

Hybrid QM/MM free energy evaluation of drug-resistant mutational effect on binding of an inhibitor Indinavir to HIV protease

Masahiko Taguchi^{1,2}, Ryo Oyama¹, Masahiro Kaneko¹, and Shigehiko Hayashi¹

¹*Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*

²*Institute for Quantum Life Science, Natural Institute for Quantum Science and Technology, Kizugawa, Kyoto 619-0215, Japan*

Email: hayashig@kuchem.kyoto-u.ac.jp, Phone: +81-75-753-4006, and Fax: +81-75-753-4000

ABSTRACT

Human immunodeficiency virus 1 (HIV-1) protease is a homo-dimeric aspartic protease essential for replication of HIV. The HIV-1 protease is a target protein in drug discovery for antiretroviral therapy, and various inhibitor molecules of transition state analog were developed. However, serious drug-resistant mutants have emerged. For understanding molecular mechanism of the drug-resistance, accurate examination of the impacts of the mutations on ligand binding as well as enzymatic activity is necessary. Here, we present a molecular simulation study on the ligand binding of Indinavir, a potent transition state analog inhibitor, to the native protein and a V82T/I84V drug-resistant mutant of HIV-1 protease. We employed a hybrid ab initio quantum mechanical/molecular mechanical (QM/MM) free energy optimization technique which combines highly accurate QM description of the ligand molecule and its interaction with statistically ample conformational sampling of MM protein environment by long-time molecular dynamics simulations. Through free energy calculations of protonation states of catalytic groups at the binding pocket and of ligand binding affinity changes upon the mutations, we successfully reproduced the experimentally observed significant reduction of the binding affinity upon the drug-resistant mutations and elucidated the underlying molecular mechanism. The present study opens the way for understanding the molecular mechanism of drug-resistance through direct quantitative comparison of ligand binding and enzymatic reaction with the same accuracy.

INTRODUCTION

Human immunodeficiency virus 1 (HIV-1) protease functions in maturation of HIV particles. The protease site-specifically cleavages peptide bonds of polyproteins of HIV through catalytic hydrolysis reactions. Since the protease-mediated maturation is essential for virus infectivity, HIV-1 protease is a target protein for drug discovery and various inhibitors that block its protease activity have been developed [1,2]. HIV-1 protease forms a homo-dimer composed of monomers denoted as A and B, respectively, and the catalytic site is located at the dimer interface (Figures 1 and S1). Two aspartic acids at the positions of 25, Asp25(A) and Asp25(B), in the catalytic site act as the catalytic groups for the hydrolysis reactions of the peptide bonds [3,4].

Inhibitors for HIV-1 protease mainly developed so far are characterized as transition state analogs [1,5,6]. Because the transition state of an enzymatic reaction is energetically stabilized for catalytic activity, the inhibitor molecule that mimics the transition state structure of the enzyme's native substrate is expected to be bound to the enzyme more strongly than the substrate in the reactant and product states. An inhibitor, Indinavir (Figures 1 and 2), is one of such transition state analogs [7-14]. Indinavir possesses an sp^3 secondary alcohol structure, instead of an sp^2 carbonyl one in the native peptide substrate, at the cleavage site in the vicinity of the catalytic Asp25 groups. The secondary alcohol structure of Indinavir mimics a diol form in the transition state of the native substrate where a lytic water molecule is attacking. Together with

bulky groups analogous to hydrophobic residues at the cleavage sites of the HIV polyprotein such as Pro-Phe, Met-Met, and Leu-Ala [15,16], Indinavir is tightly bound to the catalytic site of HIV-1 protease and in turn potently inhibits its enzymatic activity.

However, serious antiviral drug-resistances to HIV-1 protease inhibitors including Indinavir have emerged [13,14,17-27]. Among the drug-resistant mutations, the V82T/I84V mutation greatly decreases the binding affinity of Indinavir (60- to 70-fold increase of the dissociation constant), while no significant structural changes by the mutation from the native one (C_{α} -RMSD ~ 0.2 Å) were observed by an X-ray crystallographic experiment [12]. In addition, reductions of the catalytic activities of the mutant for some of the cleavage sites of the polyprotein remain much more modest [14], giving rise to the drug-resistance. Note that, if the transition state analog inhibitor literally mimicked the native substrate in the transition state, the mutation that reduces the binding affinity of the inhibitor also depressed the enzymatic activity of the mutant to a similar extent and thus abolished the drug-resistant behavior. Hence, the inhibitor that is expected to be tolerantly effective against mutations is a compound that well captures the chemical nature of the native substrate in the transition state. For molecular design of such an inhibitor tolerant to drug-resistance, therefore, simultaneous examinations of the inhibitor binding and the enzymatic activity of the native substrate are necessary.

For understanding and design of the inhibitor binding, determination of protonation state of the

key catalytic carboxyl groups, Asp25, is also necessary. Difference in the protonation state strongly alters electrostatic environment at the interface between the protein and the inhibitor, and thus significantly affects nature of their binding interaction. However, because hydrophobic groups of Indinavir fill space around Asp25 in the binding form [12], electrostatic environment around Asp25 in the inhibitor binding state is significantly different from that in the protein without the inhibitor binding. Furthermore, as the asymmetric peptide substrate and the inhibitor molecule breaks the C_2 symmetry of the homo-dimer HIV-1 protease upon their bindings, the protonation states of the two carboxylic acids of Asp25 in the homo-dimer proteins can differ depending on the substrate and inhibitor molecules bound despite that the proteins are sequentially symmetric.

Although the protonation states for the native substrate and some inhibitors were measured experimentally by biochemical, NMR, and neutron diffraction techniques [28-34], protonation states for the proteins with various inhibitors are not sufficiently examined. Theoretical determination of the protonation states is also a formidable task because the protonation changes a charge of the group and thus evaluation of the free energy associated with the protonation requires high accuracy of the electronic states of the chemical bond formation as well as large electrostatic reorganization of the extensive protein environment. In fact, despite that many theoretical investigations on the protonation states were performed until now [35-49], consensus

on the protonation state for the protein binding Indinavir is seemingly yet to be reached because of their marked discrepancy (Table S1) [36,37,40,41,44].

Here, we present a hybrid molecular simulation study on Indinavir bound to HIV-1 protease and its drug-resistant V82T/I84V mutant. We employed a hybrid quantum mechanical/molecular mechanical (QM/MM) free energy method called the QM/MM RWFE-SCF method [50,51]. The method combines highly accurate *ab initio* QM calculations for a reaction center with long-time molecular dynamics (MD) simulations using MM force fields for an extended protein system in fully variational and computationally efficient manners, allowing one to properly describe mechanistic coupling of a complex chemical event at the reaction center with global conformational changes of the protein system. The method was thus utilized for *ab initio* evaluation of a free energy activation energy of an enzymatic reaction [50,52] as well as structural modeling and design of photo-receptor proteins [53-56]. Especially, the method successfully predicted functional conformational changes of the proteins upon a photo-chemical reaction and mutations, which were later experimentally confirmed by X-ray crystallographic measurements [53,57], showing a high reliability of molecular modeling of a protein with a chemically complex ligand compound and its mutants.

In the present study, we treated the whole molecule of Indinavir and the catalytic sidechains of Asp25 quantum mechanically. The advantage of the QM/MM simulation approach over a

simulation only with MM force fields is three-fold. First, the quantum treatment avoids difficulty in developing an accurate force field of ligand molecules. Although several efficient and accurate schemes for parameterizing classical MM force fields for small ligand molecules have been proposed [58,59], it is not obvious that classical force fields are accurate enough to describe structures and energies of small ligand molecules in proteins. Especially, in the present system, Indinavir hydrogen-bonds with a possibly anionic sidechain(s) of deprotonated Asp25, which, for accurate evaluation of structure and energetics, requires description of electronic polarization and charge transfer interactions absent in conventional MM force fields. Furthermore, Indinavir exhibits several intramolecular interactions of hyperconjugation for which the QM treatment provides essentially more accurate description.

An apparent shortcoming of a conventional QM/MM MD simulation, which is able to treat the molecular interactions quantum mechanically, for a ligand binding problem is its considerably demanding computational costs which preclude long-time MD simulations for sufficient statistical sampling. However, the present approach of the QM/MM RWFE-SCF method enables optimizations of electronic wavefunction and structure of the reaction center and the ligand binding site described quantum chemically on an extensive free energy surface of the protein environment obtained by long-time MD simulations with MM force fields, and thus partly overcomes the difficulty of conventional QM/MM MD simulations in obtaining sufficient

statistical conformational samples.

Second, the protonation states of Asp25 can be determined to properly evaluate the binding affinity of the inhibitor as described above. Since the protonation of the carboxyl group involves chemical bond formation and dissociation, the QM/MM treatment is a straightforward and accurate approach. It is again noteworthy that the QM/MM RWFE-SCF method employed in the present study is capable of sufficiently sampling conformational changes of the protein environment representing reorganization in response to significant change of electrostatic potential upon the protonation change by the long-time MD simulations with MM force fields in the QM/MM optimization procedure.

Finally, to understand the molecular mechanism of drug resistance, the method allows one to treat both of the ligand binding process and the enzymatic reaction one. For the latter, one inevitably needs to employ the QM/MM approach as the chemically active transition state determines the enzymatic activity. It is therefore preferable to employ the same QM/MM approach for the ligand binding process as well, which avoids introducing a possible error due to use of computationally different methods.

In the present study, we first identified the protonation states of Asp25 for both the native protein and the V82T/I84V mutant binding Indinavir by QM/MM RWFE-SCF free energy optimizations and free energy perturbation calculations. We found that water hydration to the

symmetric carboxyl groups of Asp25 in the homo-dimer is asymmetric, which determined the protonation states of Asp25. We then calculated a difference in binding free energy of the inhibitor between the native protein and the mutant by alchemical free energy perturbation calculations [60-67]. The QM/MM calculation successfully reproduced the significant reduction of the binding free energy of the inhibitor upon the V82T/I84V mutation, which was not well described by the simulation only with the MM force fields. The QM/MM approach which enables the evaluation of the effect of drug-resistant mutations on the inhibitor binding opens the way for understanding of the molecular mechanism of the drug-resistance.

COMPUTAIONAL METHODS

To describe conformational and energetic couplings of the ligand binding with slow and extensive thermal relaxation of the native protein and its mutant, we employed the QM/MM RWFE-SCF method by which the molecular geometry of the QM part is optimized at the ab initio level of theory on an extensive mean-field free energy surface constructed from statistical samples of the MM part generated by long-time MD simulations [50,51]. In the method, iterative calculations of the ab initio QM/MM geometry optimizations and the long-time MD samplings called the sequential samplings are carried out until simultaneous convergences of the QM geometry optimization and the MD statistical samplings are achieved. All MD simulations and the QM/MM calculations were performed with AMBER [68,69] and GAMESS [70] with locally implemented

QM/MM codes, respectively.

MD simulations of the native protein. We first performed conventional MD simulations to set up the initial simulation systems for the QM/MM RWFE-SCF calculations. We utilized a crystallographic protein structures of HIV-1 protease with Indinavir (PDB ID: 1HSG) [12] and HIV-1 protease with inhibitor, JE-2147 (PDB ID: 1KZK) [71]. The amino acid sequences of the HIV-1 proteins deposited in PDB are not mutually identical because of additional mutations. We therefore took the amino acid sequence of 1KZK (Figure S1) as the reference sequence and the structure of 1HSG as the template structure of HIV-protease binding Indinavir, respectively, and modeled the structure of HIV-protease binding Indinavir by modifying the residues of 1HSG that are different from those of the reference amino acid sequence of 1KZK. Missing hydrogen atoms were then added to the protein and Indinavir with LEaP module of AMBER. We assumed standard protonated state for the protein and Indinavir, except for two aspartic acids Asp25(A) and Asp25(B) in the catalytic site. Because two aspartic acids are generally mono-protonated in the case of asymmetric type inhibitors [30-34], we modeled two mono-protonated states, AH and BH, where either Asp25(A) or Asp25(B) is protonated, respectively. Each protein was immersed in a nearly cubic boxes ($96.1 \times 96.3 \times 96.1 \text{ \AA}^3$) in a periodic boundary condition filled with TIP3P water molecules [72] and seven Cl^- ions were added to neutralize the system. The total number of atoms in the box was 76,041 (Figure 1b). The AMBER ff14SB parameter set [73] and the

parameter set reported in Ref. [74] were employed for the force fields of the protein and the Cl^- ions, respectively. The force field parameters of Indinavir were taken from AMBER parameter database [41] and missing force field parameters were supplemented by GAFF [58]. Long range part of electrostatic interactions was calculated with particle mesh Ewald method [75]. Short-range nonbonded interactions were cut off at 8 Å. Length of bonds including hydrogen atoms were constrained by SHAKE/RATTLE method [76,77]. The integral time step was set to be 2 fs. The two simulation systems were firstly potential energetically minimized. Temperature was then raised from 0 K to 300 K during 1 ns in NPT condition. Equilibrium MD simulations at 300 K for 1 μs in NPT condition were performed after the heating. Temperature and pressure were maintained with Langevin bath (collision frequency is 2 ps^{-1}) and Berendsen's method, respectively [78].

QM/MM RWFE-SCF geometry optimizations for the native protein. The QM regions consisted of Indinavir and the sidechains of Asp25(A) and Asp25(B) (Figure 2). The dangling bonds at the boundaries of the QM and MM regions were capped with dummy hydrogen atoms. The total number of the QM atoms were 105. The density functional theory (DFT) method with B3LYP-D3 functional [79,80] and 6-31G** basis set was used, except for the two carboxyl groups of Asp25 where 6-31+G** basis set was employed, and for the boundary atoms, the hydrogen atoms connecting to the boundary ones, and the dummy hydrogen atoms where 6-31G basis set

was used. The total number of the basis functions were 1,059. Restraint parameters for the restrained electrostatic potential (RESP) charge operators [81] were set to be 0.005, except for those of boundary atoms and the hydrogen atoms connecting to the boundary atoms, which were set to be 0.025, respectively. The gradient convergence tolerance for the geometry optimizations of the QM region was set to be 5.0×10^{-4} Hartree/Bohr.

To obtain the initial structures and effective atomic charges of the QM regions for the QM/MM RWFE-SCF calculations, the equilibrated MD systems described above were cooled down to 0 K and QM/MM potential energy geometry optimizations at 0 K with fixed MM coordinates taken from the last snapshots of the cooling MD simulations were carried out. For the BH system, temperature of the equilibrated MD system was decreased from 300 K to 0 K during 4 ns in the constant volume condition from the last snapshot of the equilibrium MD simulation. For the AH system, a cooling simulation for 1 ns starting at 726 ns of the equilibrium MD simulation was performed with restraints to arrest extensive conformational fluctuation of the binding site and to maintain conformation of hydrogen-bond network in the catalytic site similar to the main one (see below). The harmonic restraints with a force constant of 2.5 kcal/(mol \cdot Å²) are imposed for the distance between H₈₂ atom of Asp25(A) and O₂ atom of Indinavir and that between H₂ atom of Indinavir and O₈₂ atom of Asp25(B), respectively, when they became longer than 2 Å.

The optimized structures obtained by the QM/MM potential energy geometry optimizations at

0 K were embedded in the last snapshots of the cooling MD simulations. Temperatures of the systems where the geometries and effective atomic charges of the QM region were fixed were then increased from 0 K to 300 K during 1 ns in the constant volume condition. The first cycles of the MD simulations of the sequential samplings of the QM/MM RWFE-SCF calculations were then initiated after the heating. In each cycle of the sequential samplings, an MD trajectory calculation for 10 ns in NVT condition was performed for the fixed geometry and effective atomic charges of the QM region, and 50,000 conformational samples obtained from the last 5 ns was used for the MM statistical ensemble. The free energy geometry optimizations required 110 and 159 cycles for the protonation states AH and BH, respectively (i.e., the MD simulations for 1.1 and 1.59 μ s in total). A 1 μ s MD simulation of the AH system of which the geometry and the effective point charges of the QM region were fixed at the QM/MM free energetically optimized ones was then carried out for analysis of the protein conformations.

Calculations of the V82T/I84V Mutant. We manually modified the native protein structures in the AH and BH protonation states at 0 K to model initial structures of the V82T/I84V mutant. The native protein structures at 0 K were obtained by MD simulations for 4 ns in NVT condition which cooled down to 0 K for the simulation systems of the native proteins optimized by the QM/MM RWFE-SCF calculations described above. During the cooling simulations, the geometries and effective atomic charges of the QM regions were fixed at the free energetically

optimized ones. To check stability of the mutant proteins, equilibrium MD simulations with the MM force fields for 1 μ s in NPT condition were performed after potential energy minimization and heating to 300 K.

The QM/MM RWFE-SCF free energy optimizations were initiated by the procedures as follows. First, the MM parts of the initial mutant protein models at 0 K described above were potential energetically minimized with the fixed geometries and effective atomic charges of the QM regions. Equilibrium MD simulations for 500 ns at 300 K in NVT condition were then performed with the fixed geometries and effective atomic charges of the QM regions after heating for 1 ns, and the first cycles of the sequential samplings of QM/MM RWFE-SCF method were started after the equilibrations. During the free energy optimization of the mutant system initially in the BH protonation state, the proton at Asp25(B) was spontaneously transferred to Asp25(A) around at 5th sequential samplings and stayed in the AH protonation state after the proton transfer (see below). The free energy geometry optimizations required 96 and 147 cycles for the systems in the initial protonated states, AH and BH, respectively (MD simulations for 0.96 and 1.47 μ s in total). For the evaluation of the free energy difference between the AH and BH protonation states, we also performed a QM/MM RWFE-SCF free energy optimization of the mutant in the BH state by imposing constraint to fix bond distance between O₈₂ and H₈₂ atoms of Asp25(B) at 1.07 Å in order to prevent the spontaneous proton transfer to Asp25(A). The constraint free energy

optimization in the BH state finished at 95 cycles (MD simulations for 0.95 μ s in total).

Evaluation of free energy differences between AH and BH protonation states. To energetically characterize the free energetically optimized AH and BH protonation states, free energy differences between those protonation states were evaluated by free energy perturbation calculations as described previously [52]. The free energy differences between the free energetically optimized states X and Y are given as

$$\Delta_{Y-X}F_{\text{QM/MM}} = \Delta_{Y-X}E_{\text{QM}} + \Delta_{Y-X}F_{\text{QM-MM,MM}} \quad (1)$$

where $\Delta_{Y-X}E_{\text{QM}}$ is the difference between the expectation values of the QM Hamiltonian and $\Delta_{Y-X}F_{\text{QM-MM,MM}}$ is the free energy difference originating from QM-MM interactions and MM interactions in the MM region. The former was directly obtained by the QM/MM RWFE-SCF free energy optimization, while the latter was evaluated with Bennett acceptance ratio (BAR) method of free energy perturbation (FEP) [82-84]. For the calculation of FEP, the geometry and the effective atomic charges of the QM region were changed with linearly divided discrete N points as,

$$\begin{aligned} \mathbf{R}(i) &= \lambda_i \mathbf{R}_Y + (1 - \lambda_i) \mathbf{R}_X \\ \mathbf{q}(i) &= \lambda_i \mathbf{q}_Y + (1 - \lambda_i) \mathbf{q}_X \\ \lambda_i &= \frac{i}{N-1}, \quad i = 0, 1, \dots, N-1 \end{aligned} \quad (2)$$

where $\mathbf{R}_{X/Y}$ and $\mathbf{q}_{X/Y}$ are the coordinates and the effective atomic charges of the QM region in the X/Y states. The MM force field parameters including mainchain atomic charges of Asp25

were also changed simultaneously, because the force field parameters of a deprotonated aspartic acid are also different from those of a protonated one in Amber ff14SB force field, and the Lennard-Jones interactions and the intramolecular interactions at the QM-MM boundaries were evaluated with the MM force fields [50,52].

The MD simulation at each point was started from the last snapshot of the MD trajectory calculation at the previous point. Conformational samples of the MM region at each point were obtained by an MD trajectory calculation for 50 ns, and 50,000 samples taken from the trajectory of the last 25 ns were employed for the evaluation of free energy difference between the neighboring two points. The total length of the MD trajectory calculation is 1 μ s, i.e., $N = 20$. The forward calculation where AH was changed to BH (denoted $AH \rightarrow BH$) as i is incremented and the backward one where BH was changed to AH (denoted $BH \rightarrow AH$) as i is decremented were performed to assess the statistical convergence.

MD Simulation for inhibitor unbound states. For evaluation of ligand binding energies described below, MD simulations of the inhibitor unbound states (IUSs) were needed. Simulation systems of IUSs of the native proteins and the V82T/I84V mutant were obtained by removing the inhibitor molecule from the initial structures of the inhibitor bound states (IBSs) described above, respectively. The two carboxyl groups of Asp25 in the dimer were set to be both deprotonated based on experimental evidences [3,30]. For comparison, MD simulations for IUSs with the

mono-protonated carboxyl groups of Asp25 and the deprotonated ones were also carried out. The procedure of the MD simulations of IUSs are the same as that of IBSs, and equilibrium MD simulations for 1 μ s each for the three protonation states were performed.

Change in binding affinity of Indinavir upon the V82T/I84V mutation. The QM/MM structures and the energies were first refined by QM/MM RWFE-SCF free energy optimizations with DFT M06-2X functional [85] from those obtained with DFT B3LYP-D3 functional (see Results). The free energy geometry optimizations required 119 and 170 cycles of the sequential samplings for the native protein and the V82T/I84V mutant, respectively (MD simulations for 1.19 and 1.70 μ s in total).

Change in free energy difference between IBS and IUS upon the mutation, $\Delta_{M-N}\Delta_b F$, was evaluated through a thermodynamic cycle (Figure S2),

$$\Delta_{M-N}\Delta_b F \equiv \Delta_b F_M - \Delta_b F_N = \Delta_{M-N}F_{IBS} - \Delta_{M-N}F_{IUS} \quad (3)$$

where $\Delta_b F_M$ and $\Delta_b F_N$ are free energy differences between IBS and IUS, i.e., binding affinities, of the mutant and the native protein, respectively. However, the direct simulations of the ligand binding processes involving large conformational changes of the proteins and the ligand molecules require considerably long MD calculations, and thus are computationally too demanding. The evaluation of $\Delta_{M-N}\Delta_b F$ is therefore converted into the calculations of free energy differences between the mutant and the native protein in IBS and IUS, $\Delta_{M-N}F_{IBS}$ and

$\Delta_{M-N}F_{IUS}$, respectively, which were computed with alchemical FEP techniques [60-67].

In the case of $\Delta_{M-N}F_{IBS}$ for the QM/MM systems, QM/MM contributions are also included as

$$\Delta_{M-N}F_{IBS} = \Delta_{M-N(QM)}E_{QM} + \Delta_{M-N(QM)}F_{QM-MM,MM} + \Delta_{M-N}F_{alchemy} \quad (4)$$

$\Delta_{M-N(QM)}E_{QM}$ and $\Delta_{M-N(QM)}F_{QM-MM,MM}$ are contributions of the QM energy, and the QM-MM interaction energy and the MM one, respectively, upon change of the QM region from the native protein to the mutant while the mutation groups were unchanged. Those terms were first computed by the FEP procedures described above where the geometry and the charges of the QM region were gradually altered from those of the native protein to the mutant. The geometry and the effective atomic charges of the QM parts were divided into 10 discrete points ($N = 10$ in Eq. (2)). Conformational samples of the MM parts at each point were obtained by an MD trajectory calculation for 10 ns, and 50,000 samples taken from the trajectory of the last 5 ns were employed for the evaluation of free energy difference between the neighboring two points. The total length of the MD trajectory calculation was therefore 100 ns.

$\Delta_{M-N}F_{alchemy}$ was then computed by the alchemical FEP calculation for the system including the fixed QM region of the mutant where the mutation groups were gradually changed with the dual-topology mode implemented in AMBER program package [68]. The thermodynamic integrations were carried out with soft-core potentials [60]. The parameters of atoms in the

mutated residues were divided into 20 discrete states. At the initial end state, an energy minimization for 200 steps and a heating MD simulation for 20 ps from 50 K to 300 K were performed. Then, conformational samples of the MM parts at each discrete point were obtained by a MD trajectory calculation in NPT condition for 10 ns, and 250 samples taken from the last 5 ns were employed for the evaluation of free energy. The total length of the MD trajectory calculation is therefore 200 ns. For comparison, the alchemical FEP calculations for IBS were performed by MD simulations where the QM region was treated with the MM force field. The free energy change upon the mutation for IUS, $\Delta_{M-N}F_{IUS}$, was also evaluated by the alchemical FEP calculations with the same procedure.

The direction of the changes from the native protein to the mutant is defined as the forward direction. The calculation in the backward direction from the mutant to the native protein was also carried out to assess the statistical convergence. Note that the thermodynamic path in the backward direction does not coincide with that in the forward one. The FEP change in the forward direction is native QM/native MM \rightarrow mutant QM/native MM \rightarrow mutant QM/mutant MM, and that in the backward one is mutant QM/mutant MM \rightarrow native QM/mutant MM \rightarrow native QM/native MM. Thus, the intermediate states (mutant QM/native MM vs. native QM/mutant MM) in the thermodynamic paths are different.

RESULTS

We first compared structures of the Indinavir binding site of the native protein and the V82T/I84V mutant obtained by MD simulations with classical MM force fields and the QM/MM RWFE-SCF free energy optimizations. Next, the protonation states of the two catalytic aspartic acids, Asp25(A) and Asp25(B), were determined by free energy calculations for the free energetically optimized structures obtained by the QM/MM RWFE-SCF method. Finally, change in the binding free energy of Indinavir upon the V82T/I84V mutation was evaluated with alchemical free energy perturbation calculations.

MD simulations of the native protein and the V82T/I84V mutant. We performed MD simulations with classical MM force fields for 1 μ s for the native protein and the V82T/I84V mutant in the two mono-protonated states, AH and BH, where one of the catalytic aspartic acids, Asp25(A) and Asp25(B), is protonated, respectively (see COMPUTATIONAL METHODS). The two aspartic acids of Asp25 were experimentally suggested to be mono-protonated for asymmetric inhibitors [30-34]. One can discern in the MD trajectories of the native protein and the V82T/I84V mutant that conformations of the sidechains of the protonated Asp25(A) and Asp25(B) are distinctly different (Figure 3). The main distribution of the dihedral angle of C_{α} - C_{β} - C_{γ} - O_{δ} of the protonated Asp25(A) in AH is centered at ~ 120 degrees, while the dihedral angle of the protonated Asp25(B) in BH is sharply peaked at -60 degrees. The difference in the dihedral angle by ~ 180

degrees indicate that the hydrogen-bonds of the protonated oxygens of Asp25 to the hydroxy group of Indinavir are formed from the same side (from the behind side in the molecular images of Figure 3) and not in C_2 symmetry because of the asymmetric binding of Indinavir despite the C_2 symmetry of the protein. The sidechain of the protonated Asp25(A) is somewhat flexible, as suggested by a previous MD study [48], compared to that of the protonated Asp25(B). All the dihedral angles of C_α - C_β - C_γ - O_δ of the deprotonated Asp25 of the native protein and the mutant in AH and BH are ~ 0 degrees, indicating that conformations of the protonated and deprotonated Asp25 in each protonate state are asymmetric.

During the MD simulations for 1 μ s, water molecules move in and out of the binding site (Figures S3 and S4). In most of the periods of the trajectories, one water molecule occupied cavities in the binding site, while no water molecules were found in an X-ray crystallographic structure [12]. No significant structural differences of Asp25 and the water occupation between the native protein and the V82T/I84V mutant were found.

QM/MM RWFE-SCF geometry optimizations. Structures of the bound Indinavir and the catalytic Asp25 in the native protein were refined by QM/MM RWFE-SCF geometry optimizations starting from snapshots of the main conformations in the MD trajectories described above (see COMPUTATIONAL METHODS for details). The free energetically optimized structures of the native protein (Figure 4) showed that large conformational changes of Asp25

associated with occupations of more water molecules in the binding sites.

Prominent conformational changes appeared at the deprotonated aspartic acids of Asp25. In the initial conformations of the AH protonation state taken from the MD simulations, one of the two O_{δ} atoms of the carboxyl group of Asp25(B) was hydrogen-bonded with two N-H groups of the main chain amides of Gly27 of the protein dimer (Figure 5a). During the free energy optimization, the hydrogen-bond of the O_{δ} atom with one of the two N-H groups of Gly27 was dissociated, while that with the other N-H group was maintained (Figure 5a). Orientation of the carboxyl group of Asp25(B) was also significantly changed (Figures S5a,c and S6, and Table S2). The conformational change of the deprotonated Asp25(B) was accompanied by hydration with water molecules coming into cavities newly created by the conformational change, and consequently four water molecules occupied the cavities in the binding pocket (Figures 4a and 6a). The water hydration in the binding pocket was not observed in the X-ray crystallographic structure [12]. The discrepancy may arise from the environment around the protein; the X-ray crystallographic structure was determined in a crystal packing at high concentration of the protein, i.e., in much fewer water molecules, while the protein is fully solvated in a water box in the present simulation system. Given that no apparent counter ion group of the deprotonated Asp25 is found and the hydration in the binding pocket is closely related to the protease function, i.e., hydrolysis of a peptide bond, the hydration in the binding pocket observed in the present study seems physically

and biochemically reasonable.

Similar conformational changes of the deprotonated Asp25(A) and hydration with water molecules during the QM/MM RWFE-SCF geometry optimization were also observed for the BH protonation state (Figures 4b, 5b, 6b, and S5b,d). However, because of the asymmetric binding of Indinavir, the conformational changes of the deprotonated Asp25(A) were less pronounced (Figures 5b and S5b,d). Furthermore, although water molecules hydrating the catalytic aspartic acids increased during the free energy optimization, cavities were not fully created around Asp25(A) because of tight hydrophobic packing among the tertiary butyl and phenyl groups of Indinavir and Val82 and Ile84 of the protein, and thus the deprotonated Asp25(A) was less hydrated (Figure 6b). In addition, Asp25(B) was also less hydrated because Asp25(B) is protonated. Consequently, the number of water molecules occupying the cavities in the binding pocket in the BH state in the free energetically optimized structure was three (Figure 6b), which is less than that in the AH state, four, as described above (Figure 6a).

In addition to the conformational changes of the catalytic Asp25, molecular structures of Indinavir were also altered by the QM/MM free energy optimizations. Because of strong hydrogen-bonds of the hydroxy group of Indinavir with the carboxyl groups of Asp25, the conformational changes of Asp25 described above were correlated with a deeper binding of the hydroxy group of Indinavir inside the binding cleft (Figure 4e,f). The deeper binding of the

hydroxy moiety of Indinavir was also achieved by rotations of dihedral angles of Indinavir's backbone bonds adjacent to the hydroxyl group (Table S3). The dihedral angles of N₃-C₁₀-C₁₁-C₁₂ and C₁₀-C₁₁-C₁₂-C₁₃ of the QM/MM free energetically optimized structure in the AH state deviated remarkably from those in the X-ray crystallographic structure by 17.1 and -22.4 degrees, respectively, and in the MD simulation with the MM force fields by 26.5 and -11.1 degrees, respectively. Several dihedral angles around N₃-C₁₀, C₁₁-C₁₂, and C₁₃-C₂₁ determined by the QM/MM RWFE-SCF optimizations were also found to deviate by more than 10 degrees from those of the MD simulations with the MM force field. Those dihedral angles are influenced by the deeper binding of Indinavir described above, as they are located in the vicinity of the hydroxy group hydrogen-bonded with Asp25.

Internal molecular structures of other moieties of Indinavir were also changed by the QM/MM free energy optimizations (Table S3). The dihedral angles around C₃₁-C₃₂ in the QM/MM free energetically optimized structure changed largely from those in the MD simulation with the MM force field by ~20 degrees. The dihedral angles represent rotation of the pyridyl group which involves complex change of σ - π interaction and thus is difficult to describe with the MM force field. The high accuracy of the QM description improved the interaction and thus successfully refined the molecular structure. The difference in the conformation of the pyridyl group of Indinavir may also be correlated with hydrophobic interaction of the group with Pro81(B) and

Val82(B) of the protein (Figure 7). The hydrophobic sidechains of Pro81(B) and Val82(B) formed a more compact conformation with the aromatic pyridyl group (Figure 7b) in the QM/MM samples, while a looser conformation of those sidechains (Figure 7c) was frequently found in the MM samples, implying that a stronger hydrophobic interaction of the pyridyl group of Indinavir is established in the QM/MM free energetically optimized structure.

The dihedral angles around C₃₁-C₃₂ in the QM/MM free energetically optimized structure also largely deviated by ~37 degrees from that in the X-ray crystallographic structure (Figure 4e,f and Table S3). The deviation is attributed to difference in environment around the pyridyl group between the X-ray crystallographic structure and the simulation system. The pyridyl group is exposed to bulk water environment in the simulation system and thus its environment is expected to be very different from the environment of the X-ray crystallographic structure influenced by crystal packing. The dihedral angles of C₂₂-C₂₃ and C₂₃-C₂₄ optimized by the QM/MM RWFE-SCF calculations also deviated from those obtained by the MD simulations with the MM force field, indicating that the ab initio description improved the conformation around those dihedral angles.

The structures of Indinavir and its binding site in the V82T/I84V mutant were also refined by equilibrium MD simulations for 500 ns each and the following QM/MM RWFE-SCF geometry optimizations from starting structures obtained based on the free energetically optimized

structures of the native proteins described above (see COMPUTATIONAL METHODS). In the case of the AH state, the conformations of Asp25, the water hydration around them, and the internal structure of Indinavir for the V82T/I84V mutant were almost the same as those for the native protein (Figures 4a,c, 6c, S7a, and S8a).

In the case of the BH state, however, because a cavity around Asp25(B) is enlarged by the I84V mutation where the size of the sidechain at the position of 84 decreases, one water molecule around Asp25(B), which was excluded from the binding site upon formation of the BH state from the AH one for the native protein as described above (Figure 4a,b), partially occupied the cavity in the BH state in an equilibrium MD simulation for 500 ns before the free energy geometry optimization (Figure S7b). The occupation of the additional one water molecule may also be reinforced by increase of polarity by the V82T mutation. Consequently, the number of the hydrating water molecule in the binding pocket in the BH state of the V82T/I84V mutant fluctuates between three and four and increased from that of the native protein (Figures 6d and S7b).

Because of the increased hydration of Asp25(B) for the V82T/I84V mutant, the proton attached to Asp25(B) spontaneously transferred to Asp25(A) through the hydroxy group of Indinavir in an early cycle of the following QM/MM RWFE-SCF geometry optimization around 50 ns, leading to formation of the AH state. The spontaneous proton translocation forming the AH state clearly

indicates that a free energetically quasi stable state does not exist for the V82T/I84V mutant in the BH state, and thus the protonation state of the V82T/I84V mutant is the AH one.

For the free energy calculations of the protonation states described later, a model of the V82T/I84V mutant in the BH state was obtained by the QM/MM RWFE-SCF geometry optimization with a fixed distance of the O-H group of Asp25(B) which prevents the spontaneous proton transfer forming the AH state (see COMPUTATIONAL METHODS). The conformation of Asp25 and the internal structure of Indinavir were again almost the same as those of the native protein in the BH state (Figures 4b,d and S8b). The number of the water molecules hydrating Asp25 in the binding sites in the BH state increased to nearly four as described above, which is the same as that in the AH state (Figures 4c,d and 6c,d)

Finally, one water molecule tetrahedrally bridging two amide oxygen atoms of Indinavir (O_1 and O_3 atoms) and two amide nitrogen atoms of Ile50(A) and Ile50(B) through hydrogen-bonds was found for all of the optimized structures, respectively (Figure S9). The bridging structural water molecule was observed by various experiments including the X-ray crystallographic structure (1HSG) employed in the present study [1,4,5,12], and is considered to strengthen the binding of Indinavir.

Protonation states of the catalytic aspartic acids, Asp25. Free energy differences between the AH state and the BH one, optimized by the QM/MM RWFE-SCF methods described above

were evaluated by MD simulations of FEP calculations with BAR method (see COMPUTATIONAL METHODS). The free energy difference, $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$, was computed to be 2.8 kcal/mol (Table 1), indicating that the AH state is free energetically more stable than the BH one. Ample sampling of the long MD simulations in the forward and backward directions for 1 μ s each provided a sufficient statistical convergence represented by small difference between $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$ obtained in those directions by 0.4 kcal/mol, while shorter MD simulations led to larger deviations of $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$ by several kcal/mol (Tables 1 and S4). Dependency of density functional used in the QM/MM calculation on the free energy difference was also assessed by computing $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$ in Eq. (1) with various density functionals at the same optimized QM geometry and the same mean field of electrostatic potential from the MM region. Deviations from the value with B3LYP-D3 functional employed in the present study were less than 1.5 kcal/mol (Tables S5), and thus the conclusion that the AH state is free energetically more stable holds.

The free energy difference originates from two contributions of the QM energy, $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$, and the QM-MM interaction energy and the MM energy, $\Delta_{\text{BH-AH}}\Delta F_{\text{QM-MM,MM}}$, respectively, as shown in Eq. (1). $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$ gave a largely negative contribution of -18.1 kcal/mol, indicating that the QM energy in the BH state is considerably lower than that in the AH state. Given that the QM region includes Indinavir and the catalytic carboxylic acids of Asp25, and the internal

conformations of Indinavir in the AH and BH states are not significantly different, the large energy QM contribution is attributed to the difference in the interaction of Indinavir with Asp25 due to the asymmetric binding. The hydroxy group of Indinavir therefore energetically favors interaction with the deprotonated Asp25(A) in the BH state over the deprotonated Asp25(B) in the AH state, as the hydroxy group interacts more strongly with the anionic deprotonated carboxylic acid than the neutral protonated one.

In contrast, the other contribution of $\Delta_{\text{BH-AH}}\Delta F_{\text{QM-MM,MM}}$, 20.9 kcal/mol, was largely positive, and overcompensated that of $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$, leading to the small positive overall contribution of $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$, 2.8 kcal/mol. The large positive contribution of $\Delta_{\text{BH-AH}}\Delta F_{\text{QM-MM,MM}}$ comes from the asymmetric hydration of the catalytic carboxylic acids, Asp25, as described above (Figure 4a,b). As Asp25(B) is more hydrated than Asp25(A), the AH state where Asp25(B) is deprotonated and negatively charged, is more stabilized than the BH one.

The protonation state of the V82T/I84V mutant was determined to be the AH state by the QM/MM RWFE-SCF optimization where the proton at Asp25(B) in the BH state was spontaneously transferred to Asp25(A) as described above. The protonation state of the mutant was also verified in terms of energetics by the FEP calculations. The BH state was modeled by the QM/MM RWFE-SCF optimization with a fixed O-H bond length of the protonated Asp25(B) as described above. The free energy difference, $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$, was evaluated to be 4.6 kcal/mol

(Table 1), and thus the AH state is energetically more stable than the BH state, which agrees with the observation of the QM/MM RWFE-SCF optimization described above. Furthermore, the free energy difference of the mutant is larger by 1.6 kcal/mol than that of the native protein, which is in line with the spontaneous proton transfer during the QM/MM RWFE-SCF optimization.

As in the case of the native protein, $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$ of the mutant, -13.2 kcal/mol, was largely negative and the largely positive contribution of $\Delta_{\text{BH-AH}}\Delta F_{\text{QM-MM,MM}}$, 17.8 kcal/mol, overcompensated $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$, resulting in the positive overall free energy difference, $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$ (Table 1). However, $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$ of the mutant is significantly larger by 4.9 kcal/mol than that of the native protein, which leads to the larger $\Delta_{\text{BH-AH}}\Delta F_{\text{QM/MM}}$ of the mutant. Given that the conformations of the binding site of the native protein and the mutant in the AH state are similar, and thus the AH states are not likely responsible for the change, the increase of $\Delta_{\text{BH-AH}}\Delta E_{\text{QM}}$ of the mutant is attributed to the stronger hydration of the protonated Asp25(B) in the BH state (Figure 4b,d). As described above, the number of the water molecules hydrating the protonated Asp25(B) in the mutant is larger than that in the native protein. The stronger hydration of the protonated Asp25(B) which decreases its pKa elevates more largely the QM energy of the BH state.

Effect of the V82T/I84V drug-resistant mutation on binding affinity of Indinavir. We first refined the QM/MM structures and the energies by QM/MM RWFE-SCF free energy

optimizations with DFT M06-2X functional [85] from those obtained with DFT B3LYP-D3 functional (see COMPUTATIONAL METHODS). In an ongoing computational study on enzymatic catalysis of HIV-1 protease in our group, we found that a free energy of an active reaction intermediate is strongly dependent on free energy functionals employed, and that obtained with M06-2X functional is in better agreement with that at a higher level of theory, which will be reported elsewhere. As the binding free energies of the inhibitor molecule evaluated in the present study will be compared with those of the transition states of the enzymatic reactions to elucidate molecular mechanism of the drug resistance as described above, we obtained the formers with the same M06-2X functional as well. The free energetically optimized structures of the native proteins and the V82T/I84V mutant with M06-2X functional were found to undergo very minor changes from those obtained with B3LYP-D3 functional (Figure S10). The RMSDs of the heavy atoms of the QM regions optimized with M06-2X functional with respect to those with B3LYP-D3 one are 0.10 Å and 0.14 Å for the native protein and the V82T/I84V mutant, respectively. The compact packing of the pyridyl group of Indinavir and Pro81(B) and Val82(B) of the protein observed in the optimized structure with B3LYP-D3 functional was maintained as well in the M06-2X structure (Figure S11).

Free energy calculations of the difference in the binding free energy between the native protein and the mutant, $\Delta_{M-N}\Delta_b F$, based on Eqs. (3) and (4) showed that the binding free energy of

Indinavir increased by 3.8 kcal/mol upon the mutation of V82T/I84V (Table 2), indicating that the mutation significantly reduces the binding affinity of Indinavir. Deviation of the free energy differences obtained by the FEP calculations in the forward direction and the backward ones was small (0.2 kcal/mol) despite that their thermodynamic paths did not coincide as described above, showing a sufficient statistical convergence. The computed increase of the binding free energy is in line with the experimental observations, i.e. increase of the binding free energy by 2.5-3.0 kcal/mol [13,14], although the computed value is slightly overestimated.

As shown in Eq. (3), $\Delta_{M-N}\Delta_b F$ is given by a difference between free energy changes of the inhibitor binding state (IBS) and the inhibitor unbinding state (IUS) upon the mutation, $\Delta_{M-N}F_{IBS}$ and $\Delta_{M-N}F_{IUS}$, respectively. Furthermore, the former is composed of contributions of the QM energy, $\Delta_{M-N(QM)}E_{QM}$, the QM-MM interaction energy and the MM energy, $\Delta_{M-N(QM)}F_{QM-MM,MM}$, and the alchemical mutation, $\Delta_{M-N}F_{alch}$, as shown in Eq. (4). $\Delta_{M-N(QM)}E_{QM}$ and $\Delta_{M-N(QM)}F_{QM-MM,MM}$ represent the contributions of conformational changes of the QM region. The contribution of $\Delta_{M-N(QM)}E_{QM}$ gave an increase of 1.7 kcal/mol, and was almost compensated by that of $\Delta_{M-N(QM)}F_{QM-MM,MM}$, -1.6 kcal/mol. The contribution of the alchemical mutation between IBS and IUS, $\Delta_{M-N}F_{alch} - \Delta_{M-N}F_{IUS}$, then added an increase of 3.7 kcal/mol, and consequently the V82T/I84V mutation increased the overall free energy difference, $\Delta_{M-N}\Delta_b F$, by 3.8 kcal/mol.

The increase of $\Delta_{M-N}\Delta_b F$ is considered to originate from reduction of hydrophobic interaction of Indinavir with the V82T/I84V mutant. It is noteworthy that $\Delta_{M-N}\Delta_b F$ evaluated with the MD simulations with the MM force field of Indinavir was small, 0.3 kcal/mol (Table 2). As described above (Figures 7 and S11), the pyridyl group of Indinavir established a strong hydrophobic interaction with Val82(B) in the QM/MM samples, while the hydrophobic structure was looser in the MM samples. The observation is therefore consistent with the increase of $\Delta_{M-N}\Delta_b F$ by the QM/MM calculation as the hydrophobic interaction becomes weaker in the V82T/I84V mutant where a polar hydroxyl group is introduced at the position 82 and the bulky hydrophobic group at the position 84 is replaced with the smaller one.

Equilibrium MD simulations of IUS for 1 μ s showed large conformational fluctuation of the flap region (Figure S12), although no distinct conformational transitions to the open or semi-open conformations observed by an X-ray crystallographic measurement and previous MD simulations [86-88] were found during the simulation time, which may be a source of the slight overestimation of the increase of $\Delta_{M-N}\Delta_b F$. We also performed the alchemical FEP calculations for IUS with different protonation states of the carboxylic acids of Asp25, i.e., the mono-protonated state and the di-protonated state, for comparison, and found that the difference in the protonation state moderately altered the free energy change upon the mutation by ~ 1 kcal/mol (Table S6).

DISCUSSION AND CONCLUDING REMARKS

The present molecular simulations with the QM/MM free energy optimization and alchemical free energy calculation techniques successfully reproduced the significant reduction of binding affinity of an inhibitor, Indinavir, due to a drug resistant mutation, V82T/I84V, and elucidated the molecular mechanism underlying the affinity change. The complex molecular structures of Indinavir bound in the protein were accurately determined at the density functional level of electronic structure theory on extensive free energy surfaces of the protein represented by the long-time MD simulations on microseconds timescale. The accurate description of the molecular structures of Indinavir successfully identified the molecular origin of the reduction of the binding affinity upon the mutation.

The reduction of the binding affinity upon the mutation was, however, not well reproduced by the calculations only with MM force fields for the Indinavir binding. In fact, significant differences in the internal molecular geometry of Indinavir and its interaction with the protein between the QM/MM free energetically optimized structure and the MM force field one were observed. Especially, the position of Indinavir where the significant geometric difference appeared due to complex electronic interaction of hyperconjugation is close to the protein residues related to the V82T/I84V mutation. It is therefore suggested that the accurate description of the present approach is necessary for properly capturing the impact of the subtle structural changes

due to the drug-resistant mutation. As the present approach allows one to avoid difficulties in accurately developing MM energy functions of complex ligand molecules, the approach is capable of accurately and consistently evaluating relative binding affinities of various inhibitor molecules and drug-resistant mutants required for in silico drug design.

Although the present calculations semi-quantitatively reproduced the reduction of the binding affinity upon the drug-resistant mutation, however, the increase of the free energy difference was slightly overestimated by ~ 1 kcal/mol. Two possible sources of the error are considered. First, the protein conformation in the inhibitor unbound state might not be well modeled as described above. It is, however, noteworthy that when the binding free energy change upon mutations for the inhibitor molecule is compared to that for the native substrate in the catalytic process, through which the ability of the mutations for drug-resistance is evaluated as described above, the ligand unbound state does not appear in the thermodynamic cycle for the free energy comparison. Second, the present QM/MM free energy optimization technique neglects conformational fluctuation of the QM region, which may overestimate stability of the hydrophobic packing. The error is considered to be relatively minor in the present case of the Indinavir binding as tight binding of the inhibitor molecule in the protein was observed. However, for an inhibitor molecule with flexible groups, the lack of conformational fluctuation of the QM region may cause a problem in accurately evaluating the free energies. A correction scheme to take conformational fluctuation of

the QM region into account needs to be developed in the future study.

The protonation state of the sequentially symmetric catalytic carboxyl groups, Asp25(A) and Asp25(B), was also determined by direct calculations of free energy differences between the protonation states. The consistent combination of the high accuracy of the density functional theories for the QM description and the ample conformational samples obtained by the long-time MD simulations for microseconds for the MM protein environments thoroughly determined the protonation states and the underlying mechanism based on the highly accurate and statistically well-converged free energy differences calculated. Especially, the small overall free energy differences between the protonation states (3-5 kcal/mol) were given by sums of large positive and negative contributions of the electronic energies and the interactions with the protein and hydrating water molecules, respectively (Table 1), created by the asymmetrically bound ligand. The high accuracies of both the electronic state of the titratable groups and their interaction with the statistically reorganizing surroundings responding to large changes of electrostatic environment due to the protonation change were therefore important for the determination of the protonation state.

To understand molecular mechanism of the drug-resistance, the impacts of the mutations on both of the enzymatic activity and the inhibitor binding need to be examined as described above. The theoretical investigation on the enzymatic chemical reaction inevitably requires accurate

quantum chemical treatment. The present study demonstrated that the inhibitor binding process can be examined quantitatively by the accurate QM/MM free energy optimization approach that is directly applicable to the catalytic reaction process as well [50-52], and therefore opens the way for understanding the molecular mechanism of drug-resistance through direct quantitative comparison of those processes with the same accuracy, which is now ongoing in our group and will be reported elsewhere.

ACKNOWLEDGEMENT

We thank Taisuke Hasegawa, Cheng Cheng, Yuki Yamamoto, Takafumi Shikakura, Keiei Kumon, Justin Chan, and Shun Sakuraba for valuable discussions and comments. Molecular figures are created with VMD [89]. This work was financially supported by JSPS KAKENHI grant numbers 25104004, 16H04776, 18H05161, 19H03195, 20H05441, and 20H05098, and grants from MEXT as “Priority Issue on Post-K computer” (Building Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems) and “Program for Promoting Researches on the Supercomputer Fugaku” (Application of Molecular Dynamics Simulation to Precision Medicine Using Big Data Integration System for Drug Discovery). Some computations were performed at the Research Center for Computational Science, Okazaki, Japan, and using the computer resource offered under the category of General Projects by Research Institute for Information Technology, Kyushu University. Also, some computations were performed with

TSUBAME3.0 supercomputer provided by Tokyo Institute of Technology through the HPCI System Research Project (Project ID: hp150270).

REFERENCES

- (1) Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N. L.; Nalivaika, E. A.; Özen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A. Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* **2010**, *2*, 2509-2535.
- (2) Wensing, A. M. J.; van Maarseveen, N. M.; Nijhuis, M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antivir. Res.* **2010**, *85*, 59-74.
- (3) Ido, E.; Han, H.; Kezdy, F. J.; Tang, J. Kinetic Studies of Human Immunodeficiency Virus Type 1 Protease and Its Active-site Hydrogen Bond Mutant A28S. *J. Bio. Chem.* **1991**, *266*, 24359-24366.
- (4) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How Does a Symmetric Dimer Recognize an Asymmetric Substrate? A Substrate Complex of HIV-1 Protease. *J. Mol. Biol.* **2000**, *301*, 1207-1220.
- (5) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* **1995**, *3*, 581-590.

- (6) Brik, A.; Wong, C.-H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* **2003**, *1*, 5-14.
- (7) Trova, M. P.; Babine, R. E.; Byrn, R. A.; Casscles, Jr., W. T.; Hastings, R. C. SYNTHESIS AND BIOLOGICAL EVALUATION OF A SERIES OF HIV-1 PROTEASE INHIBITORS. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1595-1600.
- (8) Askin, D.; Eng, K. K.; Rossen, K.; Purick, R. M.; Wells, K. M.; Volante, R. P.; Reider, P. J. Highly Diastereoselective Reaction of a Chiral, Non-Racemic Amide Enolate with (*S*)-Glycidyl Tosylate. Synthesis of the Orally Active HIV-1 Protease Inhibitor L-735,524. *Tetrahedron Lett.* **1994**, *35*, 673-676.
- (9) Vacca, J. P.; Dorsey, B. D.; Schleif, W. A.; Levin, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahy, O. M.; Roth, E.; Sardana, V. V.; Schlabach, A. J.; Graham, P. I.; Condra, J. H.; Gotlib, L.; Holloway, M. K.; Lin, J.; Chen, I-W.; Vastag, K.; Ostovic, D.; Anderson, P. S.; Emini, E. A.; Huff, J. R. L-735,524: An orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4096-4100.
- (10) Dorsey, B. D.; Levin, R. B.; McDaniel, S. L.; Vacca, J. P.; Guare, J. P.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Quintero, J. C.; Lin, J. H.; Chen, I-W.; Holloway, M. K.; Fitzgerald, P. M. D.; Axel, M. G.; Ostovic, D.; Anderson, P. S.; Huff, J. R. L-735,524: The Design of a Potent and Orally Bioavailable HIV Protease Inhibitor. *J. Med. Chem.* **1994**, *37*, 3443-3451.

- (11) Dorsey, B. D.; McDaniel, S. L.; Levin, R. B.; Vacca, J. P.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Lin, J. H.; Chen, I-W.; Holloway, M. K.; Anderson, P. S.; Huff, J. R. SYNEHTIS AND EVALUATION OF PYIDYL ANALOGS OF L-735,524: POTENT HIV-1 PROTEASE INHIBITORS. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2769-2774.
- (12) Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. Crystal Structure at 1.9-Å Resolution of Human Immunodeficiency Virus (HIV) II Protease Complexed with L-735,524, an orally Bioavailable Inhibitor of the HIV Proteases. *J. Biol. Chem.* **1994**, *269*, 26344-26348.
- (13) Chen, Z.; Li, Y.; Schock, H. B.; Hall, D.; Chen, E.; Kuo, L. C. Three-dimensional Structure of a Mutant HIV-1 Protease Displaying Cross-resistance to All Protease Inhibitors in Clinical Trials. *J. Biol. Chem.* **1995**, *270*, 21433-21436.
- (14) Shock, H. B.; Garsky, V. M.; Kuo, L. C. Mutational Anatomy of an HIV-1 Protease Variant Conferring Cross-resistance to Protease Inhibitors in Clinical Trials. *J. Biol. Chem.* **1996**, *271*, 31957-31963.
- (15) Maschera, B.; Darby, G.; Palú, G.; Wright, L. L.; Tisdale, M.; Myers, R.; Blair, E. D.; Furfine, E. S. Human Immunodeficiency Virus. *J. Biol. Chem.* **1996**, *271*, 33231-33235.
- (16) Ridky, T. W.; Kikonyogo, A.; Leis, J. Drug-Resistant HIV-1 Proteases Identify Enzyme Residues Important for Substrate Selection and Catalytic Rate. *Biochemistry* **1998**, *37*, 13835-

13845.

(17) Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W.

Kinetic Characterization and Cross-Resistance Patterns of HIV-1 Protease Mutants Selected under Drug Pressure. *Biochemistry* **1995**, *34*, 9282-9287.

(18) Partaledis, J. A.; Yamaguchi, K.; Tisdale, M.; Blair, E. E.; Falcione, C.; Maschera, B. Myers,

R. E.; Pazhanisamy, S.; Futer, O.; Cullinan, A. B.; Stuver, C. M.; Byrn, R. A.; Livingston, D. J. In

Vitro Selection and Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) Isolates with Reduced Sensitivity to Hydroxyethylamino Sulfonamide Inhibitors of HIV-1 Aspartyl Protease. *J. Virol.* **1995**, *69*, 5228-5235.

(19) Munshi, S.; Chen, Z.; Yan, Y.; Li, Y.; Olsen, D. B.; Schock, H. B.; Galvin, B. B.; Dorsey, B.;

Kuo, L. C. An alternate binding site for the P1-P3 group of a class of potent HIV-1 protease inhibitors as a result of concerted structural change in the 80s loop of the protease. *Acta Cryst.* **2000**, *D56*, 381-388.

(20) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. Structural Implications

of Drug-Resistant Mutants of HIV-1 Protease: High-Resolution Crystal Structures of the Mutant Protease/Substrate Analogue Complexes. *Proteins* **2001**, *43*, 455-464.

(21) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Viability of a Drug-

Resistant Human Immunodeficiency Virus Type 1 Protease Variant: Structural Insights for Better

Antiviral Therapy. *J. Virol.* **2003**, 77, 1306-1315.

(22) Mahalingam, B.; Wang, Y-F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, 271, 1516-1524.

(23) Kovalevsky, A. Y.; Tie, Y.; Liu, F.; Boross, P. I.; Wang, Y-F.; Leshchenko, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Effectiveness of Nonpeptide Clinical Inhibitor TMC-114 on HIV-1 Protease with Highly Drug Resistant Mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, 49, 1379-1387.

(24) Rhee, S-Y.; Taylor, J.; Fessel, W. J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; Shafer, R. W. HIV-1 Protease Mutations and Protease Inhibitor Cross-Resistance. *Antimicrob. Agents Chemother.* **2010**, 54, 4253-4261.

(25) Agniswamy, J.; Louis, J. M.; Roche, J.; Harrison, R. W.; Weber, I. T. Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics. *PLoS ONE* **2016**, 11, e0168616-e0168633.

(26) Baxter, J. D.; Chasanov, W. M.; Adams, J. L. An Update on HIV-1 Protease Inhibitor Resistance. *J. AIDS Clin. Res.* **2016**, 7, 581-587.

(27) Wensing, A. M.; Calvez, V.; Günthard, H. F.; Johnson, V. A.; Paredes, R.; Pillay, D.; Shafer, R. W.; Richman, D. D. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir.*

Med. **2017**, *24*, 132-141.

(28) Hyland, L. J.; Tomaszek, Jr., T. A.; Meek, T. D. Human Immunodeficiency Virus-1 Protease.

2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects To Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454-8463.

(29) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.;

Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C-H.; Weber, P. C.

NMR and X-ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor. *J. Am.*

Chem. Soc. **1994**, *116*, 10791-10782.

(30) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.* **1996**, *3*, 946-950.

(31) Wang, Y-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.;

Kiso, Y.; Torchia, D. A. Solution NMR Evidence That the HIV-1 Protease Catalytic Aspartyl Groups Have Different Ionization States in the Complex Formed with the Asymmetric Drug KNI-

272. *Biochemistry* **1996**, *35*, 9946-9950.

(32) Adachi, M.; Ohhara, T.; Kurihara, K.; Tamada, T.; Honjo, E.; Okazaki, N.; Arai, S.; Shoyama,

Y.; Kimura, K.; Matsumura, H.; Sugiyama, S.; Adachi, H.; Takano, K.; Mori, Y.; Hidaka, K.;

Kimura, T.; Hayashi, Y.; Kiso, Y.; Kuroki, R. Structure of HIV-1 protease in complex with potent

inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc. Natl.*

Acad. Sci. USA **2009**, *106*, 4641-4646.

(33) Weber, I. T.; Waltman, M. J.; Mustyakimov, M.; Blakeley, M. P.; Keen, D. A.; Ghosh, A. K.;

Langan, P.; Kovalevsky, A. Y. Joint X-ray/Neutron Crystallographic Study of HIV-1 Protease with

Clinical inhibitor Amprenavir: Insights for Drug Design. *J. Med. Chem.* **2013**, *56*, 5631-5635.

(34) Gerlits, O.; Wymore, T.; Das, A.; Shen, C-H.; Parks, J. M.; Smith, J. C.; Weiss, K. L.; Keen,

D. A.; Blakeley, M. P.; Louis, J. M.; Langan, P.; Weber, I. T.; Kovalevsky, A. Long-Range

Electrostatics-Induced Two-Proton Transfer Captured by Neutron Crystallography in an Enzyme

Catalytic Site. *Angew. Chem. Int. Ed.* **2016**, *55*, 4924-4927.

(35) Chen, X.; Tropsha, A. Relative binding Free Energies of Peptide Inhibitors of HIV-1

Protease: The Influence of the Active Site Protonation State. *J. Med. Chem.* **1995**, *38*, 42-48.

(36) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Erickson, J. W. Calculation of Relative Binding Free

Energies of Peptide Inhibitors to HIV-1 Protease and Its I84V Mutant. *J. Am. Chem. Soc.* **1998**,

120, 8856-8863.

(37) Rick, S. W.; Topol, I. A.; Erickson, J. W.; Burt, S. K. Molecular mechanisms of resistance:

Free energy calculations of mutation effects on inhibitor binding to HIV-1 protease. *Protein Sci.*

1998, *7*, 1750-1756.

(38) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab Initio Molecular Dynamics-Based

Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem.*

Soc. **2001**, *123*, 8730-8737.

(39) Nam, K-Y.; Chang, B. H.; Han, C. K.; Ahn, S. K.; No, K. T. Investigation of the Protonated State of HIV-1 Protease Active Site. *Bull. Korean Chem. Soc.* **2003**, *24*, 817-823.

(40) Zoete, V.; Michielin, O.; Karplus, M. Protein-ligand binding free energy estimation using molecular mechanics and continuum electrostatics. Application to HIV-1 protease inhibitors. *J. Comput. Aided Mol. Des.* **2003**, *17*, 861-880.

(41) Lepšík, M.; Kříž Z.; Havlas, Z. Efficiency of a Second-Generation HIV-1 Protease Inhibitor Studied by Molecular Dynamics and Absolute Binding Free Energy Calculations. *Proteins: Struct. Funct. Bioinf.* **2004**, *57*, 279-293.

(42) Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. Mechanism of Drug Resistance Due to N88S in CRF01_AE HIV-1 Protease, Analyzed by Molecular Dynamics Simulations. *J. Med. Chem.* **2007**, *50*, 1768-1777.

(43) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study. *J. Chem. Inf. Model.* **2007**, *47*, 1590-1598.

(44) Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate Prediction of Protonation State as a Prerequisite for Reliable MM-PB(GB)SA Binding Free Energy Calculations of HIV-1 Protease Inhibitors. *J. Comput. Chem.* **2008**, *29*, 673-685.

- (45) Chen, J.; Yang, M.; Hu, G.; Shi, S.; Yi, C.; Zhang, Q. Insights into the functional role of protonation states in the HIV-1 protease-BEA369 complex: molecular dynamics simulations and free energy calculations. *J. Mol. Model.* **2009**, *15*, 1245-1252.
- (46) Kar, P.; Knecht, V. Energetic basis for drug resistance of HIV-1 protease mutants against amprenavir. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 215-232.
- (47) Yang, M.; Jiang, X.; Jiang, N. Protonation state and free energy calculation of HIV-1 protease-inhibitor complex based on electrostatic polarization effect. *Mol. Phys.* **2014**, *112*, 1659-1669.
- (48) McGee, Jr., T. D.; Edwards, J.; Roitberg, A. E. pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State. *J. Phys. Chem. B* **2014**, *118*, 12577-12585.
- (49) Bastys, T.; Gapsys, V.; Doncheva, N.; Kaiser, R.; de Groot, B. L.; Kalinina, O. V. Consistent Prediction of Mutation Effect on Drug Binding in HIV-1 Protease Using Alchemical Calculations. *J. Chem. Theory Comput.* **2018**, *14*, 3397-3408.
- (50) Kosugi, T.; Hayashi, S. QM/MM Reweighting Free Energy SCF for Geometry Optimization on Extensive Free Energy Surface of Enzymatic Reaction. *J. Chem. Theory Comput.* **2012**, *8*, 322-334.
- (51) Hayashi, S.; Uchida, Y.; Hasegawa, T.; Higashi, M.; Kosugi, T.; Kamiya, M. QM/MM

Geometry Optimization on Extensive Free-Energy Surfaces for Examination of Enzymatic Reactions and Design of Novel Functional Properties of Proteins. *Annu. Rev. Phys. Chem.* **2017**, *68*, 135-154.

(52) Kosugi, T.; Hayashi, S. Crucial Role of Protein Flexibility in Formation of a Stable Reaction Transition State in an α -Amylase Catalysis. *J. Am. Chem. Soc.* **2012**, *134*, 7045-7055.

(53) Kato, H.; Kamiya, M.; Sugo, S.; Ito, J.; Taniguchi, R.; Orito, A.; Hirata, K.; Inutsuka, A.; Yamanaka, A.; Maturana, A. D.; Ishitani, T.; Sudo Y.; Hayashi, S.; Nureki, O. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nat. Commun.* **2015**, *6*, 7177-7186.

(54) Cheng, C.; Kamiya, M.; Uchida, Y.; Hayashi, S. Molecular Mechanism of Wide Photoabsorption Spectral Shifts of Color Variants of Human Cellular Retinol Binding Protein II. *J. Am. Chem. Soc.* **2015**, *137*, 13362-13370.

(55) Kamiya, M.; Hayashi, S. Photoactivation Intermediates of a G-Protein Coupled Receptor Rhodopsin Investigated by a Hybrid Molecular Simulation. *J. Phys. Chem. B* **2017**, *121*, 3842-3852.

(56) Cheng, C.; Kamiya, M.; Takemoto, M.; Ishitani, R.; Nureki, O.; Yoshida, N.; Hayashi, S. An Atomistic Model of a Precursor State of Light-Induced Channel Opening of Channelrhodopsin. *Biophys. J.* **2018**, *115*, 1281-1291.

- (57) Oda, K.; Nomura, T.; Nakane, T.; Yamashita, K.; Inoue, K.; Ito, S.; Vierock, J.; Hirata, K.; Maturana, A. D.; Katayama, K.; Ikuta, T.; Ishigami, I.; Izume, T.; Umeda, R.; Eguma, R.; Oishi, S.; Kasuya, G.; Kato, T.; Kusakizako, T.; Shihoya, W.; Shimada, H.; Takatsuji, T.; Takemoto, M.; Taniguchi, R.; Tomita, A.; Nakamura, R.; Fukuda, M.; Miyauchi, H.; Lee, Y.; Nango, E.; Tanaka, R.; Tanaka, T.; Sugahara, M.; Kimura, T.; Shimamura, T.; Fujiwara, T.; Yamanaka, Y.; Owada, S.; Joti, Y.; Tono, K.; Ishitani, R.; Hayashi, S.; Kandori, H.; Hegemann, P.; Iwata, S.; Kubo, M.; Nishizawa, T.; Nureki, O. Time-resolved serial femtosecond crystallography reveals early structural changes in channelrhodopsin. *eLife*. **2021**, *10*, e62389-e62419.
- (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. Comp. Chem.* **2004**, *25*, 1157-1174.
- (59) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671-690.
- (60) Steinbrecher, T.; Mobley, D.L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Comp. Chem.* **2007**, *127*, 214108-214120.

(61) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Gravers, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009**, *394*, 747-763.

(62) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* **2012**, *21*, 150-160.

(63) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012**, *137*, 230901-230912.

(64) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Waster, R.; Murcko, M.; Fyfe, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.

(65) Liu, W.; Jia, X.; Wang, M.; Li, P.; Wang, X.; Hu, W.; Zheng, J.; Mei, Y. Calculations of the absolute binding free energies for *Ralstonia solanacearum* lectins bound with methyl- α -l-fucoside at

molecular mechanical and quantum mechanical/molecular mechanical levels. *RSC. Adv.* **2017**, *7*, 38570-38580.

(66) Oshima, H.; Re, S.; Sugita, Y. Prediction of Protein–Ligand Binding Pose and Affinity Using the gREST+FEP Method. *J. Chem. Inf. Model.* **2020**, *60*, 5382-5394.

(67) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee Jr., T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595-5623.

(68) Case, D. A.; Cerutti, D. S.; Cheatham, III, T. E.; Darden T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. **2017**, AMBER 2017, University of California, San Francisco.

- (69) Salmon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory. Comput.* **2013**, *9*, 3878-3888.
- (70) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, Jr J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347-1363.
- (71) Reiling, K., K.; Endres, N. F.; Dauber, D. S.; Craik, C. S.; Stroud, R. M. Anisotropic Dynamics of the JE-2147-HIV Protease Complex: Drug Resistance and Thermodynamic Binding Mode Examined in a 1.09 Å Structure. *Biochemistry* **2002**, *41*, 4582-4594.
- (72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
- (73) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory. Comput.* **2015**, *11*, 3696-3713.
- (74) Joung, I. S.; Cheatham, III, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*,

9020-9041.

(75) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(76) 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.* **1977**, *23*, 327-341.

(77) Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24-34.

(78) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(79) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(80) Grimme, S.; Antony, S.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104-150122.

(81) Hayashi, S.; Ohmine, I. Proton Transfer in Bacteriorhodopsin: Structure, Excitation, IR Spectra, and Potential Energy Surface Analyses by an *ab Initio* QM/MM Method. *J. Phys. Chem. B* **2000**, *104*, 10678-10691.

- (82) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245-268.
- (83) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003**, *91*, 140601-140604.
- (84) Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005**, *122*, 144107-144122.
- (85) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
- (86) Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y.-C.; Tam, K.; Torbett, B. E.; Elder, J. H.; McRee, D. E.; Stout, C. D. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 866-875.
- (87) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **2006**, *103*, 915-920.
- (88) Soares, R. O.; Torres, P. H. M.; da Silva, M. L.; Pascutti, P. G. Unraveling HIV protease flaps

dynamics by Constant pH Molecular Dynamics simulations. *J. Struct. Biol.* **2016**, *195*, 216-226.

(89) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33-38.

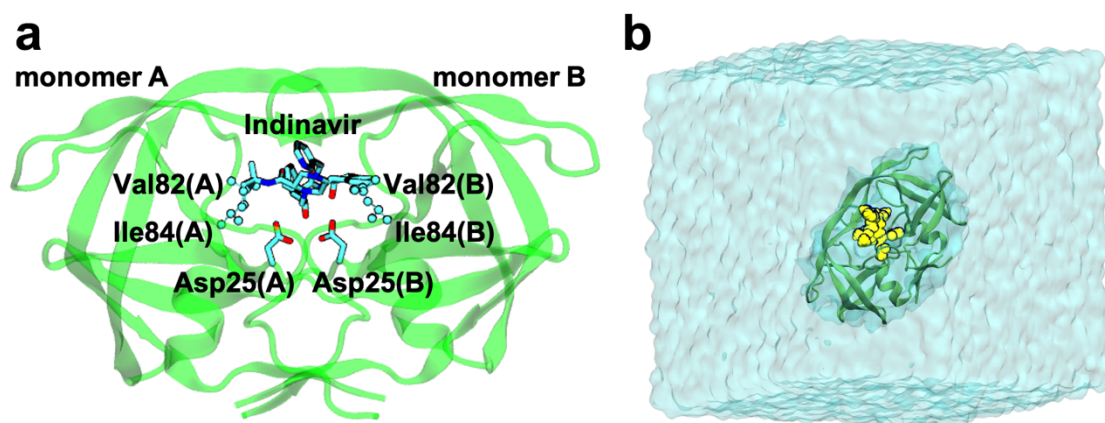


Figure 1. Structure of HIV-1 protease binding Indinavir and its molecular simulation system. **a** Structure of the native HIV-1 protease binding Indinavir. The protein backbone is drawn in a ribbon representation. Heavy atoms of Indinavir and the sidechains of Asp25(A) and Asp25(B) in the monomers A and B, respectively, are depicted in a stick representation. Heavy atoms of the sidechains of Val82 and Ile84 are depicted in a CPK representation. **b** A QM/MM simulation system for the native HIV-1 protease binding Indinavir in a water box in a periodic boundary condition. A quantum mechanically treated molecules are depicted in a van der Waals representation colored in

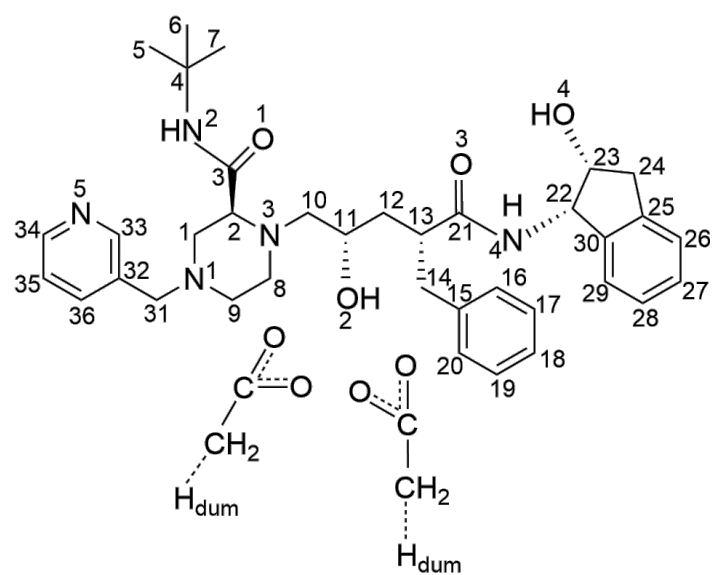


Figure 2. QM regions in the QM/MM RWFE-SCF calculations consisting of Indinavir and the sidechains of Asp25(A) and Asp25(B). Dummy hydrogen atoms to cap the QM regions, H_{dum}, are also shown.

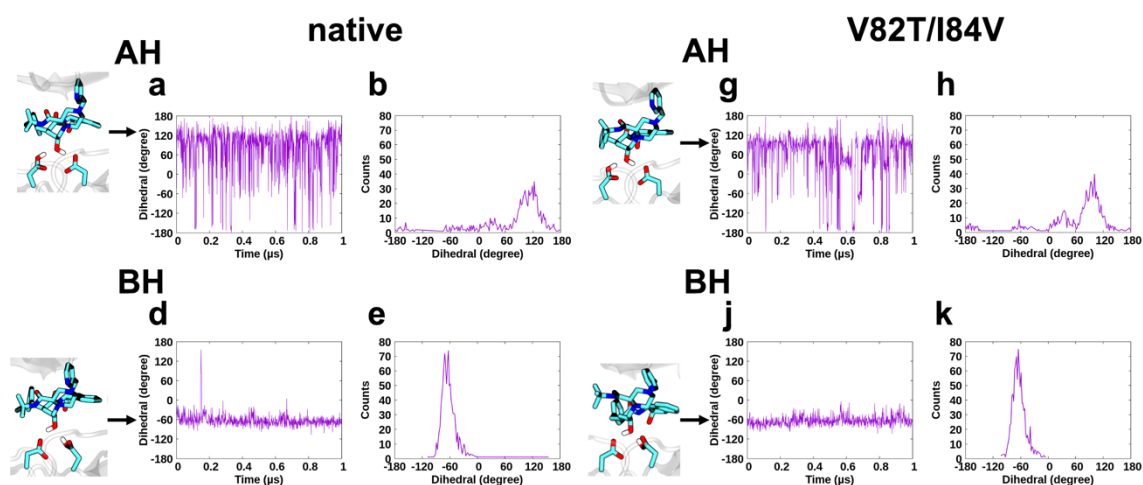


Figure 3. Conformational changes of a dihedral angle $C_{\alpha}-C_{\beta}-C_{\gamma}-O_{\delta}$ of protonated Asp25(A) in AH and Asp25(B) in BH during the 1 μ s equilibrium MD simulations. Temporal changes of the dihedral angle of the native HIV-1 protease in AH (**a**) and in BH (**d**) and the V82T/I84V mutant in AH (**g**) and in BH (**j**), and corresponding distributions of the dihedral angle (**b,e,h,k**), respectively, were shown. A bin width of 2 degrees was employed for the distributions. Heavy atoms of Indinavir and the sidechains of Asp25 are depicted in a stick representation.

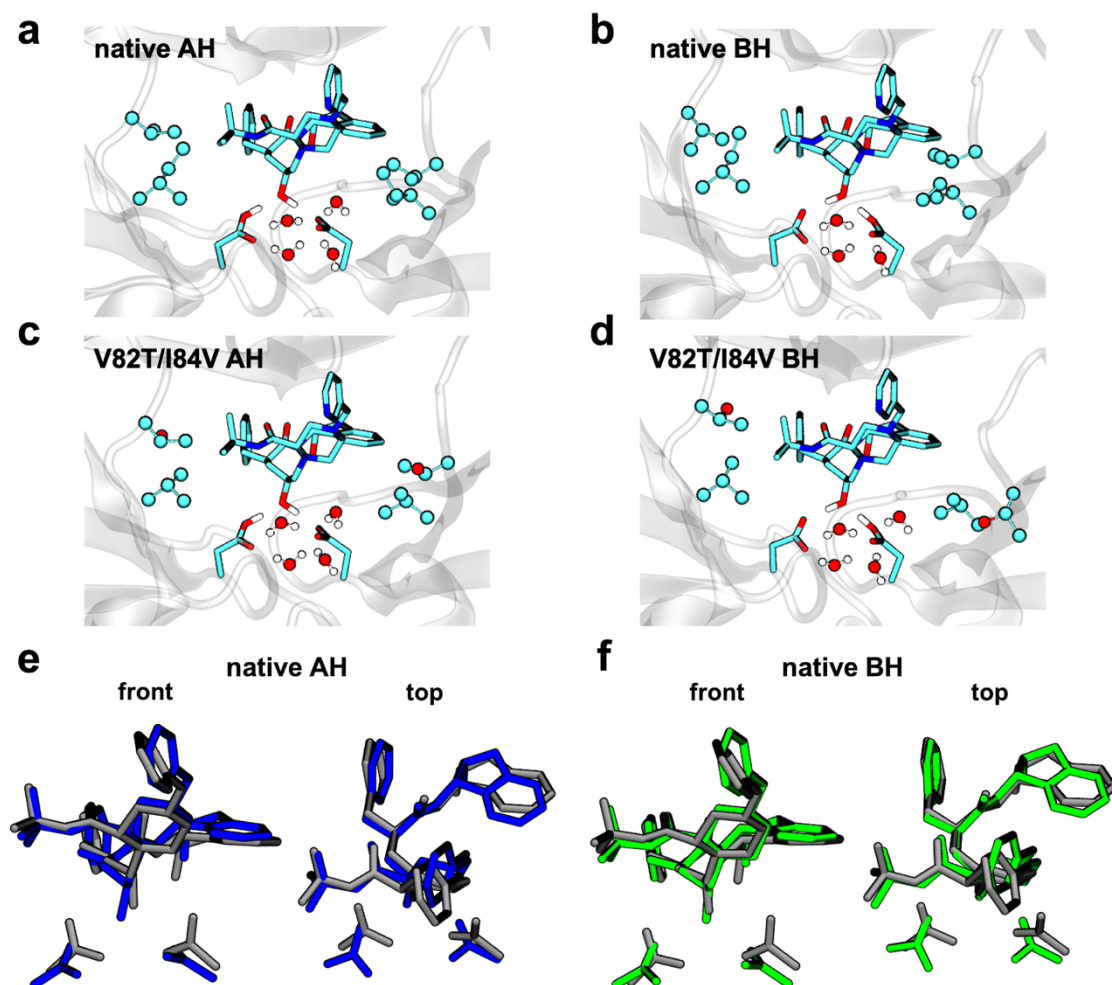


Figure 4. The free energetically optimized structures obtained by the QM/MM RWFE-SCF calculations. **a** The native structure in AH. **b** The native one in BH. **c** The V82T/I84V mutant one in AH. **d** The V82T/I84V mutant one in BH with a fixed O-H distance of protonated Asp25(B). Heavy atoms of Indinavir and the sidechains of Asp25 are depicted in a stick representation. Water molecules within 3 Å of Asp25 and heavy atoms of the sidechains of Val82/Val82Thr and Ile84/Ile84Val are shown in a CPK representation. **e** Comparison between the free energetically optimized native structure in AH (*blue*) and an X-ray crystallographic structure binding Indinavir (PDB ID: 1HSG) (*gray*). **f** Comparison between the free energetically optimized native structure in BH (*green*) and the X-ray one (*gray*). The heavy atoms of the optimized structures were RMSD-fitted to those of the X-ray one.

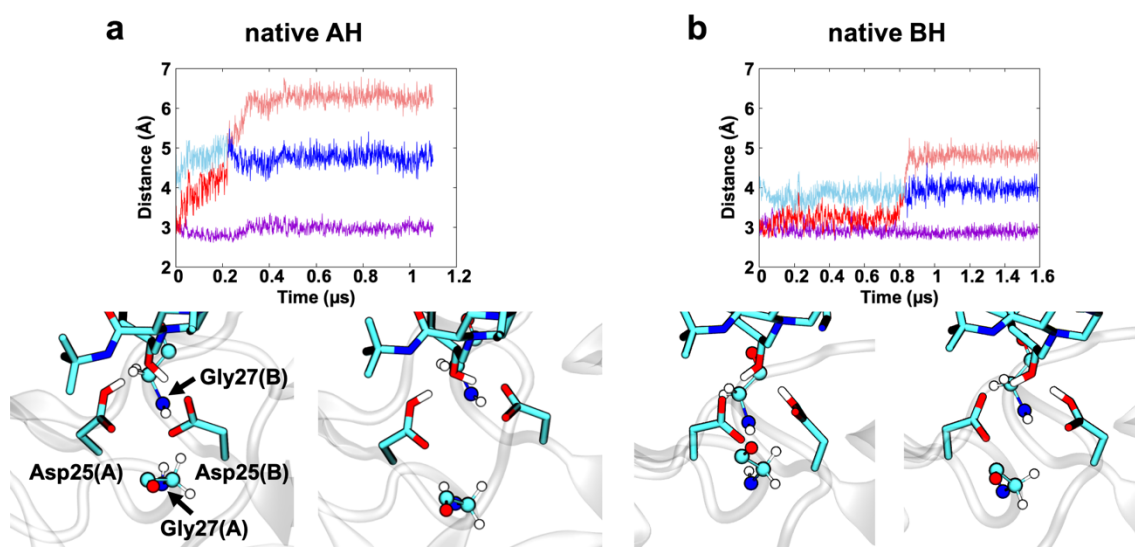


Figure 5. Temporal changes of distances between Asp25(A)/Asp25(B) and Gly27(A)/Gly27(B) during free energy optimizations. **a** The native protein in AH. A distance between O_δ atom of the sidechain of Asp25(B) and N atom of the mainchain of Gly27(B) (*purple*), and those between O_δ atom of the sidechain of Asp25(B) and N atom of the mainchain of Gly27(A) (*red* and *blue*) are depicted. **b** The native protein in BH. A distance between O_δ atom of the sidechain of Asp25(A) and N atom of the mainchain of Gly27(A) (*purple*), and those between O_δ atom of the sidechain of Asp25(A) and N atom of the mainchain of Gly27(B) (*red* and *blue*) are depicted. The distance and the angle including one of two O_δ atoms of Asp25 are colored in *red*, and those including the other are colored in *blue* in **a** and **b**. The distances of the O_δ atoms located farther from N atom of the mainchain of Gly27(A) are drawn in light colors. Bottom-left panels in **a** and **b** represent starting structures of free energy optimizations, respectively, and bottom-right ones represent the free energetically optimized ones. Heavy atoms of Indinavir and the sidechains of Asp25 are depicted in a stick representation. Gly27 are shown in a CPK representation.

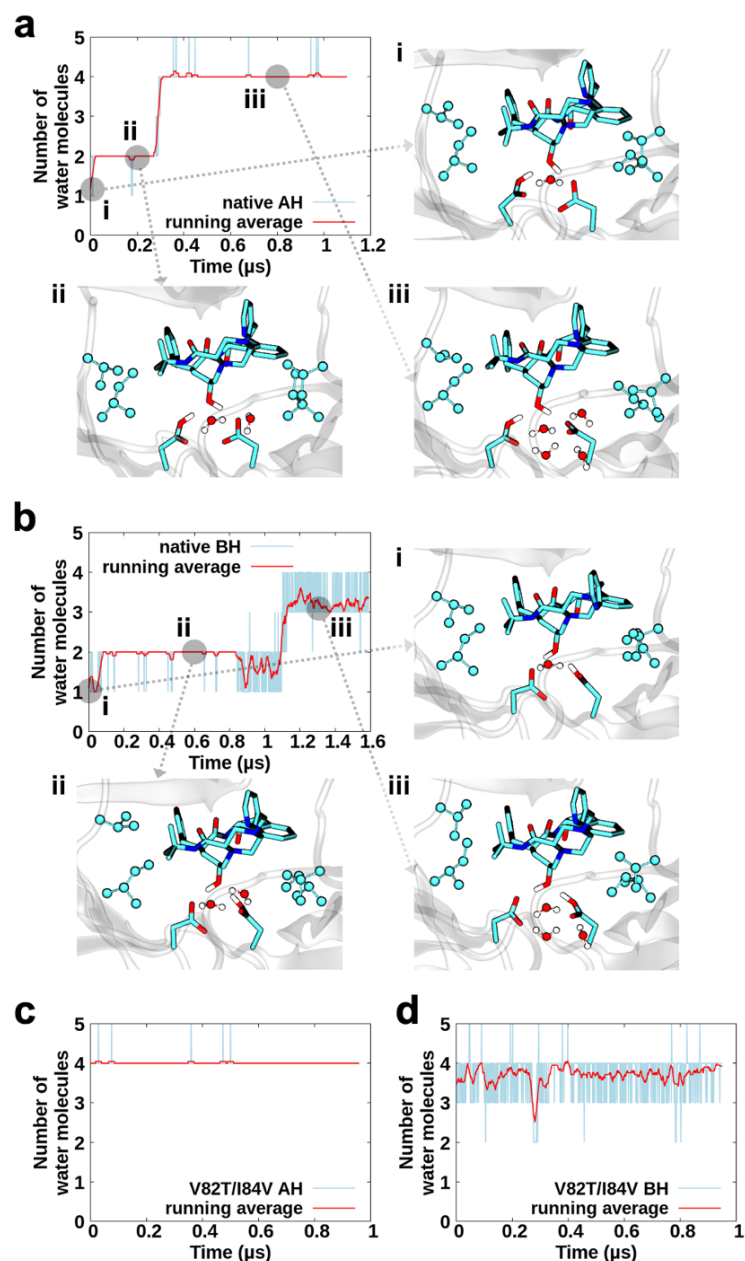


Figure 6. Temporal changes of the number of water molecules around Asp25 during the free energy optimizations. **a** The native protein in AH. **b** The native one in BH. **c** The V82T/I84V mutant in AH. **d** The V82T/I84V mutant in BH with a fixed O-H distance of protonated Asp25(B). Water molecules within 3 Å of Asp25(A) and Asp25(B) were counted. Window-averaged temporal changes are drawn with thick lines in *red*. Width of the averaging window is 20 ns. Snapshot structures of free energy optimizations are also shown in **a** and **b**. Heavy atoms of Indinavir and the sidechains of Asp25 are depicted in a stick representation. Water molecules within 3 Å of Asp25(A) and Asp25(B) and heavy atoms of sidechains of Val82/Val82Thr and Ile84/Ile84Val are shown in a CPK representation.

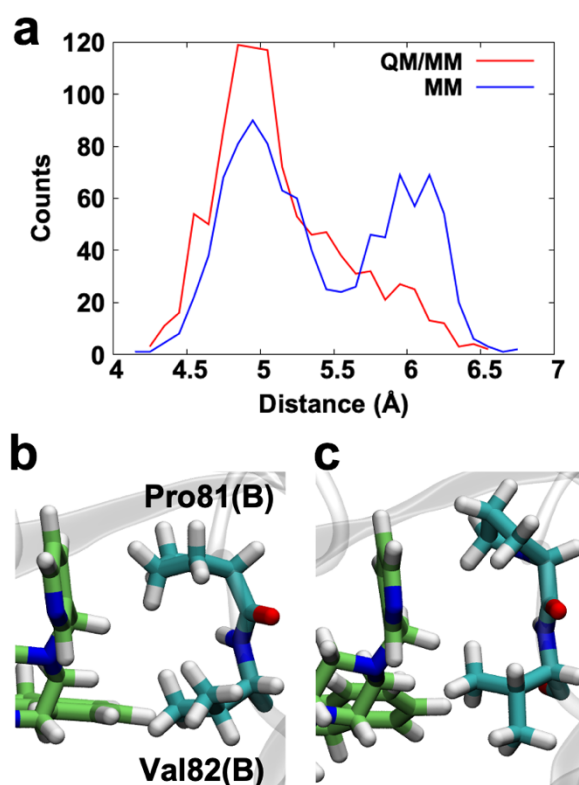


Figure 7. Hydrophobic interactions of the pyridyl group of Indinavir with the sidechains of Pro81(B) and Val82(B). **a** Distributions of distances between C_{γ} atom of Pro81(B) and C_{β} atom of Val82(B) in the QM/MM samples (*red*) and the MM ones (*blue*). The QM/MM samples were taken from a trajectory for 1 μ s obtained by an MD simulation with the QM region of which the geometry and the effective point charges were fixed at the QM/MM free energetically optimized ones with B3LYP-D3 functional. The MM samples were taken from the equilibrium MD simulation for 1 μ s (see COMPUTATIONAL METHODS). A bin width of 0.1 Å was employed for the distributions. **b** A snapshot of the QM/MM samples where the distance in **a** is ~ 5 Å. Indinavir is depicted in *green*. **c** A snapshot of the MM samples where the distance in **a** is ~ 6 Å.

Table 1. Free Energy Differences between Two Mono-protonated States AH and BH, $\Delta_{\text{BH-AH}} F_{\text{QM/MM}}$, and Their Energy Components, $\Delta_{\text{BH-AH}} E_{\text{QM}}$ and $\Delta_{\text{BH-AH}} F_{\text{QM/MM}}$, in kcal/mol.

| | $\Delta_{\text{BH-AH}} E_{\text{QM}}$ | $\Delta_{\text{BH-AH}} F_{\text{QM-MM,MM}}$ | $\Delta_{\text{BH-AH}} F_{\text{QM/MM}}$ |
|-------------------------|---------------------------------------|---|--|
| <i>Native</i> | | | |
| Forward ^a | | 20.7 | 2.6 |
| Backward ^b | -18.1 | 21.1 | 3.0 |
| Average ^c | | 20.9 | 2.8 |
| <i>V82T/I84V mutant</i> | | | |
| Forward ^a | | 18.1 | 4.9 |
| Backward ^b | -13.2 | 17.5 | 4.3 |
| Average ^c | | 17.8 | 4.6 |

^a Energy differences evaluated by forward samplings.

^b Energy differences evaluated by backward samplings.

^c Averages of energy differences evaluated by forward and backward samplings, respectively.

Table 2. Relative Binding Free Energies between the Native Protein and the V82T/I84V Mutant, $\Delta_{M-N}\Delta_b F$, and Their Free Energy Components in kcal/mol.

| QM/MM | $\Delta_{M-N(QM)}E_{QM}$ | $\Delta_{M-N(QM)}F_{QM-MM,MM}$ | $\Delta_{M-N}F_{alchemy}$ | $\Delta_{M-N}F_{IBS}$ | $\Delta_{M-N}F_{IUS}$ | $\Delta_{M-N}\Delta_b F$ |
|-----------------------|--------------------------|--------------------------------|---------------------------|-----------------------|-----------------------|--------------------------|
| Forward ^a | 1.7 | -0.8 | -12.1 | -11.2 | | 3.9 |
| Backward ^b | 1.7 | -2.4 | -10.7 | -11.4 | -15.1 | 3.7 |
| Average ^c | 1.7 | -1.6 | -11.4 | -11.3 | | 3.8 |

| MM | $\Delta_{M-N}F_{IBS}$ | $\Delta_{M-N}F_{IUS}$ | $\Delta_{M-N}\Delta_b F$ |
|-----------------------|-----------------------|-----------------------|--------------------------|
| Forward ^a | -14.4 | | 0.7 |
| Backward ^b | -15.2 | -15.1 | -0.1 |
| Average ^c | -14.8 | | 0.3 |

^a Energy differences evaluated by forward samplings.

^b Energy differences evaluated by backward samplings.

^c Averages of energy differences evaluated by forward and backward samplings, respectively.