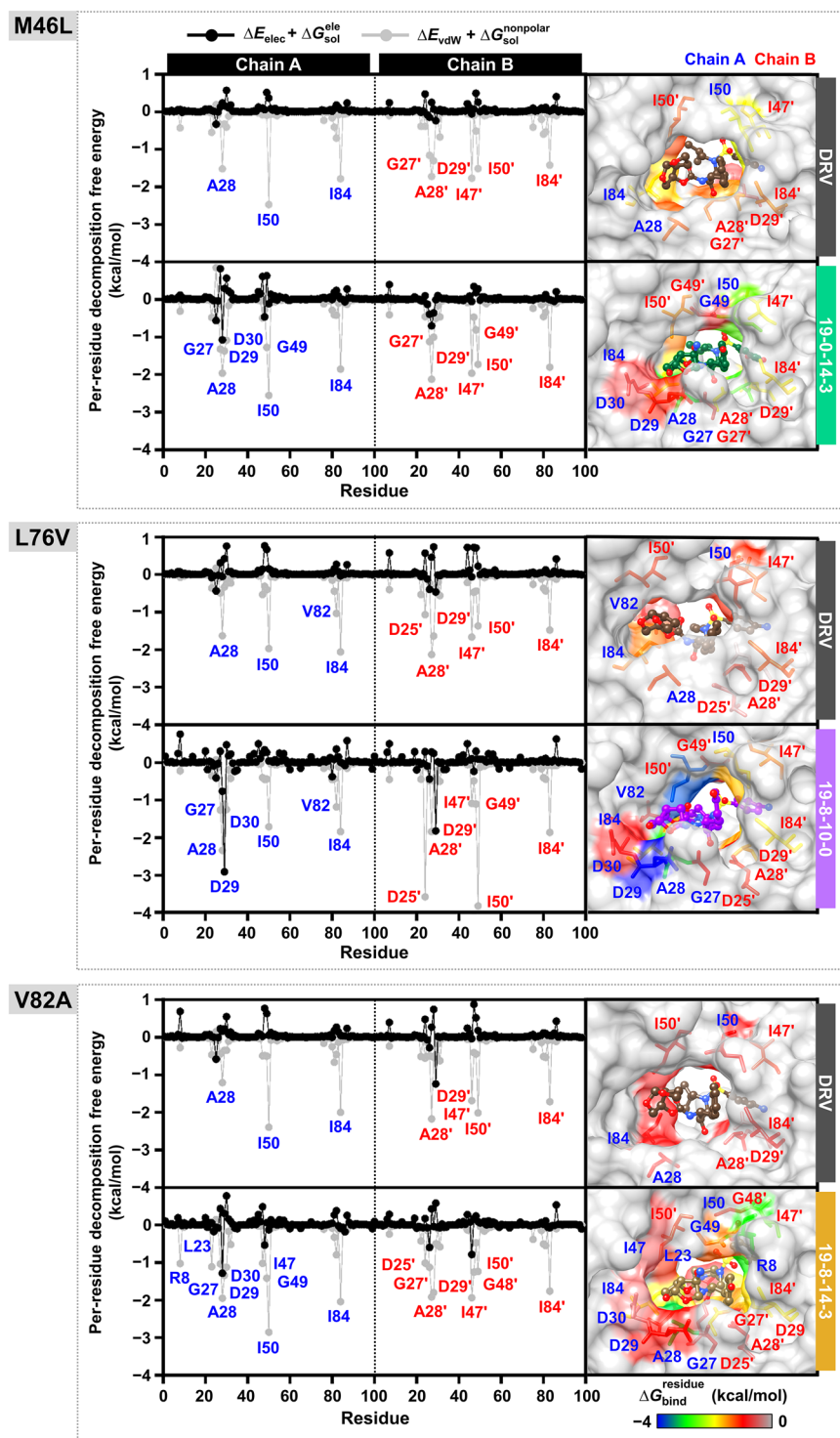


Figure 9. The values of electrostatic energy ($\Delta E_{elec} + \Delta G_{sol}^{ele}$, shown in black) and vdW ($\Delta E_{vdW} + \Delta G_{sol}^{nonpolar}$, shown in grey) contributions from each residue of HIV-1 PR to the binding of darunavir (DRV) and its analogs were analyzed. The 3D structures that represent the ligand's binding orientation at the HIV-1 PR were selected from the clustered molecular dynamics (MD) snapshot. The residues that contribute to the ligand binding are colored based on their ΔG_{bind} residue values in chain A (black) and chain B (red). The residues are arranged in order of increasing energy, which is represented by colors ranging from gray to blue, respectively.

The D25' residue interacted with the aliphatic amine of **19-8-10-0** and **19-8-14-3** analogs through vdW, H-bond, and electrostatic interactions. The catalytic residues D25 and D25' are crucial for acid–base catalysis in the aspartate active site, and most current HIV PIs mimic the substrate's transition state by using the amino group to interact with D25 and D25' residues through H-bonds. However, only the heterocyclic amine group of **19-8-14-3** analog showed H-bonds^{81,87,104–107} with both residues G48 and G49'. The aromatic amines and carboxylic acid at the P2' position of DRV and analogs bonded to the S2' pocket, forming hydrophobic contacts and electrostatic interactions with residues A28', D29', and I47'. In the **19-8-14-3/V82A** system, there was also contact with the D29' and G48' residues via a H-bond, while the residue D30' was found in the **19-8-10-0/L76V** system. The *bis*-THF moiety at P2' contributed to high-binding affinity, forming strong H-bond with D29' and D30' residues^{108,109}.

The vdW and electrostatic interactions of several residues, including A28, I50, I84, G49', and I50', were found to significantly contribute to the binding of inhibitors to both WT and mutated HIV-1 PR. Inhibitors that provide stronger interactions with other residues in the active site, such as G27, D25', and G27', via H-bonds^{110–112}, are more potent. I50 and I50' can form H-bonds with the inhibitor by utilizing water molecules in the active site, allowing the PR to assume an active conformation^{104,113,114}. The residues D25 and D25' also interact with the OH atoms of DRV and its analogs' core structure. Moreover, D30' influences the NH and COOH of the P2' moiety part, and R8 interacts with the amino group of the P1' moiety part. Although electrostatic interactions primarily contribute to the complexation of HIV-1 PIs, H-bonds between DRV/analog and their surrounding residues in the active site of HIV-1 PR can also play an important role in enzyme inhibition (Fig. S7 in Supporting Information).

Conclusions

This study demonstrates the applicability of computer-assisted combinatorial chemistry methods for designing and in silico screening of DRV analogs as HIV-1 PR inhibitors. FMO-guided chemical substructure design proposes favorable chemical fragments to facilitate drug optimization, complemented by cascade screening methodologies using molecular docking techniques. Based on SIE and MM/GBSA calculations, MD simulations indicated that among the critical criteria are binding free energy, hotspot residue, and hydrogen bonding interaction. The susceptibility of **19-0-14-3**, **19-8-10-0**, and **19-8-14-3** analogs against PR was higher than DRV in WT and mutants HIV-1 PR, supported by the energy stabilization from the individual residues, suggesting that these designed analogs could be used to combat HIV as a global health issue. Moreover, the details on the crucial interactions for analog binding have also been elucidated, which is beneficial for future design. Additionally, the FMO/combinatorial programming-guided technique presented in this study could become an efficient tool for drug discovery.

Data availability

All of the designed DRV analogs, topology coordinates, and parameter datasets that support the conclusions of this article, as well as the cascaded structure-based drug design (SBDD) pipeline utilized for data analysis, are available in the Supporting Information files and at https://github.com/fahhathai/Tutorial_CombinatorialSubstructures.git.

Code availability

An accompanying tutorial that involves a program for combining substructures is freely available at https://github.com/fahhathai/Tutorial_CombinatorialSubstructures/blob/main/Combinatorial_program.ipynb

Received: 3 June 2023; Accepted: 6 February 2024

Published online: 13 February 2024

References

1. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach. (Geneva, 2016).
2. Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161. <https://doi.org/10.1101/cshperspect.a007161> (2012).
3. Patel, M., Mandava, N. K., Vadlapatla, R. K. & Mitra, A. K. Recent patents and emerging therapeutics for HIV infections: A focus on protease inhibitors. *Pharm. Pat. Anal.* **2**, 513–538. <https://doi.org/10.4155/ppa.13.33> (2013).
4. Rusconi, S. & Viganò, O. New HIV protease inhibitors for drug-resistant viruses. *Therapy* **3**, 79–88 (2006).
5. Virgil, S. C. In *Aspartic Acid Proteases as Therapeutic Targets* 139–168 (2010).
6. Ali, A. *et al.* Molecular basis for drug resistance in HIV-1 protease. *Viruses* **2**, 2509–2535. <https://doi.org/10.3390/v2112509> (2010).
7. UNC Library. *FDA-Approved Anti-HIV Medications* (2021).
8. Mandala, D., Thompson, W. & Watts, P. Synthesis routes to anti-HIV drugs. *Tetrahedron*. <https://doi.org/10.1016/j.tet.2016.04.075> (2016).
9. Chen, J. *et al.* Revealing origin of decrease in potency of darunavir and amprenavir against HIV-2 relative to HIV-1 protease by molecular dynamics simulations. *Sci. Rep.* **4**, 6872. <https://doi.org/10.1038/srep06872> (2014).
10. Lockbaum, G. J. *et al.* Structural adaptation of darunavir analogues against primary mutations in HIV-1 protease. *ACS Infect. Dis.* **5**, 316–325. <https://doi.org/10.1021/acsinfecdis.8b00336> (2019).
11. Nakashima, M. *et al.* Unique flap conformation in an HIV-1 protease with high-level darunavir resistance. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2016.00061> (2016).
12. Gulnik, S. V. *et al.* Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochem.* **34**, 9282–9287. <https://doi.org/10.1021/bi00029a002> (1995).
13. Weber, I. T., Wang, Y.-F. & Harrison, R. W. HIV protease: Historical perspective and current research. *Viruses* **13**, 839. <https://doi.org/10.3390/v13050839> (2021).

14. Tang, M. W., Liu, T. F. & Shafer, R. W. The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology* **55**, 98–101. <https://doi.org/10.1159/000331998> (2012).
15. Yu, Y., Wang, J., Shao, Q., Shi, J. & Zhu, W. Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir. *Sci. Rep.* **5**, 10517–10517. <https://doi.org/10.1038/srep10517> (2015).
16. Maruyama, K., Sheng, Y., Watanabe, H., Fukuzawa, K. & Tanaka, S. Application of singular value decomposition to the inter-fragment interaction energy analysis for ligand screening. *Comput. Theor. Chem.* <https://doi.org/10.1016/j.comptc.2018.04.001> (2018).
17. Kitaura, K., Sugiki, S.-I., Nakano, T., Komeiji, Y. & Uebayasi, M. Fragment molecular orbital method: Analytical energy gradients. *Chem. Phys. Lett.* **336**, 163–170. [https://doi.org/10.1016/S0009-2614\(01\)00099-9](https://doi.org/10.1016/S0009-2614(01)00099-9) (2001).
18. Hengphasatporn, K. *et al.* Halogenated baicalein as a promising antiviral agent toward SARS-CoV-2 main protease. *J. Chem. Inf. Model.* **62**, 1498–1509. <https://doi.org/10.1021/acs.jcim.1c01304> (2022).
19. Ozawa, M., Ozawa, T. & Ueda, K. Application of the fragment molecular orbital method analysis to fragment-based drug discovery of BET (bromodomain and extra-terminal proteins) inhibitors. *J. Mol. Graph. Model.* **74**, 73–82. <https://doi.org/10.1016/j.jmkgm.2017.02.013> (2017).
20. Heifetz, A. Quantum mechanics in drug discovery. In *Quantum Mechanics in Drug Discovery* (2020).
21. Heifetz, A. *et al.* Using the fragment molecular orbital method to investigate agonist-orexin-2 receptor interactions. *Biochem. Soc. Trans.* **44**, 574–581. <https://doi.org/10.1042/bst20150250> (2016).
22. Wang, H. *et al.* MolHyb: A web server for structure-based drug design by molecular hybridization. *J. Chem. Inf. Model.* **62**, 2916–2922. <https://doi.org/10.1021/acs.jcim.2c00443> (2022).
23. Mario Geysen, H., Schoenen, F., Wagner, D. & Wagner, R. Combinatorial compound libraries for drug discovery: An ongoing challenge. *Nat. Rev. Drug Discov.* **2**, 222–230. <https://doi.org/10.1038/nrd1035> (2003).
24. Dounay, A. B. *et al.* Globally distributed drug discovery of new antibiotics: Design and combinatorial synthesis of amino acid derivatives in the organic chemistry laboratory. *J. Chem. Educ.* **96**, 1731–1737. <https://doi.org/10.1021/acs.jchemed.8b00942> (2019).
25. Penner, P. *et al.* FastGrow: On-the-fly growing and its application to DYRK1A. *J. Comput. Aided Mol. Des.* **36**, 639–651. <https://doi.org/10.1007/s10822-022-00469-y> (2022).
26. Gugisch, R., Kerber, A., Laue, R., Meringer, M. & Weidinger, J. MOLGEN-COMB, a software package for combinatorial chemistry. *MATCH Commun. Math. Comput. Chem.* **41**, 189–203 (2000).
27. Chávez-Hernández, A. L., Juárez-Mercado, K. E., Saldívar-González, F. I. & Medina-Franco, J. L. Towards the de novo design of HIV-1 protease inhibitors based on natural products. *Biomolecules* **11** (2021).
28. Green, D. V. S. *et al.* BRADSHAW: A system for automated molecular design. *J. Comput. Aided Mol. Design* **34**, 747–765. <https://doi.org/10.1007/s10822-019-00234-8> (2020).
29. de Souza Neto, L. R. *et al.* In silico strategies to support fragment-to-lead optimization in drug discovery. *Front. Chem.* <https://doi.org/10.3389/fchem.2020.00093> (2020).
30. Monteleone, S. *et al.* Hotspot identification and drug design of protein-protein interaction modulators using the fragment molecular orbital method. *J. Chem. Inf. Model.* **62**, 3784–3799. <https://doi.org/10.1021/acs.jcim.2c00457> (2022).
31. Dwipayana, I. D., Syah, A. P., Aditama, Y., Ferialiana, R. & Fibriani, A. Development of a dimer-based screening system for dimerization inhibitor of HIV-1 protease. *J. Microbiol. Biotechnol.* **2**, 1–11. <https://doi.org/10.37604/jmsb.v2i2.42> (2020).
32. Schmidt, M. W. *et al.* General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347–1363. <https://doi.org/10.1002/jcc.540141112> (1993).
33. Hatada, R. *et al.* Fragment molecular orbital based interaction analyses on COVID-19 main protease—inhibitor N3 Complex (PDB ID: 6LU7). *J. Chem. Inf. Model.* **60**, 3593–3602. <https://doi.org/10.1021/acs.jcim.0c00283> (2020).
34. Fedorov, D. G. & Kitaura, K. Pair interaction energy decomposition analysis. *J. Comput. Chem.* **28**, 222–237. <https://doi.org/10.1002/jcc.20496> (2007).
35. Deetanya, P. *et al.* Interaction of 8-anilinoanthracene-1-sulfonate with SARS-CoV-2 main protease and its application as a fluorescent probe for inhibitor identification. *Comput. Struct. Biotechnol. J* **19**, 3364–3371. <https://doi.org/10.1016/j.csbj.2021.05.053> (2021).
36. Nutho, B. *et al.* Why are lopinavir and ritonavir effective against the newly emerged coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochem.* **59**, 1769–1779. <https://doi.org/10.1021/acs.biochem.0c00160> (2020).
37. Fedorov, D. G. The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications. *WIREs Comput. Mol. Sci.* **7**, e1322. <https://doi.org/10.1002/wcms.1322> (2017).
38. Dolinsky, T. J. *et al.* PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525. <https://doi.org/10.1093/nar/gkm276> (2007).
39. Wittayanarakul, K. *et al.* Insights into saquinavir resistance in the G48V HIV-1 protease: Quantum calculations and molecular dynamic simulations. *Biophys. J.* **88**, 867–879. <https://doi.org/10.1529/biophysj.104.046110> (2005).
40. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748. <https://doi.org/10.1006/jmbi.1996.0897> (1997).
41. Bihani, S. C., Gupta, G. D. & Hosur, M. V. Molecular basis for reduced cleavage activity and drug resistance in D30N HIV-1 protease. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2021.1982007> (2021).
42. Kovalevsky, A. Y. *et al.* Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. *J. Mol. Biol.* **363**, 161–173. <https://doi.org/10.1016/j.jmb.2006.08.007> (2006).
43. Liu, F. *et al.* Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J. Mol. Biol.* **381**, 102–115. <https://doi.org/10.1016/j.jmb.2008.05.062> (2008).
44. Louis, J. M. *et al.* The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts. *Biochem.* **50**, 4786–4795. <https://doi.org/10.1021/bi200033z> (2011).
45. Tie, Y. *et al.* High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* **338**, 341–352. <https://doi.org/10.1016/j.jmb.2004.02.052> (2004).
46. Bandaranayake, R. M. *et al.* The effect of clade-specific sequence polymorphisms on HIV-1 protease activity and inhibitor resistance pathways. *Virology* **84**, 9995–10003. <https://doi.org/10.1128/JVI.00505-10> (2010).
47. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. <https://doi.org/10.1002/jcc.20084> (2004).
48. BIOVIA. (San Diego: Dassault Systèmes, 2021).
49. Chuntakaruk, H. *et al.* Machine learning-guided design of potent darunavir analogs targeting HIV-1 proteases: A computational approach for antiretroviral drug discovery. *J. Comput. Chem.* <https://doi.org/10.1002/jcc.27298> (2024).
50. Bisong, E. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (ed. Bisong, E.) 59–64 (Apress, 2019).
51. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33. <https://doi.org/10.1186/1758-2946-3-33> (2011).
52. Landrum, G. *RDKit: Open-Source Cheminformatics Software*, <<https://github.com/rdkit/rdkit>> (2016).
53. Riniker, S. & Landrum, G. A. Better informed distance geometry: using what we know to improve conformation generation. *J. Chem. Inf. Model.* **55**, 2562–2574. <https://doi.org/10.1021/acs.jcim.5b00654> (2015).

54. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. III. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035. <https://doi.org/10.1021/ja00051a040> (1992).
55. Manchester, J., Walkup, G., Rivin, O. & You, Z. Evaluation of pK(a) estimation methods on 211 drug like compounds. *J. Chem. Inf. Model.* **50**, 565–571. <https://doi.org/10.1021/ci100019p> (2010).
56. Kammarabutr, J., Mahalapbutr, P., Nutho, B., Kungwan, N. & Rungrotmongkol, T. Low susceptibility of asunaprevir towards R155K and D168A point mutations in HCV NS3/4A protease: A molecular dynamics simulation. *J. Mol. Graph. Model.* **89**, 122–130. <https://doi.org/10.1016/j.jmgm.2019.03.006> (2019).
57. Sanachai, K. *et al.* Insights into the binding recognition and susceptibility of Tofacitinib toward Janus Kinases. *ACS Omega* **5**, 369–377. <https://doi.org/10.1021/acsomega.9b02800> (2020).
58. 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.* **25**, 1157–1174. <https://doi.org/10.1002/jcc.20035> (2004).
59. Mahalapbutr, P. *et al.* A theoretical study on the molecular encapsulation of luteolin and pinocembrin with various derivatized beta-cyclodextrins. *J. Mol. Struct.* **1180**, 480–490. <https://doi.org/10.1016/j.molstruc.2018.12.025> (2019).
60. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935. <https://doi.org/10.1063/1.445869> (1983).
61. Chari, R., Jerath, K., Badkar, A. & Kalonia, D. Long- and short-range electrostatic interactions affect the rheology of highly concentrated antibody solutions. *Pharm. Res.* **26**, 2607–2618. <https://doi.org/10.1007/s11095-009-9975-2> (2009).
62. York, D. M., Darden, T. A. & Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **99**, 8345–8348. <https://doi.org/10.1063/1.465608> (1993).
63. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5) (1977).
64. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert. Opin. Drug Discov.* **10**, 449–461. <https://doi.org/10.1517/17460441.2015.1032936> (2015).
65. Sulea, T., Cui, Q. & Purisima, E. O. Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 2. Benchmark in the CSAR-2010 scoring exercise. *J. Chem. Inf. Model.* **51**, 2066–2081. <https://doi.org/10.1021/ci2000242> (2011).
66. Naïm, M. *et al.* Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model.* **47**, 122–133. <https://doi.org/10.1021/ci600406v> (2007).
67. Foulkes-Murzycki, J. E., Rosi, C., Kurt Yilmaz, N., Shafer, R. W. & Schiffer, C. A. Cooperative effects of drug-resistance mutations in the flap region of HIV-1 protease. *ACS Chem. Biol.* **8**, 513–518. <https://doi.org/10.1021/cb3006193> (2013).
68. Mittal, S. *et al.* Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *Virology* **87**, 4176–4184. <https://doi.org/10.1128/JVI.03486-12> (2013).
69. Hengphasatporn, K. *et al.* Multiple virtual screening strategies for the discovery of novel compounds active against dengue virus: A hit identification study. *Sci. Pharm.* **88** (2020).
70. Lockbaum, G. J. *et al.* HIV-1 protease inhibitors with a P1 phosphonate modification maintain potency against drug-resistant variants by increased interactions with flap residues. *Eur. J. Med. Chem.* **257**, 115501. <https://doi.org/10.1016/j.ejmech.2023.115501> (2023).
71. Kneller, D. W., Agniswamy, J., Ghosh, A. K. & Weber, I. T. Potent antiviral HIV-1 protease inhibitor combats highly drug resistant mutant PR20. *Biochem. Biophys. Res. Commun.* **519**, 61–66. <https://doi.org/10.1016/j.bbrc.2019.08.126> (2019).
72. Mahalingam, A. K. *et al.* HIV-1 protease inhibitors with a transition-state mimic comprising a tertiary alcohol: Improved antiviral activity in cells. *J. Med. Chem.* **53**, 607–615. <https://doi.org/10.1021/jm901165g> (2010).
73. Weber, I. T., Kneller, D. W. & Wong-Sam, A. Highly resistant HIV-1 proteases and strategies for their inhibition. *Future Med. Chem.* **7**, 1023–1038. <https://doi.org/10.4155/fmc.15.44> (2015).
74. Hohlfeld, K., Tomassi, C., Wegner, J. K., Kesteleyn, B. & Linclau, B. Disubstituted bis-THF moieties as new P2 ligands in non-peptidic HIV-1 protease inhibitors. *ACS Med. Chem. Lett.* **2**, 461–465. <https://doi.org/10.1021/ml2000356> (2011).
75. Aoki, M. *et al.* A novel HIV-1 protease inhibitor, GRL-044, has potent activity against various HIV-1s with an extremely high genetic barrier to the emergence of HIV-1 drug resistance. *Glob Health Med* **1**, 36–48. <https://doi.org/10.35772/ghm.2019.01003> (2019).
76. Ghosh, A. K., Osswald, H. L. & Prato, G. Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. *J. Med. Chem.* **59**, 5172–5208. <https://doi.org/10.1021/acs.jmedchem.5b01697> (2016).
77. O'Hagan, S. & Kell, D. Structural similarities between some common fluorophores used in biology, marketed drugs, endogenous metabolites, and natural products. *Marine Drugs* **18**, 582. <https://doi.org/10.3390/md18110582> (2020).
78. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142. <https://doi.org/10.1016/j.chembiol.2014.08.013> (2014).
79. Chaudhury, S. & Gray, J. J. Identification of structural mechanisms of HIV-1 protease specificity using computational peptide docking: implications for drug resistance. *Structure* **17**, 1636–1648. <https://doi.org/10.1016/j.str.2009.10.008> (2009).
80. Coburn, C. A. *et al.* Aromatic P1 replacements for the highly potent HIV-1 protease inhibitor CRXIVAN[®]. *Bioorg. Med. Chem. Lett.* **6**, 1937–1940. [https://doi.org/10.1016/0960-894X\(96\)00345-9](https://doi.org/10.1016/0960-894X(96)00345-9) (1996).
81. Lv, Z., Chu, Y. & Wang, Y. HIV protease inhibitors: A review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* **7**, 95–104. <https://doi.org/10.2147/hiv.S79956> (2015).
82. Ghosh, A. & Chapsal, B. Design of the anti-HIV protease inhibitor darunavir. In *Introduction to Biological and Small Molecule Drug Research and Development*, 355–384. <https://doi.org/10.1016/B978-0-12-397176-0.00013-3> (2013).
83. Reddy, G. S. K. K. *et al.* Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres. *J. Med. Chem.* **50**, 4316–4328. <https://doi.org/10.1021/jm070284z> (2007).
84. Potempa, M. *et al.* HIV-1 protease uses bi-specific S2/S2' subsites to optimize cleavage of two classes of target sites. *J. Mol. Biol.* **430**, 5182–5195. <https://doi.org/10.1016/j.jmb.2018.10.022> (2018).
85. Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M. & Schiffer, C. A. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J. Virol.* **78**, 12446–12454. <https://doi.org/10.1128/jvi.78.22.12446-12454.2004> (2004).
86. Lange, O. F. & Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins* **62**, 1053–1061. <https://doi.org/10.1002/prot.20784> (2006).
87. Paulsen, J. L., Leidner, F., Ragland, D. A., Kurt Yilmaz, N. & Schiffer, C. A. Interdependence of inhibitor recognition in HIV-1 protease. *J. Chem. Theory Comput.* **13**, 2300–2309. <https://doi.org/10.1021/acs.jctc.6b01262> (2017).
88. Zhang, Y. *et al.* Structures of darunavir-resistant HIV-1 protease mutant reveal atypical binding of darunavir to wide open flaps. *ACS Chem. Biol.* **9**, 1351–1358. <https://doi.org/10.1021/cb4008875> (2014).
89. Collins, J. R., Burt, S. K. & Erickson, J. W. Activated dynamics of flap opening in HIV-1 protease. *Adv. Exp. Med. Biol.* **362**, 455–460. https://doi.org/10.1007/978-1-4615-1871-6_59 (1995).
90. Harte, W. E. Jr. *et al.* Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA* **87**, 8864–8868. <https://doi.org/10.1073/pnas.87.22.8864> (1990).

91. Scott, W. R. & Schiffer, C. A. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Struct.* **8**, 1259–1265. [https://doi.org/10.1016/s0969-2126\(00\)00537-2](https://doi.org/10.1016/s0969-2126(00)00537-2) (2000).
92. Hamelberg, D. & McCammon, J. A. Fast peptidyl cis-trans isomerization within the flexible Gly-rich flaps of HIV-1 protease. *J. Am. Chem. Soc.* **127**, 13778–13779. <https://doi.org/10.1021/ja054338a> (2005).
93. Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *PNAS* **103**, 915–920. <https://doi.org/10.1073/pnas.0508452103> (2006).
94. Sherry, D., Worth, R., Ismail, Z. S. & Sayed, Y. Cantilever-centric mechanism of cooperative non-active site mutations in HIV protease: Implications for flap dynamics. *J. Mol. Graph. Model.* **106**, 107931. <https://doi.org/10.1016/j.jmgm.2021.107931> (2021).
95. Kneller, D. W., Agniswamy, J., Harrison, R. W. & Weber, I. T. Highly drug-resistant HIV-1 protease reveals decreased intra-subunit interactions due to clusters of mutations. *Febs J* **287**, 3235–3254. <https://doi.org/10.1111/febs.15207> (2020).
96. Perryman, A. L., Lin, J.-H. & McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **13**, 1108–1123. <https://doi.org/10.1110/ps.03468904> (2004).
97. Wolber, G. & Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **45**, 160–169. <https://doi.org/10.1021/ci049885e> (2005).
98. Rakers, C. *et al.* In silico prediction of human sulfotransferase 1E1 activity guided by pharmacophores from molecular dynamics simulations. *J. Biol. Chem.* **291**, 58–71. <https://doi.org/10.1074/jbc.M115.685610> (2016).
99. Sehgal, S. A., Hassan, M. & Rashid, S. Pharmacoinformatics elucidation of potential drug targets against migraine to target ion channel protein KCNK18. *Drug Des. Devel. Ther.* **8**, 571–581. <https://doi.org/10.2147/dddt.S63096> (2014).
100. Trinh, Q. & Le, L. An investigation of antidiabetic activities of bioactive compounds in *Euphorbia hirta* Linn using molecular docking and pharmacophore. *Med. Chem. Res.* **23**, 2033–2045. <https://doi.org/10.1007/s00044-013-0794-y> (2014).
101. Opo, F. A. D. M. *et al.* Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci. Rep.* **11**, 4049. <https://doi.org/10.1038/s41598-021-83626-x> (2021).
102. Hou, T., McLaughlin, W. A. & Wang, W. Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins* **71**, 1163–1174. <https://doi.org/10.1002/prot.21808> (2008).
103. Meher, B. R., Vaishnavi, M., Kumar, M. V. S., Patel, S. & Kaushik, S. R. in *Encyclopedia of Bioinformatics and Computational Biology*.
104. Heal, J. W., Jimenez-Roldan, J. E., Wells, S. A., Freedman, R. B. & Römer, R. A. Inhibition of HIV-1 protease: the rigidity perspective. *Bioinformatics* **28**, 350–357. <https://doi.org/10.1093/bioinformatics/btr683> (2012).
105. Klebe, G. *Drug design* (Springer, 2013).
106. Baldwin, E. T. *et al.* Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenyl-norstatine. *Struct.* **3**, 581–590. [https://doi.org/10.1016/s0969-2126\(01\)00192-7](https://doi.org/10.1016/s0969-2126(01)00192-7) (1995).
107. Silva, A. M., Cachau, R. E., Sham, H. L. & Erickson, J. W. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* **255**, 321–346. <https://doi.org/10.1006/jmbi.1996.0026> (1996).
108. Surleraux, D. L. N. G. *et al.* Design of HIV-1 protease inhibitors active on multidrug-resistant virus. *J. Med. Chem.* **48**, 1965–1973. <https://doi.org/10.1021/jm049454n> (2005).
109. Surleraux, D. L. N. G. *et al.* Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **48**, 1813–1822. <https://doi.org/10.1021/jm049560p> (2005).
110. Bastys, T. *et al.* Consistent prediction of mutation effect on drug binding in HIV-1 protease using alchemical calculations. *J. Chem. Theory Comput.* **14**, 3397–3408. <https://doi.org/10.1021/acs.jctc.7b01109> (2018).
111. Aoki, M. *et al.* A novel central nervous system-penetrating protease inhibitor overcomes human immunodeficiency virus 1 resistance with unprecedented aM to pM potency. *Elife* <https://doi.org/10.7554/eLife.28020> (2017).
112. Brik, A. & Wong, C. H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* **1**, 5–14. <https://doi.org/10.1039/b208248a> (2003).
113. Badaya, A. & Sasidhar, Y. U. Inhibition of the activity of HIV-1 protease through antibody binding and mutations probed by molecular dynamics simulations. *Sci. Rep.* **10**, 5501. <https://doi.org/10.1038/s41598-020-62423-y> (2020).
114. Chang, C. A., Huang, Y. M., Mueller, L. J. & You, W. Investigation of structural dynamics of enzymes and protonation states of substrates using computational tools. *Catalysts* <https://doi.org/10.3390/catal6060082> (2016).

Acknowledgements

H.C. thanks the Second Century Fund, Chulalongkorn University (C2F) for a Ph.D. scholarship. The authors would like to thank to the NSTDA Supercomputer Center (ThaiSC), the Cygnus and middlemol supercomputer, Center for Computational Sciences, and Center for Quantum and Information Lifesciences, University of Tsukuba, Japan for providing the computational resources. This research was partially supported by Tsukuba Basic Research Support Program Type S, University of Tsukuba, Japan.

Author contributions

Conceptualization, K.H., T.R. S.H.; data curation, H.C., K.H., T.K., T.R.; formal analysis, H.C., K.H.; funding acquisition, Y.S., P.M., T.R., S.H.; investigation, H.C., K.H., T.R.; methodology, H.C., K.H., C.A.; project administration, T.R., S.H.; resources, Y.S., T.K., S.H.; software, H.C., K.H., Y.S., T.R.; supervision, T.R., S.H.; validation, H.C., K.H., Y.S., T.K., T.R.; visualization, H.C., K.H.; writing—original draft preparation, H.C., K.H., T.R.; writing—review and editing, H.C., K.H., T.R.; All authors have read and agreed to the published version of the manuscript.

Funding

This Research is funded by Thailand Science research and Innovation Fund Chulalongkorn University (HEA662300079) to T.K. and T.R. We also thank the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund, GCUGR1125652064D), Chulalongkorn University (CE66_036_2300_008), Center of Excellence in Natural Products Chemistry, and Center of Excellence in Computational Chemistry (CECC). The computational resource was supported by a high-performance computing infrastructure project (Grant Number hp200157) to K.H. and Y.S. The calculation was funded by the Kakehashi project of Tsukuba Innovation Arena (TIA) collaborative research program, CREST JST, Japan (Grant Number JP20338388), and the AMED, Japan (Grant Number JP21ae0101047h0001) to Y.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53940-1>.

Correspondence and requests for materials should be addressed to K.H., T.K., T.R. or S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024