# Consistent Prediction of Mutation Effect on Drug Binding in HIV-1 Protease Using Alchemical Calculations

Tomas Bastys,[†,‡] Vytautas Gapsys,[¶] Nadezhda T. Doncheva,[†,§] Rolf Kaiser,[‖] Bert L. de Groot,[¶] and Olga V. Kalinina*[,†]

[†]Department for Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, D-66123 Saarbrücken, Germany

[‡]Saarbrücken Graduate School of Computer Science, University of Saarland, D-66123 Saarbrücken, Germany

[¶]Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany

[§]Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

[‖]Institute for Virology, University Clinic of Cologne, D-50935 Köln, Germany

Ⓢ *Supporting Information*

**ABSTRACT:** Despite a large number of antiretroviral drugs targeting HIV-1 protease for inhibition, mutations in this protein during the course of patient treatment can render them inefficient. This emerging resistance inspired numerous computational studies of the HIV-1 protease aimed at predicting the effect of mutations on drug binding in terms of free binding energy $\Delta G$, as well as in mechanistic terms. In this study, we analyze ten different protease-inhibitor complexes carrying major resistance-associated mutations (RAMs) G48V, I50V, and L90M using molecular dynamics simulations. We demonstrate that alchemical free energy calculations can consistently predict the effect of mutations on drug binding. By explicitly probing different protonation states of the catalytic aspartic dyad, we reveal the importance of the correct choice of protonation state for the accuracy of the result. We also provide insight into how different mutations affect drug binding in their specific ways, with the unifying theme of how all of them affect the crucial drug binding regions of the protease.
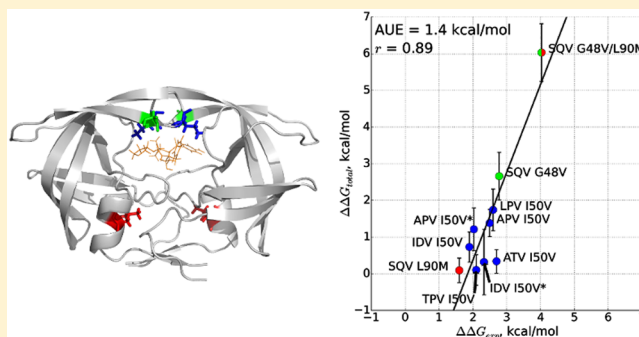
## INTRODUCTION

HIV-1 (human immunodeficiency virus-1, further denoted as HIV) has caused a global epidemic that affects approximately 37 million people worldwide.[1] There is no vaccine or cure available against HIV, but antiretroviral therapy (ART) is recommended for every infected individual[1] to suppress the virus. Success of ART has led to a near-normal life expectancy of HIV infected patients.[2] Nevertheless, resistance toward drugs remains a major issue, making it necessary to switch therapy during the course of treatment of a single patient.[1]

One of the main targets of ART is the HIV protease, a protein responsible for cleaving HIV polyproteins during virus maturation (Figure 1a). HIV protease is a homodimer with two 99 residues long subunits. The major structural components of the binding pocket of this protein are the active site at its bottom, which is followed in sequence by a loop and a short $\beta$-sheet (residues 26−32) that constitute the side of the pocket together with the so-called 80s loop (residues 79−84) (Figure 1a). On the top the pocket is covered by the so-called flap region. In the framework of ART, HIV protease can be inhibited by protease inhibitors (PIs) (Figure 1b), a class of competitive

inhibitors which are currently recommended as second- and third-line ART treatment.[1] Specific substitutions in 13 different positions in the protease are considered to be major resistance-associated mutations (RAMs).[3]

I50V is a RAM toward protease inhibitors Lopinavir (LPV)[4,5] and Darunavir (DRV)[6,7] and has been associated with particularly strong resistance to Amprenavir (APV)[8−10] and correspondingly to its prodrug Fosamprenavir (FPV). It also has been suggested to cause resistance toward Indinavir (IDV),[9,11] and at the same time, there is evidence pointing that this mutation has been associated with sensitivity toward Atazanavir (ATV)[12,13] and Tipranavir (TPV).[14] I50 is located in the protease flap region (Figure 1a). This flap is highly flexible, and its opening is thus thought to control the drug entry to the pocket.[15] Another flap mutation, G48V, as well as a more distant mutation in the hydrophobic core of the protease, L90M, are RAMs conferring resistance to the inhibitor Saquinavir (SQV).[16−20] These three mutations were analyzed in this study.
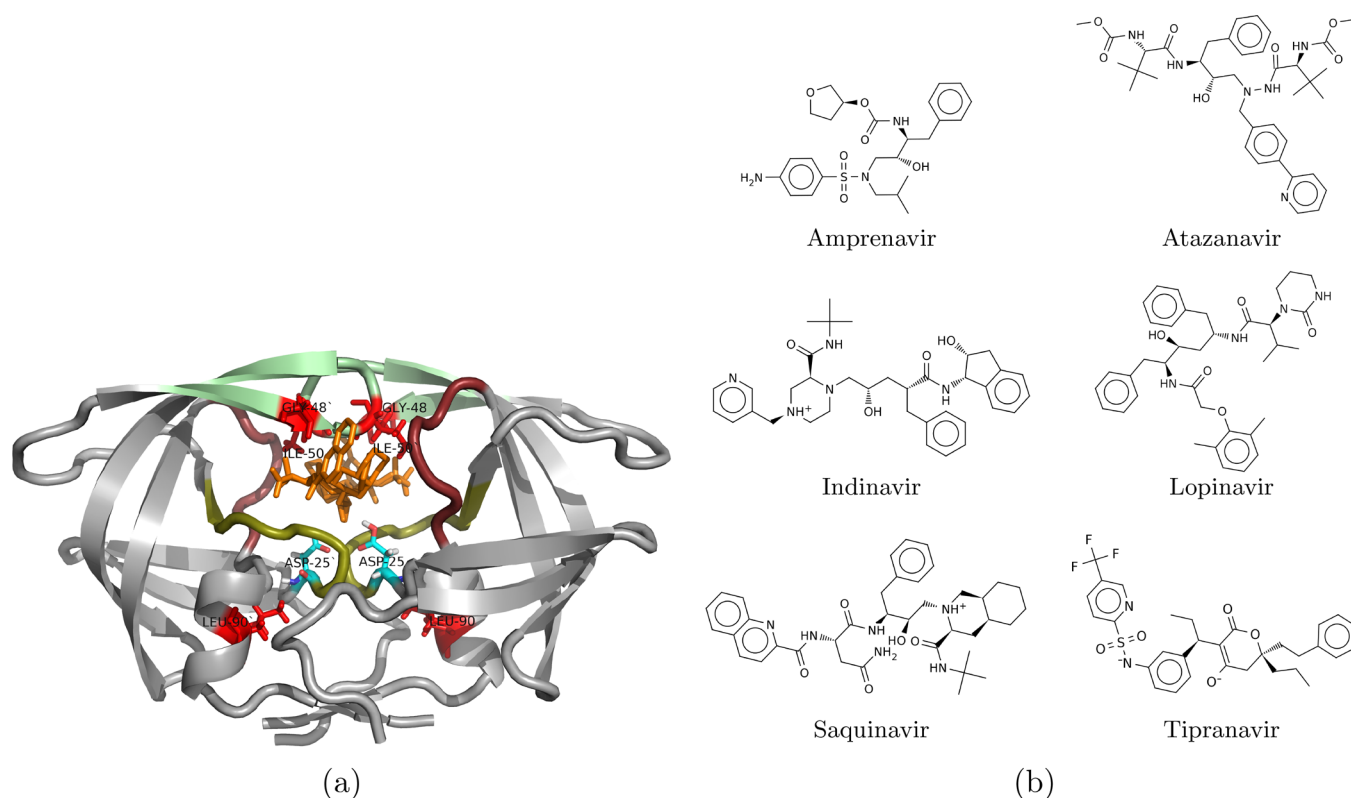
**Figure 1.** (a) HIV protease structure. The flap region is shown in pale green, the 80s loop is shown in brown, and the active-site proximate loop is shown in olive. Mutations which are analyzed in this study (red), and the catalytic site residue (cyan) and bound inhibitor (orange) are shown in the sticks model. (b) Protease inhibitors which are analyzed in this study.

Other RAMs occurring in the flap of the protease are at the positions M46, I47, and I54. RAMs at residues V82 and I84 are placed in the 80s loop, which interacts with the flap region and was suggested to be involved in supporting the closed conformation of protease.[21−23] Other RAM sites located in the binding pocket shared by the substrate and PIs are D30 and V32, which are found in the loop with a short $\beta$-sheet proximal to the active site of the protease. Apart from L90, there are several other locations in the protease hydrophobic core where RAMs occur: this includes L33 and L76. Finally, N88, found in the same $\alpha$-helix as L90, is also a RAM occurrence site. Mutations in all of the resistance-associated mutation sites, with the exception of D30, have been linked with simultaneous resistance to multiple drugs, but the resistance conferred to different drugs is of varying magnitude; likewise different mutations at the same site do not confer resistance to the same drug to the same extent.

The change in the free energy of binding, $\Delta\Delta G$, of inhibitors to the protease upon mutations in it is directly related to the change of the affinity of the inhibitor to the mutant protease and thus to the resistance phenotypes. Hence accurately predicting the effect of the mutation on $\Delta\Delta G$ is highly desirable. Previously, for estimating $\Delta\Delta G$ in HIV protease complexes, methods such as Molecular Mechanics/Poisson−Boltzmann surface area (MM-PBSA) and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) have been used due to their computational efficiency.[24−29] However, these studies are limited in their scope, i.e. in the number of mutations/inhibitors analyzed, or often have poor correlation to experimental estimates. It has been shown that the performance of these methods in terms of reproducing $\Delta G$ for ligand binding to protein varies depending on the system.[30−32]

In this study, we demonstrate that alchemical free energy calculations based on molecular dynamics (MD) simulations allow for an accurate estimation of the impact of mutations on inhibitor binding for a set of different RAMs in combination with different inhibitors. The alchemical mutation of residues is achieved using a nonequilibrium approach,[33,34] which has been demonstrated to yield highly accurate results for amino acid mutations,[35−37] DNA mutations,[38,39] and ligand binding free energies.[40] We refer to other work for a more detailed comparison between the equilibrium and nonequilibrium free energy calculations.[38,41] We show an overall good correlation of the estimated and empirical $\Delta\Delta G$ values and emphasize the importance of the correct choice of the protonation state for the two aspartic acid residues in the active site. We then analyze how the effects of the mutation propagate in the protease structure to affect the inhibitor binding at a site that can be remote from the mutated amino acid residue. Finally, we show that the dynamics of residues corresponding to major RAMs is strongly affected even if the residues are not mutated in a particular viral strain. Thus, we provide novel and general insights into the development of HIV-1 resistance to drugs.

## ■ METHODS

**System Preparation.** Crystal structures of protease-inhibitor complexes were obtained from the Protein Data Bank[42] (PDB IDs 1HPV (APV), 1HXB (SQV), 1SDT (IDV), 2BPX (IDV), 2O4K (ATV), 2O4P (TPV), 2O4S (LPV), and 3NU3 (APV)). Mutation I50V in structures 1HPV, 1SDT, 2BPX, 2O4K, 2O4P, 2O4S, and 3NU3 as well as mutations G48V, L90M, and G48V/L90M in structure 1HXB were introduced using Modeler[43] version 9.12. Additionally in structures 1HPV and 2BPX a

background mutation Q7K was introduced to match the genotype of the reference experimental study. In two cases, we have analyzed viral strains containing additional mutations L33I, L63I, C67A, and C95A, for which we used structures 1SDT and 3NU3. Denoted I50V* in this study, these mutations are introduced to protect the protease against autocatalysis as well as cysteine thiol oxidation; however, this practice is not universally applied.[44] All of the aforementioned mutations were introduced in both monomers of the protease in consistency with data from the experimental studies of HIV protease and HIV infection *in vivo*.[45,46]

To validate our approach to choosing the active site protonation state, the following structures resolved by neutron crystallography were considered: 2ZYE (with a not used clinically inhibitor KNI-272), 5E5J (DRV), 5E5K (DRV), and 5T8H (APV). Hydrogen atoms were assigned to aspartic acid, glutamic acid, and histidine residues as they were resolved in the corresponding structures, with the exception of 5E5J, where D30 was assigned a deprotonated state as the deuterium atom in that structure had 50% occupancy on D30 and K45 residues,[47] and 5E5K, where the E34 was assigned a protonated state as this state was inferred in the original study.[47]

The Gromacs simulation software package was used to set up (version 4.6.2), carry out, and analyze the MD simulations (versions 5.0.2 and 5.1.2).[48,49] All crystallographic water and ion molecules were retained. The $pK_a$ of residues was predicted using Propka,[50] and protease was assigned a monoprotonated state on either D25/D25′, where the prime refers to the second subunit of the protein, with the exception of structures 3NU3 and 1SDT, where previously suggested protonation was used,[51] and 1HXB, where similar $pK_a$ values (6.42 and 6.47) were predicted for both active site residues and protonation of D25 was chosen. The mutant reference protein protonation state was considered to be the same as that of the wildtype protein. The summary of protonated aspartic acids for all the complexes is reported in Table S1. The Amber99SB*-ILDN force field was used for parametrization of the protease. Chemaxon Calculator[52] was used to determine ligand protonation. Ligands ATV, LPV, and APV were assigned a neutral charge, IDV and SQV were assigned a +1 charge, and TPV was assigned a −2 charge. Gaussian09[53] was used to optimize ligand geometry and calculate electrostatic potential at the HF/6-31G* level of theory. Partial charges were assigned by performing a restrained electrostatic potential (RESP) fit.[54] Bonded parameters and atom types were obtained from the Generalized Amber Force Field (GAFF).[55] ACPYPE[56] was used for file conversion into the Gromacs format. The complex was solvated in TIP3P water molecules with 1.4 nm buffer in each dimension with 0.15 mol/L concentration of Cl and Na ions[57] to neutralize the system.

**Equilibrium MD Simulations.** Each system was subjected to energy minimization using the steepest descent algorithm. A maximum of 5000 steps was performed until a maximum force of 1000.0 kJ mol$^{-1}$ nm$^{-1}$ was achieved. Ten replica 200 ns simulations for each complex were performed at 300 K with a time constant of 0.1 ps using the velocity-rescaling thermostat,[58] at a constant pressure of 1 atm with a time constant of 4 ps using the Parrinello−Rahman barostat.[59] Electrostatic interactions were calculated at every step with the particle-mesh Ewald method,[60] and short-range repulsive and attractive dispersion interactions were simultaneously described by a Lennard-Jones potential with a cutoff of 1.2 nm. All bonds were constrained using the LINCS algorithm,[61] and a time step of 2 fs was adopted. Simulated annealing in length of 1 ns was performed for a mutant structure of 2BPX in both protonation states as well as 5E5J when

protonated on D25′ to avoid too close of contact between atoms before equilibrium simulation. For all of the analyses that followed, the first 20 ns of the simulations were considered to be a part of the system equilibration process and thus discarded, with the exception of free energy calculations, where the first 10 ns were discarded.

**Free Energy Calculations.** The protocol for free energy calculations was adjusted from the nonequilibrium simulation approach used in assessing changes in protein thermal stabilities and protein−protein interactions upon amino acid mutation.[36] From each of the equilibrium simulations described above, trajectory frames were extracted equidistantly in time every 10 ns. For every snapshot hybrid structures and topologies were generated using the pmx[62] framework for all of the residues to be mutated. Subsequently, short 20 ps simulations were performed to equilibrate velocities. Finally, alchemical transitions were carried out in 50 ps. During these transitions mutations in both of the monomers were introduced simultaneously, resulting in four mutations for the double mutant G48V/L90M and two mutations for the rest of the complexes. The simulation parameters for the 20 ps equilibration and alchemical transitions were identical to those used in the 200 ns equilibrium simulations. During the transitions, nonbonded interactions were soft-cored.[63] The Crooks Fluctuation Theorem[34] was used to relate the obtained work distributions to the free energy values by employing a maximum likelihood estimator.[64] The error estimates were obtained by the bootstrap approach.

**Partial Least-Squares Regression.** Partial least-squares regression was performed using the functional mode analysis tool.[65] The following sets of input atoms (excluding hydrogens) were used for the model: backbone, ligand, protein, and side chain. Constants 0 and 1 have been used as response variables for trajectories corresponding to mutant and wildtype protein simulations, respectively.

Cross-validation for each mutation and drug combination was done as follows: all trajectories for wildtype and mutant complexes were concatenated, superimposed to minimize the variance over the ensemble,[66] and divided into five equal parts. In each iteration, a model was trained on four parts of labeled input in equal parts from wildtype and mutant simulations, after which it was used to make prediction for the last part. The Pearson correlation between the actual signal and prediction was used to measure the prediction quality. The number of components in the final model has been selected using the so-called "elbow method", where model complexity is increased by boosting the number of components until adding further components only marginally improves the quality of prediction.[67]

**Mutual Information.** Mutual information $I$ between pairs of backbone $\phi$ and $\psi$ as well as side chain $\chi$ dihedral angles of residues was estimated from their individual and joint entropies using the MutInf[68] method. Twenty-four bins were used to get the discrete distributions of the dihedral angles. Dihedrals from 10 simulations for each case were kept separate for later evaluation of $I^{ind}$ for pairs of torsion angles from independent simulations. Monte Carlo sampling with adaptive partitioning of all torsion angle pairs was used to obtain a background distribution of $I$, and only those $I$ entries, which have a $p$-value of at most 0.01 according to this distribution, were retained. False positives were removed based on $P(I < E[I_{ind}])$ criteria, namely that the true mutual information is lower than expected for independent torsion angles, $I_{ind}$. $I^{ind}$, or the excess mutual information, is then subtracted from the actual $I$ so that incomplete sampling resulting from memory effects in the

**Table 1. Experimental and Estimated $\Delta\Delta G$ Values in kcal/mol**

| drug | mutation | $\Delta\Delta G_{exp}$ | $\Delta\Delta G_{calc}$ | $\Delta\Delta G_{WT}^{prot}$ | $\Delta\Delta G_{MUT}^{prot}$ | $\Delta\Delta G_{total}$ |
|------|----------|------------------------|-------------------------|------------------------------|-------------------------------|--------------------------|
| ATV | I50V | 2.7[a] | 0.02 ± 0.19 | −1.80 ± 0.16 | −1.16 ± 0.2 | 0.34 ± 0.32 |
| LPV | I50V | 2.6[a] | 1.39 ± 0.15 | −0.81 ± 0.26 | −0.1 ± 0.48 | 1.74 ± 0.57 |
| TPV | I50V | 2.1[a] | 0.9 ± 0.18 | 1.04 ± 0.3 | −0.56 ± 0.21 | 0.1 ± 0.41 |
| APV | I50V | 2.5[a] | 1.16 ± 0.17 | −1.24 ± 0.17 | −0.81 ± 0.27 | 1.38 ± 0.37 |
| IDV | I50V | 1.9[a] | 0.82 ± 0.12 | 0.91 ± 0.29 | 0.73 ± 0.27 | 0.73 ± 0.41 |
| APV | I50V* | 2.03[b] | 2.11 ± 0.15 | 1.47 ± 0.35 | −0.33 ± 0.45 | 1.21 ± 0.58 |
| IDV | I50V* | 2.33[c] | 0 ± 0.1 | −2.49 ± 0.45 | −1.82 ± 0.76 | 0.32 ± 0.89 |
| SQV | G48V | 2.78[d] | 3 ± 0.58 | 0.67 ± 0.15 | 0 ± 0.24 | 2.66 ± 0.65 |
| SQV | L90M | 1.60[d] | −0.09 ± 0.24 | 0.67 ± 0.15 | 1.03 ± 0.18 | 0.09 ± 0.33 |
| SQV | G48V/L90M | 4.03[d] | 5.32 ± 0.74 | 0.67 ± 0.15 | 2.1 ± 0.21 | 6.03 ± 0.79 |

[a]$\Delta\Delta G_{exp}$ taken from Muzammil et al.[44] [b]$K_i$ values were taken from Shen et al.[78] and $\Delta G_{exp}$ was calculated using $\Delta G_{exp} = -RT \ln K_i$. [c]$K_i$ values were taken from Liu et al.[79] and $\Delta G_{exp}$ was calculated using $\Delta G_{exp} = -RT \ln K_i$. [d]$K_i$ values were taken from Ermolieff et al.[80] and Maschera et al.[81] $\Delta G_{exp_1}$ and $\Delta G_{exp_2}$ were calculated using $\Delta G_{exp_j} = -RT_j \ln K_{i_j}$ and $\Delta G_{exp}$ at 300 K temperature and then calculated by linear interpolation between $\Delta G_{exp_1}$ and $\Delta G_{exp_2}$.

simulations can be corrected for. Mutual information for any two residues, $I_{res}$, is then estimated as the sum of the mutual information between pairs of those residues' torsion angles. For more details on the methods described refer to the original paper.[68]

Bootstrap sets were created for each set, by randomly selecting with replacement $n$ frames from a simulation of length $n$. This was done for each of the 10 simulation runs while preserving the same order across the different dihedral angles and simulation runs, repeating this procedure 10 times. For each resulting bootstrap set, $I$ was estimated. The mean and standard deviation of $I_{res}$ from the 10 bootstrap sets, $\mu_{I_{res}}$ and $sd_{I_{res}}$, were then calculated. When comparing the actual residue pairs, $I_{res_{WT}}$ and $I_{res_{MUT}}$, for wildtype and mutant simulations for some complex, only those cases were retained, where
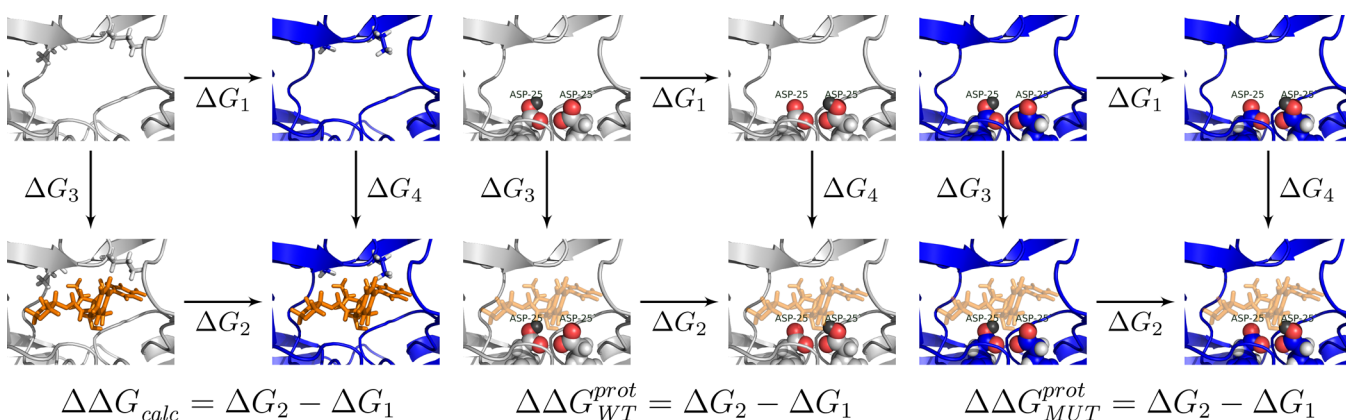
$$|\mu_{I_{res_{MUT}}} - \mu_{I_{res_{WT}}}| > \sqrt{\frac{sd_{I_{res_{MUT}}}^2}{10} + \frac{sd_{I_{res_{WT}}}^2}{10}} \qquad (1)$$

from the bootstrapped sets. As a final filter, residues whose differences in eq 1 were smaller than or equal to 0.3 kT were not considered for further analysis.

## ■ RESULTS

**Estimation of the Lowest Free Energy Protonation State of the Active Site.** The catalytic site of aspartyl proteases comprises two aspartic acid residues, of which only one is protonated in its active form.[69] In HIV protease these residues are D25 and D25′ from the two subunits of the dimer that lie next to each other in the ligand binding pocket (Figure 1a). The three-dimensional structures of HIV protease considered in this study have been obtained by X-ray crystallography, without hydrogen atoms coordinates resolved. Hence it is unknown which of the catalytic aspartate residues is protonated in each case. The apo form of the enzyme is symmetric, thus switching the protonation state of the two aspartic acid residues results in an identical dimer. On the other hand, protease inhibitors considered here are not symmetric molecules, and upon their binding to the protease, the resulting complex is no longer symmetric. Previous studies have suggested that the protonation state of the protease depends on the inhibitor bound.[70−75] For a complex with an experimental inhibitor including a diol group, a diprotonated HIV protease at its active site has been suggested,[76] but such a state seems to be rather an exception.[77]

Here we consider both monoprotonated active site states in order to find the complex with the lowest free energy. In each case studied, we conducted MD simulations in both alternative protonation states (DH25/D25′ and D25/DH25′) for the wildtype and the mutant. The reference protonation state was identified with Propka,[50] with the exception of complexes analyzed for I50V*, where previously suggested protonation was used.[51] Further we estimated the difference of the free binding energy in the wildtype and the mutant complexes between the two alternative protonation states, $\Delta\Delta G_{WT}^{prot}$ and $\Delta\Delta G_{MUT}^{prot}$ (Table 1, Figure 2). Large differences in the free binding energy between the two protonation state alternatives can be observed in several cases, e.g. −2.49 kcal/mol for IDV bound to the wildtype protease. Interestingly, the reference protonation state corresponds to the lowest free energy complex only in 44% (8 out of 18) of the cases. For some inhibitor/mutant combinations, the lowest free energy complex for the mutant has the opposite protonation state compared to the wildtype, i.e., mutation affects the protonation probabilities upon inhibitor binding. Concomitantly, the interactions in the vicinity of the active site differ between the wildtype and the mutant enzymes, as demonstrated by the interaction energy analysis below. Supporting our assertion, an experimental structure of HIV-1 protease in complex with APV resolved with X-ray/neutron crystallography[82] had the same sequence as the wildtype in our analysis of I50V* mutation. The hydrogen atoms on its aspartic dyad were resolved in the study, and the protonated aspartic acid residue matched the lowest free energy prediction for both complexes with APV in our study. To further test the accuracy of our method to predict the energetically favorable active site protonation state, we searched the Protein Data Bank for structures of HIV protease with experimentally resolved hydrogen atoms. This resulted in a data set of four additional protease complexes, including one with APV, two complexes with DRV resolved under different pH conditions, and one with an experimental inhibitor KNI-272. In all cases, the free energy change $\Delta\Delta G^{prot}$ of switching the proton from active site aspartic acid, which was resolved experimentally, to the active site aspartic acid on the opposite monomer was estimated to be energetically unfavorable (Table S2). This included the two complexes with DRV, where the only manifestation of differences in experimental pH values in our simulations was in terms of different protonation states of several residues. Here we were able to correctly predict the opposite active site protonation states of these complexes.

$$\Delta\Delta G_{calc} = \Delta G_2 - \Delta G_1 \qquad \Delta\Delta G_{WT}^{prot} = \Delta G_2 - \Delta G_1 \qquad \Delta\Delta G_{MUT}^{prot} = \Delta G_2 - \Delta G_1$$

**Figure 2.** Thermodynamic cycles for $\Delta\Delta G$ estimation. The first cycle estimates the mutation effect on ligand binding, the second cycle estimates the effect of switching the proton between D25/D25′ for the wildype protein, and the third cycle switches correspondingly for the mutant protein. Only the binding pocket is shown in the figures; the wildtype protein is shown in gray, the mutant is shown in blue, and the inhibitor is shown in orange. In the second and third cycles the inhibitor is set to semitransparent for the sake of clarity, and the proton in question for D25/D25′ is shown as a black sphere.

Taking into account the alternative protonation states of the active site, we measured the total difference of the free binding energy between the wildtype and the mutant complex as

$$\Delta\Delta G_{total} = \Delta\Delta G_{calc} + \frac{1}{\beta}\ln(1 + e^{-\beta\Delta G_{WT}^{prot}}) - \frac{1}{\beta}\ln(1 + e^{-\beta\Delta G_{MUT}^{prot}})$$

(2)

where $\beta$ is the inverse of the product of the gas constant and the temperature. Considering the alternative, protonation states result in correlation between $\Delta\Delta G_{exp}$ and $\Delta\Delta G_{total}$ of 0.89 (average unsigned error 1.4 kcal/mol) (Figure 3).
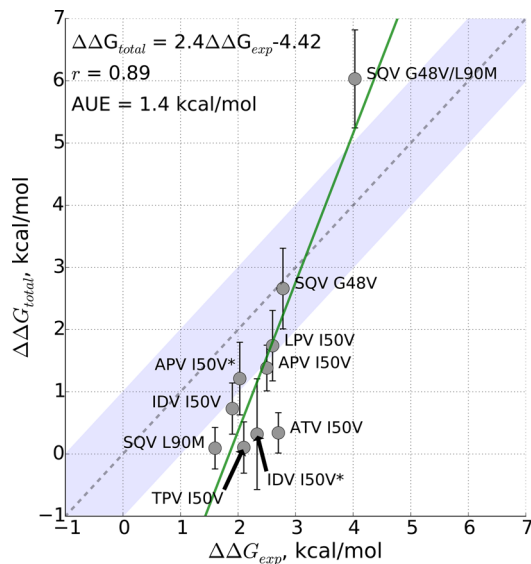


**Figure 3.** Calculated total estimate and experimental measurement of free binding energy binding change upon mutation.

It must be noted, that for many of the complexes studied, there is a compensation of the changes of the free binding energy caused by switching the active site protonation state between the complexes. That is, if simulation of a wildtype protein in the reference state was in an energetically unfavorable protonation state, the corresponding mutant protein also had a tendency to be in an energetically unfavorable protonation state, which means they compensated each other to some degree in the overall estimate of $\Delta\Delta G$ (Figure S1a). This results in a high degree of

correlation between $\Delta\Delta G_{total}$ and $\Delta\Delta G_{calc}$ (Figure S1b). Thus, we wanted to evaluate how accurate can one expect to be in estimating $\Delta\Delta G$ if one were to choose randomly, which of the active site residues is protonated in both wildtype and mutant protein. For that purpose we probed all unique combinations of $\Delta\Delta G$ estimates for the set of all ten complexes, where $\Delta\Delta G$ was a result of any possible linear combination of $\Delta\Delta G_{calc}$ with $\Delta\Delta G_{WT}^{prot}$ and $\Delta\Delta G_{MUT}^{prot}$. For each of these sets we calculated its correlation $r$ with $\Delta\Delta G_{exp}$, which resulted in $4^{10}$ data points (Figure S1c). Correlation between $\Delta\Delta G_{exp}$ and $\Delta\Delta G_{total}$ ($r = 0.89$) was higher than with the mean $\Delta\Delta G$ from all possible combinations ($r = 0.81$), with $\Delta\Delta G_{calc}$ ($r = 0.79$) or $\Delta\Delta G_{propka}$ ($r = 0.81$) (the latter corresponded to the evaluation of the most likely protonation state as predicted by Propka separately for wildtype and mutant protease complexes). This confirms the benefit of combining simulations of HIV protease in both active site protonation states for best accuracy in estimating the effect of mutation on drug binding free energy.

**Energetic Contributions of Individual Residues.** To elucidate energetic interactions of individual residues with the inhibitor, for each of the residues we measured the contributions of Lennard-Jones and short-range electrostatic terms and compared them between the wildtype and the mutant complexes (Figure 2). In all cases and further throughout the analyses, we considered only the complexes with the protonation states corresponding to the lowest free energy.

For the inhibitors ATV, LPV, TPV, and APV, the mutation I50V leads to an increase in the interaction energy between the inhibitor and the protein (Figure 4), which may account for the observed resistance in this phenotype. The experimental measurements from isothermal titration calorimetry (ITC) for ATV, LPV, and APV reported an ethalpic $\Delta\Delta H$ penalty, which also includes the direct ligand-protein interactions, caused by I50V mutation.[44] This was not the case for TPV, where a reduction in $\Delta\Delta H$ was reported for the same mutation.[44] Mutation A71V, which tends to appear together with I50V,[12] is known to compensate for the loss of viral fitness due to primary drug resistance-associated mutation.[83] An ITC study of I50V+A71V double mutant[12] suggested an increase in affinity toward ATV as a result of an increase in entropy which compensated the increase in enthalpy. The same compensatory effect in that study was observed for APV but with the overall result of a binding penalty for the drug. In our study the mutation I50V* decreased the
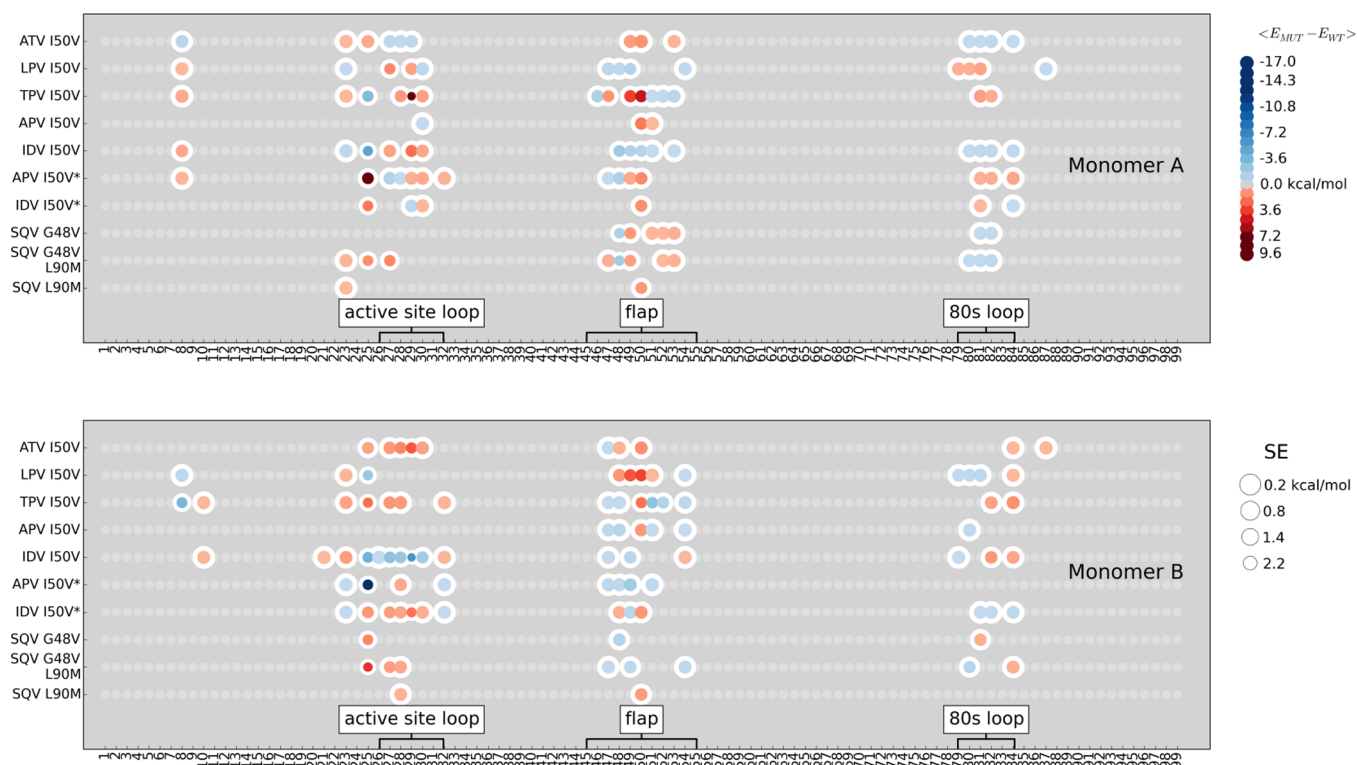
**Figure 4.** Energy differences of nonbonded interaction between protein and inhibitor in wildtype and mutant complexes. Residues, for which the difference ($E_{MUT}-E_{WT}$) between the wildtype and the mutant complexes is higher than the propagated error (SE) and its absolute value is higher than 0.1 kcal/mol, are represented as a colored circle, where the color represents the relative interaction energy and the size of the circle relates inversely to the standard error of the estimate.

direct interaction energy between the protein and APV. This was mostly the result of a stronger interaction with the active site residues due to a difference in protonation state preference between the wildtype and mutant. To test whether the addition of mutation A71V results in a different thermodynamic profile, we performed additional simulations of a 50V+71V mutant with the ATV and APV inhibitors. The resulting $\Delta\Delta G_{total}$, 0.91 and 1.27 kcal/mol, respectively, was close to the original estimates for single mutation I50V for these inhibitors, and it exhibited comparable direct protein-drug interaction profiles (Figure S2). Thus, we could not reproduce the compensating effect of the double mutation in our simulation. Similarly to APV, I50V and I50V* showed differing protein interaction profiles with IDV, with the former exhibiting stronger direct interactions and the latter having a small penalizing effect on interactions after mutation (Figures 4 and S2). In combination with SQV, mutations G48V, L90M, and G48V/L90M had an unfavorable effect on direct inhibitor-protein interaction energy, albeit to a smaller extent for the single-site mutations (Figure 4).

For all of the complexes with the mutation I50V analyzed, if the substitution had any notable effect on the direct interaction energy between protease and inhibitor, such an effect extended beyond the mutated residue itself. The largest contributions were observed for residues in the flap region, i.e. near the mutation site. Other proximal affected residues were in the 80s loop, as well as residues around the active site (Figure 4). Given such a wide distribution of mutation effects, we were interested whether this is related to structural changes in the protease or a different pose of the inhibitor. Partial least-squares (PLS) regression based functional mode analysis,[65] a supervised machine learning technique, allows for correlation of Cartesian input from MD trajectories to a desired functional property. In our case we

investigated whether the major collective motions discriminate wildtype and mutant complexes: using Cartesian coordinates as an input, we aimed to predict whether a trajectory was generated by a wildtype or mutant protein. We applied this technique with different input features: coordinates of heavy atoms of the backbone, ligand, protein, and side chain, to create statistical models. Of these, models with protein backbone input proved to be the most predictive, and models with ligand input proved to be the least predictive in the cross-validation on average (data not shown). Models trained on protein backbone atoms proved to be capable of making a satisfactory distinction for all complexes except IDV in the context of I50V* and complexes with SQV in the context of the G48V or L90M mutations (Figure S3). Mapping the regions of the backbone contributing the most to these PLS models confirmed the flap as the region that is the most important for distinguishing the corresponding trajectories of wildtype and mutant complexes (Figures 5 and S4). The 80s loop is also prone to assume different conformations between wildtype and mutant complexes, as well as the loop with the active site, particularly so for cases in which the energetically favorable protonation state differs between wildtype and mutant proteins (Figures 5a and 5b).

**Changes in the Correlated Motion of Amino Acid Residues.** The analyzed mutations seem to have an effect on direct protein-inhibitor interaction as well as protein backbone rearrangements outside of the immediate mutation site. Certainly, this must be the case for RAMs which are not in direct contact with the inhibitor, such as the mutation L90M. However, we did not observe any consistent effect of this mutation in terms of changes in the protein backbone and just a minor effect in protein-inhibitor interaction. To gain insight into other nonlocal effects of the mutations, we inspected differences in the

(a) I50V TPV                                           (b) I50V* APV
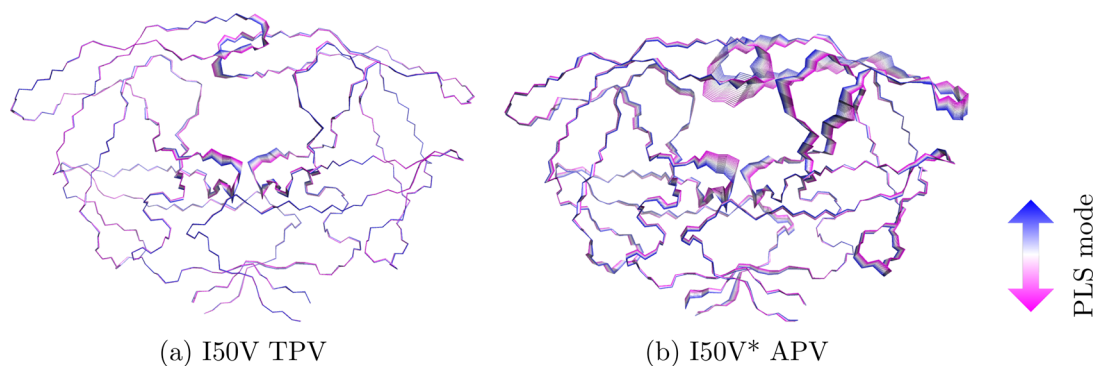
**Figure 5.** Interpolation between the extremes of the PLS models for the corresponding complexes. Blue-to-magenta bands correspond to the interpolation along the mode which relates the true label of simulation, wildtype or mutant, to the underlying differences in protein motions.

correlated motion of all pairs of residues between the wildtype and the mutant protease-inhibitor complexes. Since these correlations are not necessarily linear, we chose to evaluate the pairwise mutual information based on the distributions of dihedral angles from molecular dynamics simulations,[68] a measure that depends both on individual Shannon entropies of the residues and their joint entropy.

First we inspected in which protein regions correlations are affected most by ligand binding. For this purpose we compared the residues' mutual information between apo and holo wildtype proteins (Figure S5). The flexible flap region for all of the complexes is affected to the largest extent. Correlations in the active site loop of the protease as well as the 80s loop also seem to be affected in many cases, particularly so upon binding of TPV and SQV. These results are expected, since the flap and 80s loop regions are suggested to be involved in ligand binding,[15,21−23] while the active site D25/D25′ typically makes hydrogen bonds with the inhibitor upon its binding, thus likely affecting its interactions with other residues in the protein.

Next we estimated the mutual information networks for holo protein complexes with mutations and compared them with wildtype holo protein complexes to evaluate how the mutation affects residues' correlated motions in the protein (Figure 6). We wanted to investigate what is the overlap between such pairs with those whose correlations were affected by ligand binding in the wildtype complexes. To measure the overlap between these sets, we used the Jaccard index, which is the ratio of the size of the intersection of the two sets and the size of their union. A Jaccard index close to 1 indicates that there is a high overlap between the sets of such pairs, whereas an index close to 0 would indicate that different residue pairs show correlated movements in the processes related to binding and propagation of mutation effects. An average Jaccard index value of 0.28 indicates that overall similar residue pairs are involved in these processes, but some differences exist (Table S3). In contrast, when choosing random pairs of sets of the same respective sizes as in the original data, an average Jaccard index value is 0.003. More generally, upon closer inspection we observe that while the identity of the residues in these pairs may differ, they are located in the same protein regions: the flaps, the 80s loop, and the active site loop (Figure 6 and Figure S5).
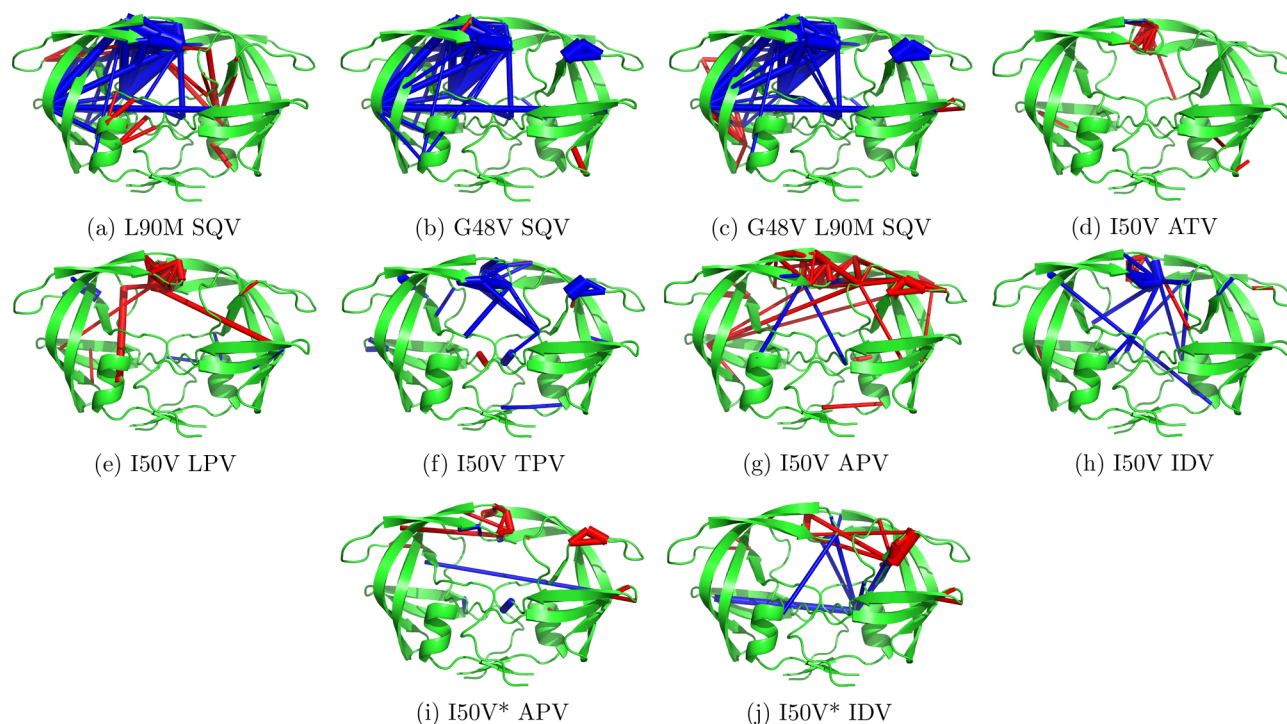
Finally we analyzed in detail the effect of mutations on the protein mutual information network. For the mutation L90M, the correlated motion of the mutation site with the active site residues D25/D25′ is changed. The active site residues directly interact with the inhibitor (Figure 6a), suggesting a path by which the mutation L90M, which is not located in the inhibitor

binding pocket, affects its binding. The changed interaction pattern between L90M and the active site residues has been observed previously in crystal structures[84] and MD simulations.[85] It has also been noticed that the side chain I84/I84′, which is typically in direct contact with D25/D25′, is oriented differently upon the mutation L90M,[85] suggesting changes in the binding pocket of the protease. Indeed, in our analysis, I84 demonstrates significant differences in the correlated motion with multiple residues in the binding pocket, as reflected by high values of the mutual information, for all three G48V, G48V/L90M, and L90M mutants in complex with SQV. These differences are possibly the result of different $\chi$ angle distributions of I84 between wildtype and mutant complexes (Figure S6). We also observe strong differences in mutual information involving the flap region of the complexes with SQV (Figures 6a-6c) and to a smaller extent in other complexes (Figures 6d−6j). Suspecting that such large differences are related to large structural rearrangements in the protease upon mutation, we performed a principal component analysis on the backbone of the complexes involving SQV. The motion along the first principal component, which accounts for most of the structural variance, indeed showed the largest shifts in the 80s loop and the flap region (Figure S7).

## ■ DISCUSSION

In this study, we have demonstrated the applicability of alchemical free energy calculations for an accurate estimation of the change of free binding energy, $\Delta\Delta G$, for different mutations in the HIV-1 protease and different inhibitors. Taking into account alternative protonation states helped to achieve an improvement in correlation with experimental $\Delta\Delta G$ estimates, compared to choosing one state based on empirical $pK_a$ calculations with Propka.[50] A recent study of protonation of aspartic proteases, including HIV-1 protease, also suggested inaccuracy of prediction of the protonation state using Propka.[86] Although the importance of choosing the correct protonation state has been noted in multiple studies,[29,70−75] predictions based on empirical $pK_a$ estimates, as well as setting the protonation state based on a previous suggestion for the protein-inhibitor complex, regardless of the specific sequence context, still remain standard practice. We suggest that explicit probing of both protonation states is needed to reproduce the correct ensemble. We observe that choosing an alternative protonation state can contribute more than 2 kcal/mol to the change in free energy of the system. We also show that a point mutation can change the preferred protonation state of the protease, contrary to the often held assumption.[73]

The effect of the I50V* mutation on the binding free energy of APV and IDV has recently been addressed by Duan et al.[51]

(a) L90M SQV     (b) G48V SQV     (c) G48V L90M SQV     (d) I50V ATV

(e) I50V LPV     (f) I50V TPV     (g) I50V APV     (h) I50V IDV

(i) I50V* APV     (j) I50V* IDV

**Figure 6.** Mutual information mapped onto the protease structure. Cylinders connecting residues represent differences in mutual information between those residues in mutant and wildtype simulations, with the width of the cylinder proportional to the difference and red indicating higher correlation in mutant and blue indicating higher correlation in the wildtype protein. This corresponds to the degree to which residue pairs exhibit differences in the correlation of their motions.

**Table 2. $p$-Values for Fisher's Exact Test for over-Representation in Holo Protein of RAM Sites Amongst Residues Showing Different Correlations with Other Residues upon Mutation or Differences in Direct Interaction Energies with Inhibitor[a]**

| drug and mutation combination | ATV I50V | LPV I50V | TPV I50V | APV I50V | IDV I50V | APV I50V* | IDV I50V* | SQV G48V | SQV G48V/L90M | SQV L90M |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$-value | 0.12 | 0.05 | 0.02 | $2 \times 10^{-3}$ | 0.2 | $1 \times 10^{-3}$ | 0.07 | 0.06 | 0.05 | $3 \times 10^{-5}$ |

[a]The absolute value is larger than 0.1 kcal/mol.

using the MM/PBSA approach. In their study, the authors could not reproduce the experimental values when sampling from 20 ns long MD simulations using the AMBER03[87] force field (Table S4). In the present study, we obtained correct trends in free energy changes for these cases and could also provide a hypothesis for the mechanism of resistance. To compare our data with the results of Duan et al., we used snapshots from the first 20 ns of the trajectories to calculate $\Delta\Delta G_{total}$ for all of our complexes, which resulted in a correlation of $-0.11$. Hence we can suggest that insufficient sampling might have been one of the issues contributing to the inaccuracies observed by Duan et al.

An unfavorable effect of mutation on inhibitor binding has been correctly predicted for all of the cases analyzed. Of these, our predictions of $\Delta\Delta G$ were less accurate for complexes involving ATV and TPV for the mutation I50V, as well as to a lesser extent for IDV for the mutation I50V and for SQV in combination with the mutation L90M, underestimating the complex destabilization of mutation. The mutation I50V is however not one of the major resistance associated mutations for either ATV, TPV, or IDV inhibitors.

In our study, the mutation I50V directly affects interaction energy between protein and inhibitor, and the resulting impact on the enthalpy can be unfavorable (APV, ATV, LPV, and TPV) and favorable (IDV). In contrast to complexes involving the mutation I50V, we do not observe large effect on direct

interactions between the protease and SQV for single-site mutations. These mutations as well as their combination, however, perturb the correlated motions in the protein, both in the mutation site and beyond: this seems to be caused to a large extent by the movement of the protease flap region as well as the 80s loop, which is particularly evident for simulations of the complexes with the resistance-causing mutations.

The resistance-associated mutations, which were studied in this work, showed a tendency to affect the same regions in terms of backbone arrangements, residues interactions with inhibitor, or correlated motions with other residues. These regions, such as the protease flap, 80s loop, and loop proximal to the active site, harbor other sites in which mutations are associated with resistance toward protease inhibitors. In fact, major RAM sites are over-represented among residues whose interaction or dynamics is changed in our simulation upon mutation, particularly in cases when the modeled mutation is a major RAM itself (Table 2). This hints at a possible relationship between residues in the protease, mutations of which are associated with viral resistance toward the drug, and suggests their collective involvement in the process of drug binding.

## ■ CONCLUSIONS

In summary, we have analyzed a data set of ten complexes of HIV-1 protease with inhibitors to assess the effect of resistance-associated mutations on the energetic properties of the

complexes and the differences in inhibitor binding. Our predictions of the effect of those mutations on the free binding energy of the ligand binding correlates well with the experimental measurements for these mutations. We observed that explicit probing of alternative protonation states of the active site of the protease contributes to this accuracy, in addition to enabling us to select the most likely protonation state of the complex. In terms of mechanisms as to how mutation contributed to resistance, we witnessed that in most cases mutation has a negative effect on direct protein-inhibitor interactions. More generally, we found that for all the investigated protein–ligand complexes the mutations consistently affected the same regions in the HIV protease: the flaps, 80s loop, and active site loop.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b01109.

> Residues interaction energies with inhibitors and residues list with differences in MI or energetic interactions with inhibitor (XLS)
>
> Summary of reference protonation states of aspartic acids in this study (Table S1), estimation of free energy change upon switching the proton between active site aspartic acid residues for structures with experimentally resolved hydrogen atoms (Table S2), plots $\Delta\Delta G_{WT}^{prot}$ vs $\Delta\Delta G_{MUT}^{prot}$, $\Delta\Delta G_{total}$ vs $\Delta\Delta G_{calc}$, and correlations of $\Delta\Delta G$ with all possible protonation states with $\Delta\Delta G_{exp}$ (Figure S1), energy differences of nonbonded interaction between protein and inhibitor in barplots representation (Figure S2), Jaccard indexes for MI differences matrices of wildtype apo vs wildtype holo and wildtype holo vs mutant holo simulations (Table S3), performance evaluation of PLS predictions (Figure S3), projections of PLS models on protease (Figure S4), MI differences between wildtype apo and holo simulations (Figure S5), side chain angles distribution of I84 in wildtype and L90M mutant complexes with SQV (Figure S6), principal components analysis of wildtype and mutant complexes with SQV (Figure S7), and experimental estimates of the drug binding energy change upon mutation cited in the current study compared with predictions from the present and other studies (Table S4) (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author
*E-mail: kalinina@mpi-inf.mpg.de.

### ORCID Ⓞ
Tomas Bastys: 0000-0002-1120-1663
Vytautas Gapsys: 0000-0002-6761-7780
Bert L. de Groot: 0000-0003-3570-3534

### Notes
The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

## ■ 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*; 2nd ed.; 2016.

(2) Antiretroviral Therapy Cohort Collaboration. Survival of HIV-Positive Patients Starting Antiretroviral Therapy Between 1996 and 2013: A Collaborative Analysis of Cohort Studies. *Lancet HIV* **2017**, *4*, e349−e356.

(3) The Stanford HIV Drug Resistance Database. https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/ (accessed May 29, 2018).

(4) Garriga, C.; Pérez-Elías, M. J.; Delgado, R.; Ruiz, L.; Nájera, R.; Pumarola, T.; Alonso-Socas Mdel, M.; Garca-Bujalance, S.; Menéndez-Arias, L. Mutational Patterns and Correlated Amino Acid Substitutions in the HIV-1 Protease after Virological Failure to Nelfinavir- and Lopinavir/Ritonavir-Based Treatments. *J. Med. Virol.* **2007**, *79*, 1617−1628.

(5) Masquelier, B.; Breilh, D.; Neau, D.; Lawson-Ayayi, S.; Lavignolle, V.; Ragnaud, J. M.; Dupon, M.; Morlat, P.; Dabis, F.; Fleury, H. Human Immunodeficiency Virus Type 1 Genotypic and Pharmacokinetic Determinants of the Virological Response to Lopinavir-Ritonavir-Containing Therapy in Protease Inhibitor-Experienced Patients. *Antimicrob. Agents Chemother.* **2002**, *46*, 2926−2932.

(6) Delaugerre, C.; Pavie, J.; Palmer, P.; Ghosn, J.; Blanche, S.; Roudiere, L.; Dominguez, S.; Mortier, E.; Molina, J. M.; de Truchis, P. Pattern and Impact of Emerging Resistance Mutations in Treatment Experienced Patients Failing Darunavir-Containing Regimen. *AIDS* **2008**, *22*, 1809−1813.

(7) Sterrantino, G.; Zaccarelli, M.; Colao, G.; Baldanti, F.; Di Giambenedetto, S.; Carli, T.; Maggiolo, F.; Zazzi, M. Genotypic Resistance Profiles Associated with Virological Failure to Darunavir-Containing Regimens: a Cross-Sectional Analysis. *Infection* **2012**, *40*, 311−318.

(8) Tisdale, M.; Myers, R. E.; Maschera, B.; Parry, N. R.; Oliver, N. M.; Blair, E. D. Cross-Resistance Analysis of Human Immunodeficiency Virus Type 1 Variants Individually Selected for Resistance to Five Different Protease Inhibitors. *Antimicrob. Agents Chemother.* **1995**, *39*, 1704−1710.

(9) Partaledis, J. A.; Yamaguchi, K.; Tisdale, M.; Blair, E. E.; Falcione, C.; Maschera, B.; Myers, R. E.; Pazhanisamy, S.; Futer, O.; Cullinan, A. B.; M, S. C.; 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**, *28*, 5228−5235.

(10) Maguire, M.; Shortino, D.; Klein, A.; Harris, W.; Manohitharajah, V.; Tisdale, M.; Elston, R.; Yeo, J.; Randall, S.; Xu, F.; Parker, H.; May, J.; Snowden, W. Emergence of Resistance to Protease Inhibitor Amprenavir in Human Immunodeficiency Virus Type 1-Infected Patients: Selection of Four Alternative Viral Protease Genotypes and Influence of Viral Susceptibility to Coadministered Reverse Transcriptase Nucleoside Inhibitors. *Antimicrob. Agents Chemother.* **2002**, *46*, 731−738.

(11) 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.

(12) Mittal, S.; Bandaranayake, R. M.; King, N. M.; Prabu-Jeyabalan, M.; Nalam, M. N.; Nalivaika, E. A.; Yilmaz, N. K.; Schiffer, C. A. Structural and Thermodynamic Basis of Amprenavir/Darunavir and Atazanavir Resistance in HIV-1 Protease with Mutations at Residue 50. *J. Virol.* **2013**, *87*, 4176−4184.

(13) Cunyat, F.; Ruiz, L.; Marfil, S.; Puig, T.; Bofill, M.; Blanco, J.; Clotet, B.; Cabrera, C. Genotypic and Phenotypic Evolution of HIV Type-1 Protease During in Vitro Sequential or Concomitant Combination of Atazanavir and Amprenavir. *Antiviral Ther.* **2010**, *15*, 431−436.

(14) Bethell, R.; Scherer, J.; Witvrouw, M.; Paquet, A.; Coakley, E.; Hall, D. Phenotypic Protease Inhibitor Resistance and Cross-Resistance

in the Clinic from 2006 to 2008 and Mutational Prevalences in HIV from Patients with Discordant Tipranavir and Darunavir Susceptibility Phenotypes. *AIDS Res. Hum. Retroviruses* **2012**, *28*, 1019−1024.

(15) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P.; Poljak, R. J. The Three-Dimensional Structure of the Aspartyl Protease from the HIV-1 Isolate BRU. *Biochimie* **1991**, *73*, 1391−1396.

(16) Jacobsen, H.; Hänggi, M.; Ott, M.; Duncan, I. B.; Owen, S.; Andreoni, M.; Vella, S.; Mous, J. In Vivo Resistance to a Human Immunodeficiency Virus Type 1 Proteinase Inhibitor: Mutations, Kinetics, and Frequencies. *J. Infect. Dis.* **1996**, *173*, 1379−1387.

(17) Craig, C.; Race, E.; Sheldon, J.; Whittaker, L.; Gilbert, S.; Moffatt, A.; Rose, J.; Dissanayeke, S.; Chirn, G. W.; Duncan, I. B.; Cammack, N. HIV Protease Genotype and Viral Sensitivity to HIV Protease Inhibitors Following Saquinavir Therapy. *AIDS* **1998**, *12*, 1611−1618.

(18) Winters, M.; Schapiro, J. M.; Lawrence, J.; Merigan, T. C. Human Immunodeficiency Virus Type 1 Protease Genotypes and in Vitro Protease Inhibitor Susceptibilities of Isolates from Individuals Who Were Switched to Other Protease Inhibitors after Long-Term Saquinavir Treatment. *J. Virol.* **1998**, *72*, 5303−5306.

(19) Watkins, T.; Resch, W.; Irlbeck, D.; Swanstrom, R. Selection of High-Level Resistance to Human Immunodeficiency Virus Type 1 Protease Inhibitors. *Antimicrob. Agents Chemother.* **2003**, *47*, 759−769.

(20) Sevin, A. D.; DeGruttola, V.; Nijhuis, M.; Schapiro, J. M.; Foulkes, A. S.; Para, M. F.; Boucher, C. A. Methods for Investigation of the Relationship between Drug-Susceptibility Phenotype and Human Immunodeficiency Virus Type 1 Genotype with Applications to AIDS Clinical Trials Group 333. *J. Infect. Dis.* **2000**, *182*, 59−67.

(21) 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.* **2004**, *13*, 1108−1123.

(22) Perryman, A. L.; Lin, J. H.; McCammon, J. A. Restrained Molecular Dynamics Simulations of HIV-1 Protease: The First Step in Validating a New Target for Drug Design. *Biopolymers* **2006**, *82*, 272−284.

(23) 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.* **2015**, *5*, 10517.

(24) Hou, T.; Yu, R. Molecular Dynamics and Free Energy Studies on the Wild-type and Double Mutant HIV-1 Protease Complexed with Amprenavir and Two Amprenavir-Related Inhibitors: Mechanism for Binding and Drug Resistance. *J. Med. Chem.* **2007**, *50*, 1177−1188.

(25) Kožíšek, M.; Bray, J.; Řezáčová, P.; Šašková, K.; Brynda, J.; Pokorná, J.; Mammano, F.; Rulíšek, L.; Konvalinka, J. Molecular Analysis of the HIV-1 Resistance Development: Enzymatic Activities, Crystal Structures, and Thermodynamics of Nelfinavir-Resistant HIV Protease Mutants. *J. Mol. Biol.* **2007**, *374*, 1005−1016.

(26) Stoica, I.; Sadiq, S. K.; Coveney, P. V. Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases. *J. Am. Chem. Soc.* **2008**, *130*, 2639−2648.

(27) Chen, J.; Zhang, S.; Liu, X.; Zhang, Q. Insights into Drug Resistance of Mutations D30N and I50V to HIV-1 Protease Inhibitor TMC-114: Free Energy Calculation and Molecular Dynamic Simulation. *J. Mol. Model.* **2010**, *16*, 459−468.

(28) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V. Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *J. Chem. Theory Comput.* **2014**, *10*, 1228−1241.

(29) 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.

(30) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040−4048.

(31) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* **2005**, *48*, 7796−7807.

(32) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.

(33) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.

(34) Crooks, G. E. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *60*, 2721−2726.

(35) Seeliger, D.; de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.* **2010**, *98*, 2309−2316.

(36) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chem., Int. Ed.* **2016**, *55*, 7364−7368.

(37) Cortes Cabrera, A.; Sanchez-Murcia, P. A.; Gago, F. Making Sense of the Past: Hyperstability of Ancestral Thioredoxins Explained by Free Energy Simulations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 23239−23246.

(38) Seeliger, D.; Buelens, F. P.; Goette, M.; de Groot, B. L.; Grubmüller, H. Towards Computional Specificity Screening of DNA-Binding Proteins. *Nucleic Acids Res.* **2011**, *39*, 8281−8290.

(39) Gapsys, V.; de Groot, B. L. Alchemical Free Energy Calculations for Nucleotide Mutations in Protein-DNA Complexes. *J. Chem. Theory Comput.* **2017**, *13*, 6275−6289.

(40) Giovannelli, E.; Cioni, M.; Procacci, P.; Cardini, G.; Pagliai, M.; Volkov, V.; Chelli, R. Binding Free Energies of Host-Guest Systems by Nonequilibrium Alchemical Simulations with Constrained Dynamics: Illustrative Calculations and Numerical Validation. *J. Chem. Theory Comput.* **2017**, *13*, 5887−5899.

(41) Goette, M.; Grubmüller, H. Accuracy and Convergence of Free Energy Differences Calculated from Nonequilibrium Switching Processes. *J. Comput. Chem.* **2009**, *30*, 447−456.

(42) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(43) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

(44) Muzammil, S.; Armstrong, A. A.; Kang, L. W.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, L. M.; Freire, E. Unique Thermodynamic Response of Tipranavir to Human Immunodeficiency Virus Type 1 Protease Drug Resistance Mutations. *J. Virol.* **2007**, *81*, 5144−5154.

(45) Josefsson, L.; King, M. S.; Makitalo, B.; Brännström, J.; Shao, W.; Maldarelli, F.; Kearney, M. F.; Hu, W.-S.; Chen, J.; Gaines, H.; Mellors, J. W.; Albert, J.; Coffin, J. M.; Palmer, S. E. Majority of CD4+ T Cells from Peripheral Blood of HIV-1−Infected Individuals Contain Only One HIV DNA Molecule. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11199−11204.

(46) Josefsson, L.; Palmer, S.; Faria, N. R.; Lemey, P.; Casazza, J.; Ambrozak, D.; Kearney, M.; Shao, W.; Kottilil, S.; Sneller, M.; Mellors, J.; Coffin, J. M.; Maldarelli, F. Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals That the Majority of CD4+ T-Cells Contain One HIV-1 DNA Molecule. *PLoS Pathog.* **2013**, *9*, e1003432.

(47) 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.

(48) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(49) Abraham, M. J.; Murtola, T.; Schulz, R.; P'all, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Super-computers. *SoftwareX* **2015**, *1−2*, 19−25.

(50) Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.

(51) Duan, R.; Lazim, R.; Zhang, D. Understanding the Basis of I50V-Induced Affinity Decrease in HIV-1 Protease via Molecular Dynamics Simulations Using Polarized Force Field. *J. Comput. Chem.* **2015**, *36*, 1885−1892.

(52) Chemaxon *Calculator,* Version 5.3.8; ChemAxon: 2010. http://www.chemaxon.com (accessed June 5, 2018).

(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09*, Revision C.01; Gaussian Inc.: Wallingford, CT, 2010.

(54) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

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

(56) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE-Antechamber Python Parser Interface. *BMC Res. Notes* **2012**, *5*, 367.

(57) Joung, I. S.; Cheatham, 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.

(58) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(59) Parrinello, M.; Rahman, A. Canonical Sampling Through Velocity Rescaling. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(60) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(61) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(62) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated Protein Structure and Topology Generation for Alchemical Perturbations. *J. Comput. Chem.* **2015**, *36*, 348−354.

(63) Gapsys, V.; Seeliger, D.; de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2373−2382.

(64) 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.

(65) Krivobokova, T.; Briones, R.; Hub, J. S.; Munk, A.; de Groot, B. L. Partial Least-Squares Functional Mode Analysis: Application to the Membrane Proteins AQP1, Aqy1, and CLC-ec1. *Biophys. J.* **2012**, *103*, 786−796.

(66) Gapsys, V.; de Groot, B. L. Optimal Superpositioning of Flexible Molecule Ensembles. *Biophys. J.* **2013**, *104*, 196−207.

(67) Tibshirani, R.; Walther, G.; Hastie, T. Estimating the Number of Clusters in a Data Set via the Gap Statistic. *J. R. Statist. Soc. B* **2001**, *63*, 411−423.

(68) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2009**, *5*, 2486−2502.

(69) Hyland, L. J.; Tomaszek, T. A. J.; 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.

(70) 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.

(71) 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.

(72) 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*, 9945−9950.

(73) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Erickson, J. W. Calculation of Relative Binding Free Energies of Peptidic Inhibitors to HIV-1 Protease and Its I84V Mutant. *J. Am. Chem. Soc.* **1998**, *120*, 8856−8863.

(74) 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. U. S. A.* **2009**, *106*, 4641−4646.

(75) 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.

(76) 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. C.; 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*, 10791−10792.

(77) McGee, T. D., Jr.; 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.

(78) Shen, C. H.; Wang, Y. F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. Amprenavir Complexes with HIV-1 Protease and Its Drug-Resistant Mutants Altering Hydrophobic Clusters. *FEBS J.* **2010**, *277*, 3699−3714.

(79) Liu, F.; Boross, P. I.; Wang, Y. F.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Kinetic, Stability, and Structural Changes in High-Resolution Crystal Structures of HIV-1 Protease with Drug-Resistant Mutations L24I, I50V, and G73S. *J. Mol. Biol.* **2005**, *354*, 789−800.

(80) Ermolieff, J.; Lin, X.; Tang, J. Kinetic Properties of Saquinavir-Resistant Mutants of Human Immunodeficiency Virus Type 1 Protease and Their Implications in Drug Resistance in Vivo. *Biochemistry* **1997**, *36*, 12364−12370.

(81) Maschera, B.; Darby, G.; Palú, G.; Wright, L. L.; Tisdale, M.; Myers, R.; Blair, E. D.; Furfine, E. S. Human Immunodeficiency Virus. Mutations in the Viral Protease That Confer Resistance to Saquinavir Increase the Dissociation Rate Constant of the Protease-Saquinavir Complex. *J. Biol. Chem.* **1996**, *271*, 33231−33235.

(82) 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.

(83) Nijhuis, M.; Schuurman, R.; de Jong, D.; Erickson, J.; Gustchina, E.; Albert, J.; Schipper, P.; Gulnik, S.; Boucher, C. A. Increased Fitness of Drug Resistant HIV-1 Protease As a Result of Acquisition of Compensatory Mutations During Suboptimal Therapy. *AIDS* **1999**, *13*, 2349−2359.

(84) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. Crystal Structure of an in Vivo HIV-1 Protease Mutant in Complex with Saquinavir: Insights into the Mechanisms of Drug Resistance. *Protein Sci.* **2000**, *9*, 1898−1904.

(85) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational Simulations of HIV-1 Proteases Multi-Drug Resistance due to Nonactive Site Mutation L90M. *J. Am. Chem. Soc.* **2006**, *128*, 7887−7895.

(86) Huang, J.; Sun, B.; Yao, Y.; Liu, J. Fast and Reliable Thermodynamic Approach for Determining Protonation State of Asp Dyad. *J. Chem. Inf. Model.* **2017**, *57*, 2273−2280.

(87) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.