CHAPTER 3

PREDICTION PROTONATION STATE OF HIV-1 PR BOUND TO 6 FDA-APPROVED DRUGS USING ENERGETIC AND STRUCTURAL ANALYSIS

3.1 Introduction

Computational structure-based drug design often plays a key role in the initial stages of developing new pharmaceutical agents. Detailed calculations of enzyme-inhibitor interactions based on physics-based all-atom force fields with or without quantum-mechanical treatments may provide a deeper understanding of the mode of interaction and suggest further modifications, for example to address the challenge of drug resistance from mutations in the target enzyme. Force field based calculations can also provide estimates of both relative and absolute binding affinities, although in practice it is often challenging to obtain accurate results.[73, 85, 86]

A widely studied example is the dimeric human immunodeficiency virus type 1 (HIV-1) aspartyl protease (Figure 3.1A). Because HIV-1 protease carries out the essential proteolytic cleavage of viral protein precursors into functional units,[87, 88] it is one of the major targets in AIDS therapies. Much efforts have been devoted to finding effective inhibitors of HIV-1 protease including many computational studies.[89] Previous studies were able to predict relative binding affinities of different inhibitors to HIV-1 protease with reasonable accuracies.[90-97] However, results vary greatly between different studies and calculated binding affinities for some inhibitors may deviate by tens of kcal/mol from experimental data, well outside

the experimental error[95, 96] unless parameterized linear interaction energy methods are employed.[97]

A central feature of the dimeric HIV-1 protease active site is the presence of two aspartates, D25 and D25', that form the catalytic dyad but also play a key role in enzyme-inhibitor interactions.[98-103] An unresolved question is the protonation state of D25 and D25' under physiological conditions. There appears to be some consensus that the OD2 oxygen of D25 is protonated in the apo-form of HIV-1 protease while D25' remains anioinic[104-108] although a dianionic state of both aspartates at pH has also been proposed.[109, 110] The protonation state upon binding of substrates and/or inhibitors appears to depend on the type of inhibitor. Negatively charged inhibitors or inhibitors that are not able to form hydrogen bonds to both aspartates or to one aspartate and a bridging water molecule seem to stabilize the doubly protonated form.[111] In most other cases it has been suggested that HIV-1 protease is protonated either at D25 or D25' upon binding of inhibitors[92, 96, 97, 104, 111, 112] with different studies not always in agreement with each other. However, the correct assignment of the protonation state in a given HIV-1 protease-inhibitor complex is critical for effective inhibitor design and for the calculation of binding affinities close to experimental values.

Molecular dynamics (MD) simulations are often applied to obtain conformational sampling in the computational prediction of protein-ligand binding free energies.[73, 113] Within the statistical thermodynamics framework, free energy perturbation (FEP) and thermodynamic integration (TI)[114] are the most rigorous methods for the calculation of relative binding free energies. At least in principle, these methods can provide a high level of accuracy within the limitations of the usually classical description of molecular interactions from a given force field.[115]

However, calculations that employ these methods are very costly because extensive sampling from initial to final states is required in a given free energy calculation for satisfactory convergence. In the case of ligand binding, the protein and ligand are at infinite separation in the initial state and fully associated in the final state so that direct application of FEP or TI methods would require simulation of the ligand association or dissociation process. In a more practical, yet still costly approach, relative binding affinities can be obtained from FEP or TI methods through computational alchemy involving the gradual conversion between two ligands.[116]

Alternative methods have been devised where only the initial and final states are compared so that sampling can be reduced significantly.[117-124] A particularly popular method is the MMPB/SA (Molecular Mechanics Poisson-Boltzmann and Surface Area) approach.[75] In this method, configurations of the protein, the ligand, and the protein-ligand complex are obtained from explicit solvent molecular dynamics simulations and subsequently rescored with a free energy estimate based on an implicit solvent dielectric continuum model.[77]

The present study[125] also follows the MMPB/SA approach to determine binding free energies for HIV-1 protease in complex with the six FDA-approved inhibitors saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, and lopinavir. However, instead of deciding on a protonation state a priori as in previous studies, binding free energies are compared between different protonation states and combined with a structural analysis of the simulated protein-inhibitor complexes in the context of the crystallographic data in order to assign the D25/D25' protonation states more reliably. As a result, the calculated binding free energies are found to be more consistent with experimental data.

A related technical issue that is being addressed in this study is to what extent MMPB/SA calculations could be substituted with the more efficient MMGB/SA method that uses the Generalized Born approximation[79] instead of solutions to the Poisson equation. Previous studies have used either the MMPB/SA[92, 126] or MMGB/SA[95] approach, but a direct comparison between both methods for HIV-1 protease ligand binding calculations is not available.

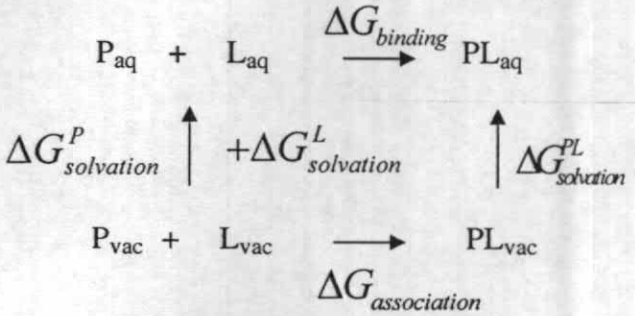
3.2 Methods

3.2.1 Explicit solvent molecular dynamics simulations of HIV-1 PR ligand complexes

Conformational sampling of HIV-1 protease in complex with lopinavir (LPV), ritonavir (RTV), saquinavir (SQV), indinavir (IDV), amprenavir (APV), and nelfinavir (NFV) was obtained from molecular dynamics simulations with Amber 8.[127] Each complex was simulated with four different protonation states: 1) anionic form of both aspartates (called "D-"), 2) protonated D25, anionic D25' (called "D25"), 3) protonated D25', anionic D25 (called "D25'"), and 4) protonated D25 and D25' (called "D25,25'"). The combination of six ligands with the four protonation states resulted in a total of 24 simulations. Crystallographic structures of the complexes were used as the starting structures (PDB codes: 1MUI[128] (LPV), 1HXW[129] (RTV), 1HXB[130] (SQV), 1HSG[131] (IDV), 1HPV[132] (APV), and 1OHR[133] (NFV). Each complex was solvated in a rectangular periodic box using the TIP3P water model[134] and neutralized with 4 (D-), 5 (D25 or D25'), or 6 (D25,D25') Na^+ and 10 Cl^- counterions. At least 10 Å of solvent were allowed between the edge of the box and the closest atom of the solute. The Amber94[135] force field was used to model the enzyme. Standard RESP methodology[136] was

applied to obtain force field parameters for the inhibitors that fit quantum mechanical electrostatic potentials at the HF/6-31G* level, obtained with Gaussian.[137] During the simulations, electrostatic interactions were calculated using the Particle-Mesh Ewald (PME) method[138] with a direct space cutoff distance of 12 Å.

Before starting the molecular dynamics runs, all of the systems were subjected to the following minimization protocol: Water molecules and counterions were minimized first while restraining the solute over 500 steps each of steepest descent and conjugate gradient minimization. The same protocol was then repeated without solute restraints for 2,500 steps of steepest descent and 1,000 steps of conjugate gradient minimization. Molecular dynamics simulations in the NVT ensemble were subsequently started in which each system was heated from 0K to 298K over 20 ps. During the initial heating phase the solute was restrained harmonically with a force constant of 5 kcal.mol^{-1}.Å^{-2}. The solute restraint was then removed and molecular dynamics simulations in the NPT ensemble at a temperature of 298K and a pressure of 1 atm were continued for another 3 ns. Bonds involving hydrogen atoms were constrained using SHAKE[139] so that a time step of 2 fs could be used during all of the simulations.



Scheme 3.1 Thermodynamic cycle for the calculation of binding free energy.

3.2.2 Ligand Binding Free Energy Calculations

Absolute binding free energies were estimated using the MMPB/SA or MMGB/SA scheme according to the free energy cycle shown in scheme 3.1. Instead of a direct calculation, the binding free energy of the complex ($\Delta G_{binding}$) in aqueous solution can be obtained as follows:

$$\Delta G_{binding} = \Delta G_{association} + \Delta G^{PL}_{solvation} - \Delta G^{P}_{solvation} - \Delta G^{L}_{solvation} \qquad (3.1)$$

$\Delta G_{association}$ is the Gibbs energy of protein-ligand association in vacuum, which can be computed from enthalpic and entropic terms as $\Delta G_{association} = \Delta H_{association} - T\Delta S_{association}$. $\Delta H_{association}$ is calculated from the force field using a combination of the standard bonded and non-bonded terms. The change in entropy upon association ($\Delta S_{association}$) reflects translational, rotational, and vibrational contributions that are calculated based on classical statistical thermodynamics.[140] The loss of translational and rotational degrees of freedom upon protein-ligand association is easily estimated from masses and the moments of inertia of the different ligands as in a previous study.[141] The vibrational contribution to the entropy of ligand association can be estimated under the harmonic approximation from normal-mode analyses of the protein, ligand, and protein-ligand complex following extensive minimization. Normal mode calculations were performed with a distance dependent dielectric according to *4r*, where *r* is the pair wise distance between atoms.

$\Delta G^{P}_{solvation}$, $\Delta G^{L}_{solvation}$ and $\Delta G^{PL}_{solvation}$ are the solvation free energies of protein (P), ligand (L), and complex (PL) including both polar and non-polar contributions. The polar contributions are commonly obtained from Poisson-Boltzmann (PB) calculations or can be approximated with a generalized Born (GB) model formalism. In this study, we calculated Poisson-based electrostatic solvation energies in the MMPB/SA approach using DelPhi with the dielectric interface defined

from the molecular surface[142] and a grid spacing of 0.4 Å. Electrostatic solvation energies in the MMGB/SA scheme were obtained with the standard pair wise descreening GB method implemented in Amber (*igb*=1)[143] as well as with the GBMV[144] method implemented in CHARMM.[145] Bondi radii[146] were used to define the dielectric boundary with the Amber GB method and with GBMV. In order to be able to use the GBMV method from CHARMM for this analysis the Amber force field parameters for the ligands were translated to the CHARMM force field format and used along with the standard Amber94 potential that has been used in CHARMM previously.[147] A dielectric constant of 80 was used in all cases. The non-polar contribution to the solvation free energy was estimated from

$$\Delta G_{non-polar} = \gamma \cdot SASA + b \qquad (3.2)$$

where SASA is the solvent-accessible surface area, γ is set to 0.00542 cal/mol/Å^2 and b was set equal to 0.92 kcal/mol/Å^2 following previously suggested values.[148] Ligand binding free energies were averaged over 100 snapshots taken from the last 1 ns of the explicit solvent molecular dynamics simulations to allow for ample equilibration time. Due to computational limitations, normal-mode calculations where only carried out for the first, 50th, and 100th snapshot from each trajectory.

3.2.3 *Explicit/implicit hybrid scheme*

In addition to the standard MMPB(GB)/SA scheme, we also investigated whether the inclusion of a limited number of explicit water molecules near the protein-ligand binding site has a significant effect on the results. In order to address this question we included water molecules within 6 Å from the central hydroxyl group of the inhibitors in each of the 100 snapshots. Typically, this resulted in the addition 3-5 water molecules. In order to obtain a conformationally averaged estimate of water-solute interactions, each snapshot including the selected explicit

water molecules was subjected to additional short molecular dynamics simulations over 100 ps each. These simulations were carried out with CHARMM using the GBMV implicit solvent model to provide a realistic solvent environment beyond the explicitly included water molecules. Because the sole purpose of these additional simulations was to sample water configurations, both the protein and ligand were fixed while only the explicit water molecules were allowed to move freely. 20 snapshots were taken from each of these simulations to estimate and average binding free energies according to the scheme described above but with the explicit water molecules included along with the protein in the energy decomposition.

3.3 Results

Binding free energies of HIV-1 protease in complex with inhibitors LPV, RTV, SQV, IDV, APV, and NFV were estimated according to MMPB(GB)/SA schemes from single-trajectory molecular dynamics simulations of each of the complexes as described in the methods section. Table 3.1 presents the calculated binding affinities for each ligand and protonation state with different MMPB/SA and MMGB/SA schemes. A detailed break-up of individual energetic contributions is given in Tables 3.2 and 3.3. All of the values include the -TΔS contribution at 300K and are therefore meant to be considered as estimates of absolute binding affinities. However, it should be pointed out, that the reported binding affinities are calculated for a fixed protonation state of the enzyme. This calculation neglects the energetic penalty of changing the protonation state from the assumed monoprotonated, D25, form of the enzyme in the absence of a ligand[104-107], which will be addressed in more detail below. Because conformations were extracted from a single trajectory, the effect of possibly substantial conformational rearrangements in the enzyme in the

absence of any of the ligands is also neglected. Estimates of the statistical uncertainties in Table 3.1 were calculated from the difference in block averages over the first and second half of the snapshots. Because of possible structural and energetic correlation between subsequent snapshots this measure is expected to provide a better estimate of the statistical error than an error estimate based on $\sigma/\sqrt{N}$, where σ is the standard deviation and N is the number of samples and independency of all samples is assumed. Errors calculated based on the standard deviations are generally lower in the range of 0.2-0.6 kcal/mol.

3.3.1 Comparison of MMPB/SA and MMGB/SA Schemes

Although the standard MMPB/SA scheme is well-established, it is limited by the relatively high expense of carrying out Poisson-Boltzmann calculations for each snapshot. As a consequence only a relatively small number of snapshots from the explicit solvent molecular dynamics simulations are typically evaluated. In order to speed up the calculations and gain the ability to evaluate a larger number of snapshots, the Generalized Born approximation[79, 143, 149, 150] has been used instead.[95] However, few studies have systematically examined how well the Generalized Born approximation performs in ligand-binding affinity calculations[151, 152] in comparison with the Poisson reference energies. The different columns in Table 3.2 compare the effect of obtaining the electrostatic solvation energy from PB or GB methods in the calculation of binding affinities for a given ligand and protonation state. Note that numbers in blanket refer to the free energy change from free D25 to D25,25', $\Delta G_{free(D25\rightarrow 25,25')}$ which defined in schematic 3.2. They are not taken into consideration under these sections. At a first glance it can be seen that the GB methods result in significantly different absolute binding affinity estimates,

especially with the GB method based on the pairwise descreening approximation from Amber[143], where the difference is on the order of 40-50 kcal/mol. Such deviations may be acceptable, however, if the relative ranking between different protonation states and ligands is maintained. In that respect, the data in Table 3.1 shows that the overall ranking is largely maintained with the GBMV method but the GB method in Amber has more difficulties in providing the same qualitative ranking as the MMPB/SA method between different protonation states as well as different ligands. For example, the ranking of ligands according to the lowest binding energy to a monoprotonated enzyme from PB is RTV=LPV>IDV>SQV=NFV>APV, but according to the Amber-GB method it isIDV>RTV>LPV>SQV>NFV>APV. The GBMV method switches the order of SQV and NFV, but it confirms the MMPB/SA finding that the binding energies are very similar for both inhibitors. Furthermore, the difference between RTV and LPV binding affinities is much larger with GBMV compared to MMPB/SA results.

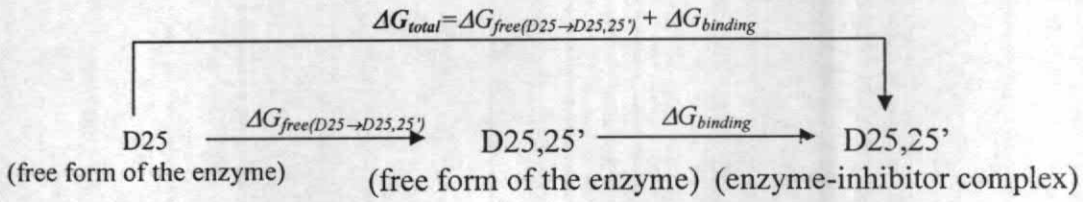
A more detailed comparison of the performance of the GB methods in comparison with the PB reference is shown in Figure 3.2. Both GB methods display significant noise when comparing individual snapshot energies between PB and GB methods. However, overall, the GBMV method is in good absolute agreement with PB and also maintains the relative ranking between different snapshots fairly well, while the results from the GB method in Amber are far from the diagonal line and do not rank some of the snapshots correctly. This is especially obvious in the binding energies for the unprotonated enzyme (shown in blue) with the ligands LPV, RTV, and SQV that are much less favorable than for the protonated enzyme according to the PB and GBMV calculations but have about the same energy with the GB method in Amber.

3.3.2 *Prediction of Enzyme Protonation State from Binding Affinites*

The protonation state of HIV-1 protease when binding to a given inhibitor may be predicted according to the lowest calculated binding free energy for a given inhibitor. However, before the values in Table 3.1 can be compared, the energetic cost of altering the protonation state of the enzyme from the assumed D25 protonation state for the apo-form at physiological pH has to be considered using the following assumptions. *(i)* This contribution is expected to be significant for both the unprotonated and diprotonated forms of the enzyme. A closer look at the binding affinities for binding to the dianionic form of the aspartic dyad reveals uniformly less favorable binding that the mono- or diprotonated forms of the enzyme. Adding the cost of deprotonating D25 would result in even less favorable binding energies. Consequently, it does not appear likely that both catalytic aspartates are deprotonated when any of the ligands studied here are bound to HIV-1 protease. *(ii)* On the other hand, moving the proton from D25 to D25' is assumed to incur only a slight energetic cost that is within the error of the calculations presented here according to previous results[111] and will be neglected in the following.

In order to assess whether the diprotonated form is in fact the most likely protonation state for these inhibitors, the free energy change between the monoprotonated (D25) and diprotonated form of the free enzyme ($\Delta G_{free(D25\rightarrow D25,25')}$ according to schematic 2) is needed. Note that the data shown in Table 3.1 excludes the $\Delta G_{free(D25\rightarrow D25,25')}$ term. This can lead consequently to a misunderstanding. As can be seen from Table 3.2, however, that the binding affinities for the mono- and diprotonated forms are comparable, with more negative binding affinities to the

diprotonated forms for SQV, IDV, APV, and NFV. To overcome this discrepancy, the $\Delta G_{free(D25\rightarrow D25,25')}$ were calculated according to equations (3.3-3.5).



Scheme 3.2 The total binding affinity (ΔG_{total}) of the complexes for the D25,25' state calculated as a summation between the free energy change from D25 to D25,25' in free form ($\Delta G_{free(D25\rightarrow D25,25')}$) according to equation 3.3 and $\Delta G_{binding}$ according to equation 3.1.

Following basic thermodynamics of equilibrium processes, the free energy change of protonating an ionizable group, in this case D25', with a given pK_a value in aqueous solvent with a certain pH can be calculated from

$$\Delta G = -RT\ln 10^{-(pH-pK_a)} \qquad (3.3)$$

The pK_a value of D25' is given as $pK_a = pK_{a,free} + pK_{a,shift}$ where $pK_{a,free}$= 3.9 for a free aspartate and $pK_{a,shift}$ is the shift due to the enzyme environment. $pK_{a,shift}$ is related to the free energy difference

$$\Delta\Delta G = \Delta G(EnzH \rightarrow Enz) - \Delta G(AspH \rightarrow Asp). \qquad (3.4)$$

of protonating aspartate by itself and in the context of the enzyme according to

$$\Delta\Delta G = -RT\ln K_{a,shift} \qquad (3.5)$$

The quantity $\Delta G(EnzH \rightarrow Enz)$ can be determined from the difference in MMPB/SA energies of the free enzyme with D25 and D25,D25' protonations. If one assumes that

ligand binding does not substantially alter the sampling of enzyme conformations, one can use the same data that was used to calculate ligand binding affinities. $\Delta G(AspH \rightarrow Asp)$ is estimated from evaluating minimized aspartate dipeptide in anionic and neutral forms with the MMPB/SA formalism. While classical force fields cannot provide reasonable estimates of absolute pK_a values, the difference $\Delta\Delta G$ can be determined more reliably when the same force field is used in the calculation of $\Delta G(EnzH \rightarrow Enz)$ and $\Delta G(AspH \rightarrow Asp)$. The free energy change between the D25 and D25' states of the free enzyme $\Delta G_{free(D25 \rightarrow D25,25')}$ calculated according schematic 2 for the diprotonated form of all complexes at $pH = 7$ and T = 298 K were calculated using equation 3.3-3.5. The results are shown in Table 3.5. While inhibitors were not included in these calculations, the obtained free energies depend on the inhibitor that was present during the sampling of enzyme conformations.

The combination of values from Table 3.4 ($\Delta G_{free(D25 \rightarrow D25,25')}$ calculated according to schematic 2) with the binding free energies ($\Delta G_{binding}$ according to equation 3.1) for the diprotonated enzyme state from Table 3.2 were summarized in the blanket in Table 3.2. The total binding free energies for the D25,25' (in blanket) are less favorable than the binding energies for the monoprotonated form. Consequently, we would predict that HIV-1 protease remains monoprotonated when either one of the six inhibitors studied here are bound. However, it appears possible that the proton shifts from D25 to D25'. According to the MMPB/SA results, the lowest binding affinity is found for D25 protonation when LPV, RTV, SQV, or IDV is bound. For APV, either D25 or D25' protonation is possible within the uncertainties of the calculation and for NFV D25' appears to be the preferred protonation state.

The standard MMPB/SA formalism does not include any explicit water molecules in the energetic analysis of the snapshots from the molecular dynamics simulation. The advantage of the implicit model is that the mean-field, or instantaneously averaged interactions between a given snapshot and the solvent are available without the need for further sampling. However, it is not entirely clear to what extent this approximation affects the accuracy of the obtained free energies. Rather than including all water molecules, one possibility is to include few selected water molecules. This idea of a hybrid explicit/implicit solvation model has been pursued before with inconclusive results as to whether the inclusion of explicit water improves the accuracy of the results.[153] Here, this idea was also tested by including a small number of water molecules in the MMPB/SA analysis. In order to obtain solvent-averaged energy contributions the explicit water molecules were allowed to sample conformational space in the context of the frozen enzyme-inhibitor complex and with implicit solvent to represent the remaining solvent environment. Such an approach is more expensive than the standard MMPB/SA method, but the additional computational cost is moderate compared to additional sampling with full explicit solvent. Table 3.1 shows the results obtained with the hybrid implicit/explicit MMPB/SA scheme. The resulting energies are generally similar to the results with the standard MMPB/SA method. However, in the case of LPV and IDV binding to the monoprotonated enzyme, the difference between the binding affinities for D25' or D25 protonation is within the statistical uncertainties, while binding to D25 is clearly more favorable according to the MMPB/SA results. Furthermore, the relative ordering of different inhibitors is changed slightly from the MMPB/SA results to RTV>IDV>LPV>SQV=NFV>APV.

Based on the analysis of binding affinities, D25 protonation is predicted for RTV and SQV at *pH*=7 while D25' protonation is predicted for NFV. D25 protonation is also more likely for LPV and IDV. However, the results from the hybrid implicit/explicit MMPB/SA calculations permit D25' protonation as well. In the case of APV the binding affinity calculations cannot distinguish between binding to HIV-1 protease protonated either at D25 or D25'.

In addition to the ΔG_{total} at *pH*=7 and 298 K for the D25,25' state as shown in the blanket in Table 3.1, the ΔG_{total} of pH=3-6 were also calculated for the 6 drugs. The results were given in Table 3.5 and 3.6 for the implicit and implicit/explicit MMPB/SA, respectively. In the case of SQV, the hybrid MMPB/SA scheme (Table 3.6) results in a very favorable binding affinity for the diprotonated form of the enzyme that could become more favorable over the monoprotonated form around a *pH* of 5 when the cost of protonating the second aspartate in the free enzyme is included. Both the standard and hybrid MMPB/SA calculations support a similar conclusion for the diprotonated complex with IDV as well that could become favorable around a *pH* of 3 according to the calculations presented here (Table 3.5-3.6).

3.3.3 Prediction of Protonation State from Structural Analysis

Crystal structures of HIV-1 protease complexed with all of the inhibitors examined in this study indicate a well-defined inhibitor structure in a certain conformation. The major degrees of freedom of the inhibitors are the torsion angles at different sub sites as indicated in Figure 3.1B. An analysis of the simulated inhibitor conformations in terms of these torsion angles may provide insight into which protonation states of the enzyme lead to structures that are incompatible with

the crystallographic data and therefore less favorable.[112] The results of such an analysis are shown in Figure 3.3 where the predominant torsion angles in each of the simulations are compared with the experimental data. While most torsion angles remain very near to the experimental values, we will focus on the cases where the sampled conformations deviate from experiment. In general, the unprotonated form of the enzyme leads to different sampling from experiment in at least one torsion angle for most inhibitors (except for IDV and RTV). Furthermore, we find that sampling of the complexes with the diprotonated enzyme is fully compatible with the crystallographic data for all of the inhibitors. D25' monoprotonation is incompatible with experiment for LPV and SQV and D25 monoprotonation is in disagreement with experiment when RTV, APV, or NFV are bound.

As another aspect, the structure of the enzyme may also be compared between the simulations and crystallography. Overall, the simulated structure of HIV-1 protease remains close to the crystal structure with ~1 Å backbone root mean square deviation, but the position of the catalytic aspartates warrants a closer look. Table 3.7 shows averaged side chain RMSD values for D25 and D25' from the simulations for different inhibitors and different protonation states. It is interesting that the diprotonated enzyme universally leads to better agreement with the crystal structures than the monoprotonated forms. Among the monoprotonated states, the lower RMSD values are around 1.1 Å except for LPV, where a value of 1.7 Å suggests more significant structural deviations.

Based on the structural analysis, the enzyme would be predicted to be diprotonated. However, if the diprotonated state is excluded the structural analysis would predict D25 protonation for LPV, SQV, and IDV, but D25' protonation for

RTV, APV, and NFV. In the following section these results are combined with the energetic analysis to form a more complete picture.

3.3.4 *Prediction of Protonation State from Combined Energetic and Structural Analysis*

Neither the analysis of binding affinities nor the comparison of conformational sampling of the inhibitors with the crystallographic data provides an unambiguous picture of HIV-1 protease protonation upon binding of the inhibitors studied here. A consensus can be found by combining the energetic and structural analysis as shown in Table 3.8. Accordingly, it is predicted that the enzyme is protonated at D25 when LPV, SQV, or IDV are bound, but protonated at D25' when APV or NFV is bound. In the case of IDV, D25' protonation could also be possible. The protonation state of HIV-1 protease when RTV is bound is inconclusive because the structural analysis favors the energetically less favorable D25' protonation state. Diprotonation is predicted for SQV and IDV at low pH.

These results may be compared with previous studies that have assigned protonation states to HIV-1 protease when bound to SQV, IDV, APV, and NFV.[92, 96, 97] Agreement is found at least in part for SQV and APV binding, but not for NFV binding where D25 protonation has been suggested before.[97] In the case of IDV, three different studies suggest D25' protonation[92, 96, 97], which is possible but less likely according to this study.

3.3.5 *Comparison between calculated and experimental binding affinities*

After assignment of the enzyme protonation state for each of the inhibitors, the corresponding calculated binding affinities from Table 3.1 can be compared to the experimental data. The resulting root mean square deviations between calculated and experimental binding affinities are given in Table 3.9. Because the protonation states of RTV and IDV could not be assigned unambiguously, four different sets of protonation states were considered. Better agreement with the experimental data is found when the enzyme is protonated at D25' when RTV is bound. Based on this finding, D25' protonation with the RTV inhibitor appears to be more likely. There is little difference between D25 and D25' protonation for the IDV inhibitor, although D25' protonation leads to slightly better results within the margin of error.

The data in Table 3.9 also shows that the error in calculating binding affinities is smallest with the MMPB/SA methods but increases with the Generalized Born methods. While the error with the MMGBMV/SA method may still be acceptable, the deviation with the MMGB/SA method is much larger than with the MMPB/SA method. The inclusion of explicit water molecules in the MMPB/SA method results in slightly larger relative deviations but better absolute agreement with the experimental data. Considering the overall uncertainties of the computational approach, the good absolute agreement with experiment for the ligand binding energies obtained with the hybrid explicit/implicit MMPB/SA method is remarkable.

The average deviations in Table 3.9 do not show whether the relative ranking of inhibitors matches the experiments. This information is most important for practical applications of computational ligand binding free energy calculations. The ranking based on the calculated binding affinities for the final assignment of enzyme protonation states is LPV>RTV>IDV>NFV=SQV>APV with the MMPB/SA method

and changes to IDV>LPV>RTV>SQV=NFV>APV when explicit waters are included in the MMPB/SA analysis. This result may be compared with the experimental ranking LPV>RTV>SQV>APV=IDV>NFV. The ranking from the calculations does not completely agree with the experimental data. In particular, binding to APV is underestimated and binding to NFV is overestimated. Furthermore, IDV binding is overestimated, especially with the hybrid explicit/implicit MMPB/SA method.

3.4 Discussion

In this study, protein-ligand binding affinity calculations following the established MMPB/SA scheme were applied to the well-studied example of HIV-1 protease inhibitor binding. It is found that the choice of enzyme protonation state significantly affects the results and that excellent agreement with experimental data is obtained when the protonation states are assigned carefully based on energetic and structural considerations. With the final assignment of protonation states (LPV: D25, RTV: D25', SQV: D25, IDV: D25, APV: D25', NFV: D25') relative binding affinities are calculated with a root mean square deviation of 3.4 kcal/mol. Such a degree of uncertainty may be sufficient for discriminating inhibitors from non-binding molecules. However, the ranking of inhibitors according to their experimental binding affinities is reproduced only to some part in the calculations. This indicates that the MMPB/SA protocol used here may be useful for identifying new inhibitors but might be challenged when applied to inhibitor optimization and in particular the design of new HIV-1 inhibitors that are more resistant to viral mutations.

There are a number of approximations in the MMPB/SA scheme that likely have an effect on the obtained results. First, a single trajectory for each inhibitor was used to sample both enzyme and ligand conformations. This approach neglects

possible conformational change of the enzyme and the inhibitor when dissociated but has the advantage of cancelling errors when obtaining free energy estimates as the difference between complex and enzyme energies. Previous studies have tried to address this issue, but could not reach sufficient convergence with nanosecond simulations.[96]

A critical issue with any force-field based quantitative studies is the choice of parameters, in particular the partial charges of the inhibitors. Compared to previous studies of the same systems, better absolute agreement with the experimental data was found in this study despite similar methodology. Although this may be fortuitous to some extent, it is mainly a function of the particular choice of partial charges for the inhibitors. Similar conclusions have also been made in different contexts.[154] A more robust framework would involve the use of QM/MM methods[86, 154] or at least the inclusion of atomic polarizability in order to allow inhibitor and protein charges to respond appropriately during complex formation.[155] It is likely that such methods are needed to substantially improve upon the level of accuracy reported here.

The MMPB/SA method samples enzyme complex conformations with explicit solvent but estimates the solvation contribution to the free energy of binding from an implicit model. This study tests whether the addition of selected explicit water molecules could improve the accuracy of the binding affinity estimates. The overall conclusion is that a hybrid explicit/implicit MMPB/SA scheme offers little advantage over the standard MMPB/SA when it comes to estimating relative binding affinities between different inhibitors. However, the absolute agreement with the experimental data is better with the hybrid scheme suggesting that the inclusion of explicit water molecules does add substantial contributions to the purely implicit model. It is also noteworthy that the hybrid method brings the D25' protonation state for the IDV

inhibitor into the range of possibilities. The pure implicit MMPB/SA scheme excludes that state based on energetics. Although D25 protonation is predicted for the complex with IDV based on the data in this study, several other studies have proposed D25' protonation.[92, 96, 97]

A related question is whether computationally more advantageous MMGB/SA methods could be used to obtain similar results as with the MMPB/SA method. The data from this study suggests that the use of the Generalized Born approximation does indeed present a compromise between speed and accuracy. With the GB method from Amber the results differ quantitatively and qualitatively. The absolute binding affinities are very far from the experimental values and even relative binding affinities deviate substantially. The newer and more expensive GBMV method from CHARMM fares significantly better by reaching similar qualitative results as with the MMPB/SA scheme, but with a slightly reduced accuracy in estimating relative binding affinities and larger deviations when absolute results are compared.

The focus of this study and others is to obtain binding affinity calculations for the single correct enzyme protonation state. The similar binding energies between different protonation states for some of the inhibitors (e.g. APV) indicate, however, that multiple protonation states could also be in equilibrium at physiological or experimental *pH*. While this possibility has not been addressed here, recently introduced constant-*pH* sampling methods could be used to examine this issue in more detail in future studies.[156, 157]

Finally, the results presented here may be compared with previous calculations. The most similar study is a calculation of binding free energies for five (RTV, SQV, APV, IDV, APV) out of the six inhibitors considered here.[90] based on an MMPB/SA analysis. The resulting binding affinities that were calculated without

the entropic contribution have root mean square deviations of 10.6 kcal/mol and 1.9 kcal/mol for absolute and relative comparisons, respectively. The relative binding affinities are in surprisingly good agreement with the experimental data considering the very short simulation time (200 ps) in that study and no apparent consideration of different enzyme protonation states depending on the bound inhibitor. Other, more recent studies have estimated absolute and relative binding free energies for SQV and IDV[96] and NFV, IDV, and APV[95] that are much further from the experimental data and differ much more between each other compared to experiment and the results found in this study.

The final prediction of protonation states agrees with previous studies in the case of APV and SQV, but disagrees in the case of NFV and IDV. This is not necessarily a concern since previous attempts to determine the enzyme protonation state with these inhibitors were based on minimization and approximate energetic analyses.[92, 96, 97] On the other hand, the results presented here are the consequence of a full MMPB/SA analysis for all protonation states and a comparison of structural features of the enzyme-inhibitor complex. All of the predicted protonation states are monoprotonated since diprotonation was excluded based on the assumption that the free enzyme is monoprotonated at D25 and that the energetic cost of adding a second proton to D25' in the free enzyme is relatively high on the order of 10 kcal/mol. Some studies have indicated that the diprotonated form may be preferred for some inhibitors.[111] The combined structural and energetic data presented here suggests this possibility only when forming a complex with IDV or SQV at reduced pH. However, the structural analysis alone would favor diprotonation for all of the inhibitors and it is conceivable that uncertainties in estimating the pK_a shift of D25' in

the free enzyme along with uncertainties in the binding affinity calculations may bring the equivalence point between mono- and diprotonation near the physiological range.

3.5 Conclusions

The main finding of this study is that accurate predictions of binding affinities for HIV-1 protease inhibitors are possible with an MMPB/SA-type free energy analysis when enzyme protonation states for each inhibitor are determined carefully. HIV-1 protease is predicted to maintain D25 protonation of the free enzyme when LPV, SQV, or IDV are bound, but is assumed to change protonation to D25' when the inhibitors RTV, NFV, or APV are bound. Furthermore, it is found that the inclusion of explicit water molecules in a hybrid implicit/explicit MMPB/SA scheme improves absolute agreement of calculated binding affinities with the experimental data. On the other hand, the use of the Generalized Born approximation significantly diminishes the accuracy of the obtained results.

Table 3.1 Absolute binding free energies of HIV-1 proteases in kcal/mol estimated from MMPB/SA and MMGB/SA schemes in comparison with experimental results (ΔGexp =-RTln Ki ; T=300K). Energies are averaged over 100 snapshots from the last 1 ns of molecular dynamics simulations of the enzyme-inhibitor complexes. Statistical errors calculated from standard deviations are given in parentheses. The monoprotonation state(s) with the lowest binding energies within the error estimates is indicated in bold face. Numbers in blanket refer to the free energy change from free D25 to D25,25', $\Delta G_{free(D25\rightarrow 25,25')}$ which defined in schematic 3.2.

Inhibitor	*Protonation state*	*MMPB /SA*	*MMPB/SA +explicit H_2O*	*MMGBMV/SA*	*MMGB/SA*	ΔG_{exp}
LPV	D25	**-14.7** *(0.8)*	**-13.1** *(1.1)*	**-8.53** *(1.3)*	**-59.5** *(0.6)*	-16.2
	D25'	-11.1 *(1.9)*	**-12.1** *(1.4)*	0.13 *(2.3)*	**-59.8** *(0.2)*	[158]
	D25,25'.	-12.8 *(0.0)*	-12.6 *(0.0)*	-3.72 *(0.6)*	-58.9 *(0.4)*	
		[-1.06]	[-0.86]			
	D-	13.2 *(0.1)*	13.5 *(0.4)*	16.8 *(1.1)*	-57.6 *(0.6)*	
RTV	D25	**-16.4** *(2.8)*	**-21.9** *(2.2)*	**-15.0** *(2.5)*	**-62.4** *(2.5)*	-14.9
	D25'	-9.19 *(0.0)*	-12.1 *(2.7)*	-3.51 *(0.7)*	**-61.9** *(2.4)*	[159]
	D25,25'	-11.1 *(0.5)*	-17.1 *(1.4)*	-5.55 *(0.4)*	-56.2 *(0.1)*	
		[-3.54]	[-4.54]			
	D-	19.3 *(3.6)*	12.0 *(4.1)*	23.3 *(2.2)*	-56.0 *(0.0)*	
SQV	D25	**-8.10** *(0.3)*	**-10.5** *(0.4)*	**2.90** *(1.2)*	-51.8 *(0.0)*	-14.3
	D25'	0.48 *(1.1)*	-8.36 *(1.3)*	5.87 *(0.6)*	**-55.8** *(0.2)*	[160]
	D25,25'	-10.8 *(0.9)*	-17.9 *(0.7)*	-4.95 *(1.0)*	-51.2 *(0.4)*	
		[0.07]	[-7.03]			
	D-	17.3 *(1.6)*	11.1 *(1.1)*	20.4 *(5.7)*	-54.9 *(0.0)*	
IDV	D25	**-10.9** *(0.3)*	**-17.6** *(0.9)*	**-5.24** *(0.8)*	**-64.9** *(0.5)*	-13.3
	D25'	-7.99 *(1.5)*	**-15.8** *(1.2)*	-1.15 *(1.0)*	-60.1 *(3.1)*	[131]
	D25,25'	-16.8 *(0.8)*	-22.2 *(0.6)*	-10.7 *(0.4)*	-60.2 *(0.3)*	
		[-4.46]	[-9.86]			
	D-	6.48 *(0.9)*	-3.73 *(0.3)*	15.9 *(0.2)*	-42.9 *(0.9)*	
APV	D25	**-0.69** *(0.3)*	**-4.55** *(0.7)*	**4.79** *(0.1)*	**-43.2** *(1.2)*	-13.1
	D25'	**-1.07** *(0.1)*	**-4.34** *(0.8)*	8.63 *(2.1)*	-37.5 *(1.8)*	[132]
	D25,25'	-6.96 *(2.7)*	-2.85 *(2.6)*	2.89 *(4.0)*	-40.8 *(2.2)*	
		[4.31]	[8.42]			
	D-	7.57 *(3.9)*	8.65 *(3.4)*	17.4 *(3.7)*	-28.2 *(0.0)*	
NFV	D25	-4.69 *(0.2)*	-7.52 *(0.1)*	2.33 *(0.0)*	**-48.2** *(0.2)*	-12.4
	D25'	**-8.18** *(0.8)*	**-10.2** *(1.1)*	**-2.14** *(2.0)*	-40.3 *(0.5)*	[133]
	D25,25'	-11.0 *(0.9)*	-11.2 *(0.4)*	-5.31 *(0.4)*	-51.0 *(0.7)*	
		[12.67]	[5.57]			
	D-	-4.10 *(0.6)*	-6.14 *(0.3)*	3.94 *(1.5)*	-39.2 *(0.1)*	

Table 3.2 Energetic contributions to MMPB(GB)/SA analysis in kcal/mol.

		Non-bonded Interaction in Vacuum	Polar Solvation Energy (PB)	Polar Solvation Energy (GBMV)	Polar Solvation Energy (GB)	Nonpolar Solvation Energy	-TΔS
LPV	D25	-119.8	95.5	101.7	50.7	-7.8	17.4
	D25'	-121.3	100.8	112.1	52.2	-8.0	17.3
	D25,25'	-123.5	101.3	110.4	55.1	-8.1	17.6
	D-	-130.5	133.1	136.7	63.3	-7.8	18.4
RTV	D25	-127.9	100.9	102.4	55.0	-8.1	18.6
	D25'	-125.0	104.9	110.5	52.2	-8.3	19.3
	D25,25'	-122.9	100.6	106.2	55.5	-8.4	19.6
	D-	-125.0	133.3	137.4	58.1	-8.5	19.5
SQV	D25	-111.4	91.7	102.7	48.0	-8.3	19.9
	D25'	-120.8	112.0	117.4	55.8	-8.2	17.4
	D25,25'	-122.2	99.9	105.8	59.5	-8.5	20
	D-	-135.9	140.7	144.1	68.5	-8.1	20.5
IDV	D25	-140.7	117.2	122.8	63.3	-8.2	20.8
	D25'	-128.0	109.3	116.1	57.3	-8.2	18.9
	D25,25'	-132.3	104.9	111.1	61.5	-8.2	18.8
	D-	-116.8	112.4	121.8	62.1	-8.1	19
APV	D25	-104.3	92.6	98.1	50.1	-6.7	17.8
	D25'	-96.0	82.8	92.8	46.4	-7.0	19.2
	D25,25'	-101.2	81.7	91.5	47.8	-7.1	19.7
	D-	-94.7	90.4	100.2	54.6	-6.4	18.4
NFV	D25	-99.3	83.7	90.7	40.2	-7.3	18.2
	D25'	-92.5	72.9	78.9	40.8	-7.0	18.5
	D25,25'	-114.0	92.7	98.3	52.7	-7.4	17.8
	D-	-107.8	91.5	99.6	55.4	-7.3	19.5

Table 3.3 Energetic contributions to MMPB/SA analysis with selected explicit water molecules in kcal/mol.

		Non-bonded Interaction in Vacuum	Polar Solvation Energy(PB)	Nonpolar Solvation Energy
LPV	D25	-128.8	106.3	-8.0
	D25'	-127.6	106.2	-8.0
	D25,25'	-124.6	102.5	-8.1
	D-	-130.6	133.5	-7.8
RTV	D25	-140.3	108.0	-8.2
	D25'	-125.9	102.7	-8.3
	D25,25'	-126.3	98.0	-8.4
	D-	-140.7	141.8	-8.6
SQV	D25	-114.2	92.2	-8.3
	D25'	-136.0	118.5	-8.3
	D25,25'	-140.5	111.1	-8.4
	D-	-137.4	136.1	-8.1
IDV	D25	-153.9	123.8	-8.2
	D25'	-141.3	114.9	-8.2
	D25,25'	-146.1	113.3	-8.2
	D-	-135.3	120.6	-8.1
APV	D25	-113.7	98.2	-6.8
	D25'	-109.2	92.8	-7.1
	D25,25'	-98.3	85.5	-9.8
	D-	-104.6	101.4	-6.5
NFV	D25	-114.5	96.2	-7.4
	D25'	-108.3	86.9	-7.2
	D25,25'	-122.2	100.7	-7.5
	D-	-129.3	111.3	-7.6

Table 3.4 Estimated free energy change in kcal/mol from free, monoprotonated (D25) HIV-1 protease to the diprotonated (D25,D25') ($\Delta G_{free(D25\rightarrow D25,25')}$ *at pH=7 and T=298 K*) form using data from simulations of different inhibitor complexes.

Inhibitor	*ΔG(AspH->Asp)*	*ΔG(EnzH->Enz)*	$pK_{a,shift}$	*ΔG (pH=7, T=298K)*
LPV	-12.12	-19.64	-5.51	**11.74**
RTV	-12.12	-15.45	-2.44	**7.56**
SQV	-12.12	-18.76	-4.87	**10.87**
IDV	-12.12	-20.23	-5.95	**12.34**
APV	-12.12	-19.16	-5.16	**11.27**
NFV	-12.12	-24.66	-9.20	**16.77**

Table 3.5 The binding affinity (ΔG_{total}) of the complexes for the D25,25' state calculated as a summation between $\Delta G_{free(D25\rightarrow D25,25')}$ and $\Delta G_{binding}$ according at a given pH of 3-7 using MMPB/SA.

	Drug	**PH=7**	PH=6	PH=5	PH=4	PH=3
	LPV	**-1.06**	-2.43	-3.79	-5.15	-6.52
dipro	RTV	**-3.54**	-4.91	-6.27	-7.64	-9.01
	SQV	**0.07**	-1.3	-2.66	-4.03	-5.39
	IDV	**-4.46**	-5.83	-7.19	-8.55	**-9.92**
	APV	**4.31**	2.94	1.57	0.21	-1.15
	NFV	**12.67**	11.3	9.94	8.58	7.21

Table 3.6 The binding affinity (ΔG_{total}) of the complexes for the D25,25' state calculated as a summation between $\Delta G_{free(D25\rightarrow D25,25')}$ and $\Delta G_{binding}$ according at a given pH of 3-7 using hybrid MMPB/SA.

	Drug	**PH=7**	PH=6	PH=5	PH=4	PH=3
	LPV	**-0.86**	-2.23	-3.59	-4.95	-6.32
dipro	RTV	**-4.54**	-5.91	-7.27	-8.64	-10.01
	SQV	**-7.03**	-8.4	**-9.76**	-11.13	-12.49
	IDV	**-9.86**	-11.23	-12.59	-13.95	**-15.32**
	APV	**8.42**	7.05	5.68	4.32	2.96
	NFV	**5.57**	4.2	2.84	1.48	0.11

Table 3.7 Average root mean square deviation in Å for catalytic aspartate (D25 and D25') side chains from snapshots during the last 1ns of the molecular dynamics simulations compared to the crystallographic structures of the complexes with the respective inhibitors. The entire complex was fitted to the experimental structure before deviations were calculated. The minimum RMSD values for monoprotonated states are highlighted in bold.

	D25	*D25'*	*D25,D25'*	*D-*
LPV	**1.68**	1.69	0.86	1.13
RTV	1.22	**1.15**	0.92	0.68
SQV	**1.11**	1.20	0.69	1.35
IDV	**1.22**	1.49	0.81	0.94
APV	1.33	**0.99**	0.95	1.25
NFV	1.50	**1.19**	0.67	1.53

Table 3.8 Prediction of HIV-1 protease protonation states in complex with different inhibitors based on energetic and structural analysis.[92, 96, 97] '++' indicates the most favorable binding energies. '+' indicates less favorable binding energies within the estimated uncertainty intervals. 'o' denotes that conformational sampling of inhibitors is compatible with the X-ray structure. Average RMSD values for catalytic aspartates below 1.0 Å are indicated with 'oo' and values between 1 Å and 1.25 Å are indicated

with 'o'.

		MMPB/SA	MMPB/SA +expl. H_2O	Inhibitor structure	Catalytic aspartates	Consensus	Other studies
LPV	*D25.*	++	++	o	-	*	
	D25'		+	-	-		
	D25,25'			o	oo		
	D-			-	o		
RTV	*D25*	++	++	-	o	?	
	D25'			o	o	?	
	D25,25'			o	oo		
	D-			o	oo		
SQV	*D25*	++	++	o	o	**	X[96]
	D25'			-	o		X[92]
	D25,25'		+ (low pH)	o	oo	* (low pH)	
	D-			-	-		
IDV	*D25*	++	++	o	o	**	
	D25'		+	o	-	*	X[92, 96, 97]
	D25,25'	+ (low pH)	+ (low pH)	o	oo	* (low pH)	
	D-			o	oo		
APV	*D25*	+	++	-	-		
	D25'	++	+	o	oo	**	X[97]
	D25,25'			o	oo		
	D-			-	-		
NFV	*D25*			-	-		X[97]
	D25'	++	++	o	o	**	
	D25,25'			o	oo		
	D-			-	-		

Table 3.9 Root mean square deviations in kcal/mol between calculated and experimental binding affinities for different enzyme protonation states. Both absolute (first value) and relative deviations (second value) are given. Relative deviations were evaluated after subtracting the difference between the experimental and calculated average binding affinities for all inhibitors to allow for a constant offset between experiment and calculation. 'Consensus' protonation means LPV: D25, RTV: D25 or D25', SQV: D25, IDV: D25 or D25', APV: D25', and NFV: D25'.

	Consensus IDV:D25, RTV:D25	Consensus IDV:D25, RTV:D25'	Consensus IDV:D25', RTV:D25	Consensus IDV:D25', RTV:D25'
MMPB/SA	5.9/4.2	6.3/3.4	6.2/4.2	6.6/3.2
MMPB/SA+expl. H20	5.4/5.3	4.7/3.8	5.2/5.0	4.5/3.3
MMGBMV/SA	12.9/7.0	13.7/5.1	13.4/6.9	14.2/4.7
MMGB/SA	39.9/9.9	39.8/9.8	38.9/9.0	38.8/8.9

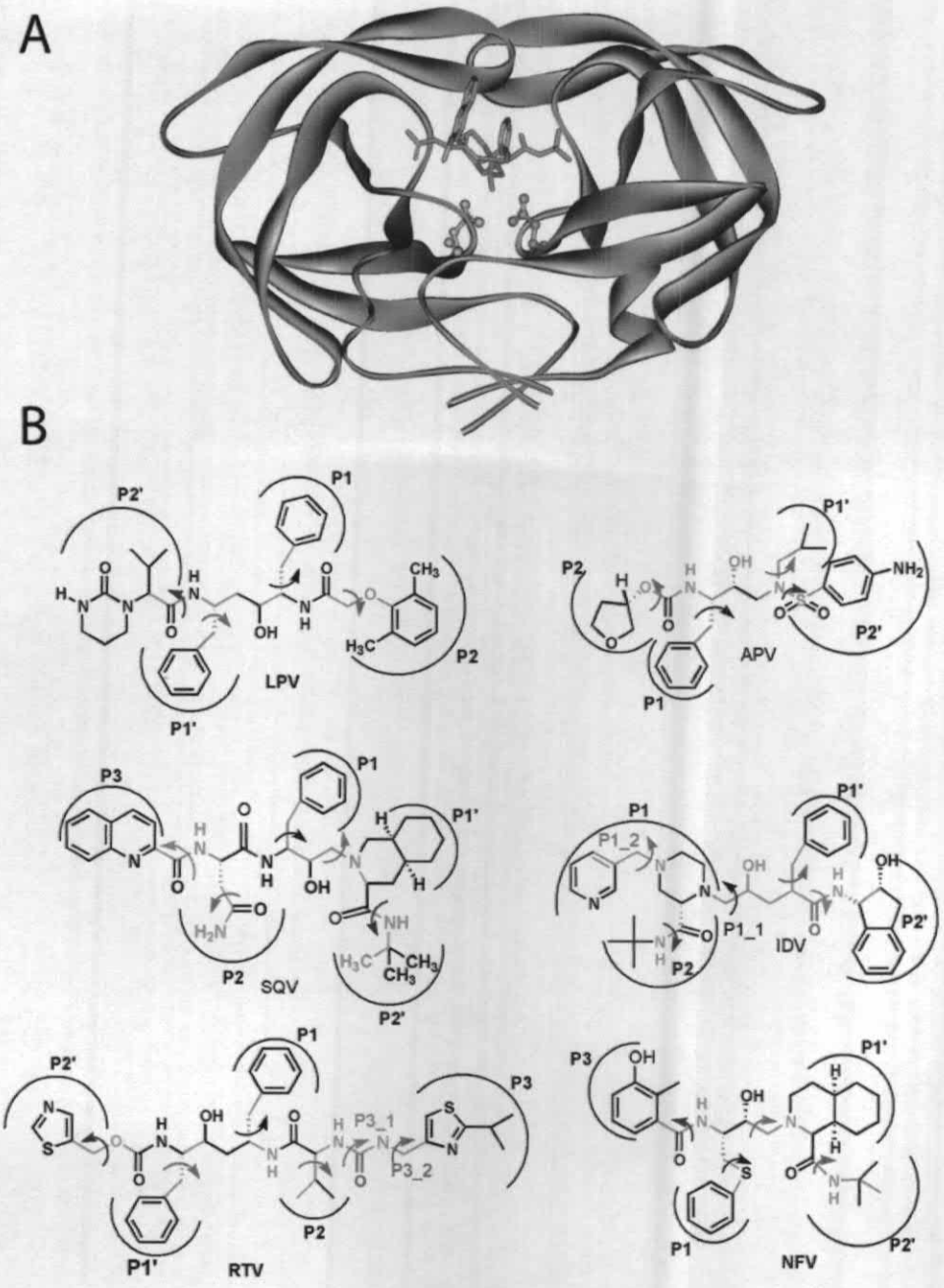


Figure 3.1 (A) Structure of HIV-1 protease in complex with inhibitor (shown in blue). (B) Structures of inhibitors considered in this study: lopinavir (LPV), ritonavir (RTV), saquinavir (SQV), indinavir (IDV), amprenavir (APV), and nelfinavir (NFV). Subsites are labeled as P1, P2 etc. Selected torsion angles discussed in the text are shown in color and indicated by arrows.

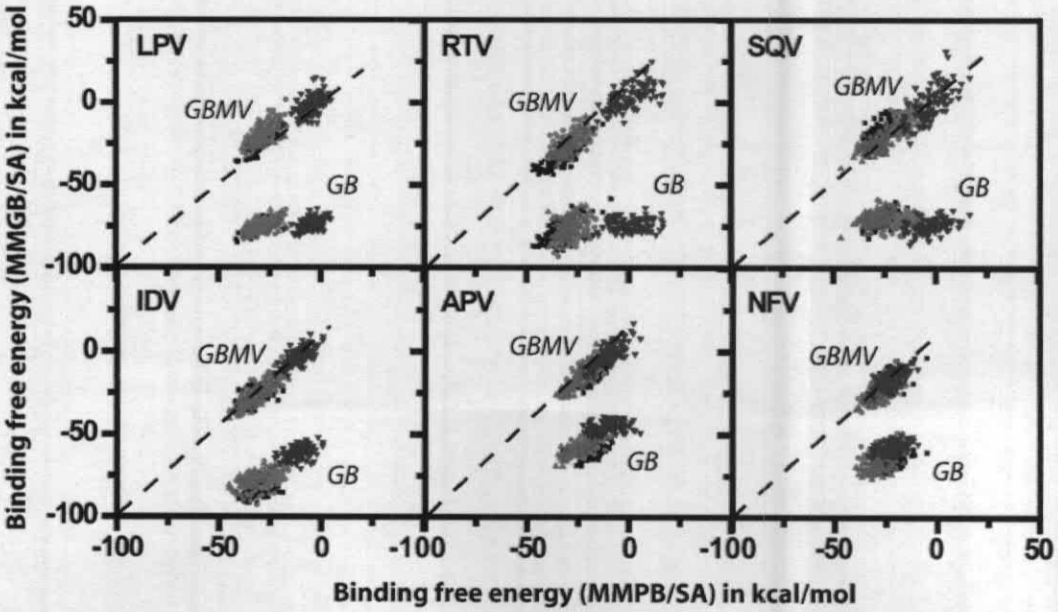


Figure 3. 2 Comparison of calculated binding free energies between MMPB/SA and MMGB/SA approaches. Colors indicate results for different protonation states: D25 (black), D25' (red), D25,25' (green), D(-) blue.

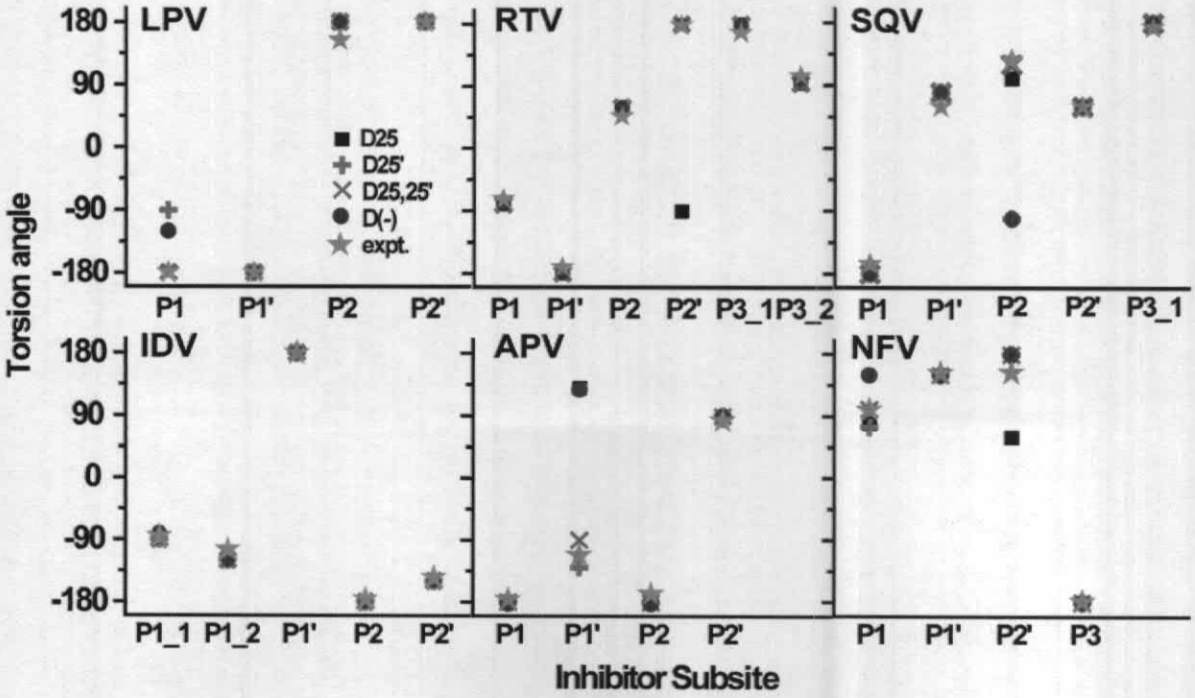


Figure 3.3 Sampling of inhibitor torsion angles as identified in Fig. 1 during simulations of HIV-1 protease inhibitor complexes with different enzyme protonation states. Only the dominant torsion angles are shown and compared with the angles found in the respective crystal structures of HIV-1 protease with each of the inhibitors.