Published in final edited form as:

J Chem Theory Comput. 2018 May 08; 14(5): 2784–2796. doi:10.1021/acs.jctc.8b00097.

Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease

Florian Leidner¹, Nese Kurt Yilmaz¹, Janet Paulsen¹, Yves A. Muller², and Celia A. Schiffer^{1,*}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

²Division of Biotechnology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Abstract

Water is an essential in many biological processes, and the hydration structure plays a critical role in facilitating protein folding, dynamics and ligand binding. A variety of biophysical spectroscopic techniques have been used to probe the water solvating proteins, often complemented with molecular dynamics (MD) simulations to resolve the spatial and dynamic features of the hydration shell, but comparing relative water structure is challenging. In this study 1 microsecond MD simulations were performed to identify and characterize hydration sites around HIV-1 protease bound to an inhibitor, darunavir (DRV). The water density, hydration site occupancy, extent and anisotropy of fluctuations, coordinated water molecules, and hydrogen bonds were characterized and compared to the properties of bulk water. The water density of the principal hydration shell was found to be higher than bulk, dependent on the topology and physio-chemical identity of the biomolecular surface. The dynamics of water molecules occupying principal hydration sites was highly dependent on the number of water-water interactions, and inversely correlated with hydrogen bonds to the protein–inhibitor complex. While many waters were conserved following the symmetry of homodimeric HIV protease, the asymmetry induced by DRV resulted in asymmetric lower-occupancy hydration sites at the concave surface of the active site. Key interactions between water molecules and the protease, that stabilize the protein in the inhibited form, were altered in a drug resistant variant of the protease indicating that modulation of solvent–solute interactions might play a key role in conveying drug resistance. Our analysis provides insights into the interplay between an enzyme inhibitor complex and the hydration shell and has implications in elucidating water structure in a variety of biological processes and applications including ligand binding, inhibitor design and resistance.

*Corresponding author: celia.schiffer@umassmed.edu.

Supporting Information

Root mean square deviation of the HIV-1 Protease C α atoms is presented for all ten 100ns simulations (Figure S1). The dependence of the number of hydration sites on density cutoff and total simulation time is presented in Figure S2. In Figure S3 the nearest neighbor distance distribution is shown for crystallographic water molecules in the 1t3r structure (WT HIV-1 Protease in complex with DRV) and hydration sites predicted from MD simulations of the same structure. The interdependence between protein interactions and consensus between experimental and computational waters is demonstrated and outliers are analyzed. In Figure S4 the correlation between hydration site properties presented in Figure 3 is shown as a scatterplot. Table S1 summarizes occupancy and hydrogen bonds made by high occupancy hydration sites shown in Figure 5 & 6. Table S2 shows conservation of HS-1A & HS-1B in different retroviral proteases. This information is available free of charge via the Internet at <http://pubs.acs.org>

Introduction

Water is ubiquitous in all aspects of life, consequently evolution has steered biomolecules to adapt and integrate their aqueous surrounding. As such, the interplay between biomolecules as solute and water as a solvent is pivotal to many biomolecular processes. In addition to being involved in chemical reactions catalyzed by enzymes, water molecules mediate non-covalent interactions to drive biomolecular folding and binding processes. In protein folding, water plays a key role both as a driving force during the hydrophobic collapse of the newly synthesized protein chain and by facilitating intramolecular interactions.^{1, 2} Similarly, interactions between proteins, small molecules, and nucleic acids utilize solvent molecules to either lower the entropic costs or create favorable enthalpic interactions.³

The principal interactions between water and proteins have been studied by a plethora of different techniques. NMR and a variety of biophysical spectroscopic techniques have been used to describe water dynamics at biomolecular surfaces.^{4, 5} The consensus from these studies is that compared to bulk water, dynamics around biomolecules slow down significantly, and that this change persists for several layers of hydration beyond the protein surface. Most experimental techniques lack the ability to report on the spatial resolution of individual water molecules but measure bulk properties instead.⁶ On the single particle level, molecular dynamics (MD) simulations have been used to analyze distribution and dynamics of protein hydration, often supplementing experimental techniques.^{7–10} Due to water's significance in ligand binding and in-silico screening, several computational techniques have been developed to predict the thermodynamic profile of structural waters on the protein surface.^{11–13}

High-resolution x-ray crystal structures of proteins provide spatial distribution of water molecules surrounding the protein in the crystalline state. However, most of these water molecules cannot be detected in solution by NMR dispersion experiments due to rapid exchange with bulk solvent. In contrast, slowly exchanging but spatially less restrained water molecules detected by NMR may not be resolved by x-ray crystallography.¹⁴ This lack of correlation between spatial and temporal distribution of water molecules has also been observed in MD simulations.¹⁵ MD analysis can provide both spatial and dynamic features of the water molecules, providing insights into apparent discrepancies in experimental measurements and enabling a detailed analysis of the hydration shell of proteins.

HIV-1 protease is an aspartyl protease that is essential for the maturation of HIV and a key target in antiretroviral therapy. This homodimeric viral protein is comprised of two chains of 99 amino acids each, with the active site located between the two monomers. (Figure 1A) There are 10 FDA-approved drugs targeting this viral protease, all of which are transition-state-mimic competitive inhibitors binding at the active site. Darunavir (DRV) is the most recently approved and most potent inhibitor with low picomolar inhibition constant. Upon binding DRV, the otherwise highly dynamic protease “flaps” close down to lock in the inhibitor and dynamically restrict the protease.^{16–18} A conserved buried water molecule, revealed by x-ray crystal structures of HIV-1 protease bound to both substrates and

peptidomimetic FDA-approved inhibitors, coordinates the intermolecular interactions between the protease flaps and the ligand in the bound state, and is termed the “flap water”. In addition, several other conserved water molecules have been identified in crystal structures of both wild type and highly mutated drug resistant protease variants; however, the crystallographic waters were overall highly variable depending on the crystal lattice, protease variant, and the ligand bound.^{19–21} Although likely crucial for inhibitor binding, the dynamic and spatial properties of hydration layer around HIV-1 protease, as well as any asymmetry induced by inhibitor binding to this symmetric protease has not been explored.

In this work, we present a molecular dynamics based protocol to describe the solvent structure around HIV-1 protease bound to DRV. From the water density distribution principal sites of hydration were resolved as volume elements centered on local maxima. Hydration sites were analyzed in terms of water density, occupancy, hydrogen bonds, and protein versus water interactions. The analysis reveals how the binding of the asymmetric inhibitor, DRV, is reconstituted in the solvent structure. We describe key interactions between DRV-bound protease and highly localized water molecules. Comparison of the wildtype hydration structure with a drug resistant protease variant revealed the impact of resistance mutations on these key interactions and active site hydration.

Results

Identification of Hydration Sites

The average water density distribution for HIV-1 protease was reconstructed from ten 100 ns molecular dynamics simulation trajectories. All simulations reached convergence and were stable (Figure S1). A rectangular grid of equidistant points was constructed for the whole simulation box containing the protein, inhibitor and water molecules. The spacing between grid points was 0.5 Å. For each grid point the average water density (ρ) was calculated over the ten trajectories. Of the grid points occupied by water molecules, 92% had a density within 50% of the density expected for bulk water ($\rho_{\text{bulk}} = 0.033 \text{ Å}^{-3}$), and 0.4% had a density that was at least twice the value of ρ_{bulk} . These high-density cells ($\rho > 2\rho_{\text{bulk}}$), displayed in Figure 1B, formed defined peaks in the 3D water density distribution. Cross-sections through the protease (Figure 1D and E) showed high solvent density around the protein surface relative to bulk water, and a number of well-defined water molecules in the protein interior including the flap water that is key for inhibitor binding.

To analyze the spatial distribution and describe the properties of water quantitatively, discrete peaks in the water density were assigned as hydration sites. First, individual grid cells that correspond to local maxima in the water density were identified, then those that were less than 1.5 Å apart were combined into a single hydration site (see Methods for details). The assignment of maxima as hydration sites necessitated the definition of a minimum threshold for water density to minimize noise from regions of homogeneous density. A density threshold of $\rho_N/\rho_0 = 2$ was chosen, at which isolated regions of water density could be observed while only losing minimal information in the regions around the protein surface.

A total of 347 hydration sites were identified at the $\rho_N/\rho_0 = 2$ threshold. To evaluate the impact of simulation time and density threshold on the assignment of hydration sites, the calculations were repeated on non-overlapping subsets of the entire ten 100 ns simulations. The density threshold was increased incrementally for each subset, and the number of hydration sites calculated (Figure S2). The number of identified hydration sites converged rapidly, and even at the minimum threshold of $\rho_N/\rho_0 = 2$ including more than three 100 ns simulations did not produce any significant changes in the number of identified hydration sites. In subsequent analysis, the full set of ten 100 ns simulations (total of 1 μ s) was used. In addition to testing internal convergence, the nearest neighbor distance was calculated between water molecules in the crystal structure and hydration sites. Hydration sites were largely located within close proximity to experimental crystallographic waters in the DRV-bound protease structure. (Figure S3A). The consensus between computational and experimental water sites increased with stronger interactions with the protein surface residues (Figure S3B). In all cases the majority of hydration sites had a crystallographic water molecule located within 1.73 Å which corresponds to the minimum distance at which points can be distinguished on a rectangular mesh with 0.5 Å spacing. Cases where there were no crystallographic water molecules were within 4 Å were examined individually. All outliers were found to either correspond to hydration sites located at crystallographic interfaces (Figure S3C) or cases where the solvation of an alternative conformation was not modeled in the crystal structure (Figure S3D).

Surface Topology and Hydration Sites

The hydration sites identified above were all proximal to the protein surface, with no site being further than 4.9 Å away from the protease–inhibitor complex, and 274 sites located within 3 Å of the reference structure. The spatial distribution of high-density hydration sites had an irregular pattern around the protease surface (Figure 1B and C). Regions of convex surface topology were deprived while concave surfaces were enriched in high-density water. Convex surfaces are largely formed by the relatively long side chains of arginine and lysine residues protruding from the protein surface. Despite relatively high electrostatic potential and the ability to form hydrogen bonds with water molecules, the water density around these side chains is not well resolved, which can be attributed to high structural flexibility and solvent accessibility, which favors a more dynamic water network.

In contrast, concave surface regions had higher water density and corresponding hydration sites (Figure 1B). Regions of concave topology can be subdivided into 3 classes: holes, channels and ridges. Holes are surface inversions occupied by a well-differentiated region of high water density. The density peak at the active site of the protease located between the inhibitor and the flaps is a principal example of a hole that contains a hydration site. Ridges are formed either by side chain protrusions or by proximal protein chains. Examples are the flap tips and the ridge between the active site water channel and the flap elbows. These ridges are relatively narrow and occupied by “beads on a string” of high water density. Channels are similar to ridges in that they are stretched regions that are limited by surface protrusions. In HIV-1 protease, there is a large surface channel region extending from the active site toward the termini. The surface of this channel is largely formed by the single alpha helix of the protease monomers, and limited by the active site loop and the terminal

region. The channel in both monomers is defined by a high local water density, and the density peaks in the channel are more connected with each other indicating relative mobility. In general, the concave surface topology limits diffusion of water molecules in the hydration layer, restricting the ability of the water to rearrange, and offers a broader surface for solvent-solute interactions.²² Overall, the protein surface topology considerably affected the spatial distribution of hydration sites.

Characterization of Hydration Sites

To characterize the hydration structure and dynamics of HIV-1 protease, key properties of the hydration sites were determined for the occupying water molecules. The results are visualized in Figure 2 where the hydration site properties are displayed as histograms and compared to the corresponding value for bulk water.

Occupancy—Although all hydration sites identified had been chosen to have mean water density at least twice that of bulk water, the fraction of time the hydration site was occupied by a water molecule (mean occupancy) varied from 1 for buried water molecules to as low as 0.16 (Figure 2A). HS-1A (Figure 6) was consistently occupied by a water molecule in all 10 trajectories. The occupancy distribution has a sharp peak around the mean occupancy of 0.59, which is larger than that calculated for bulk water, 0.45. Of the 347 hydration sites resolved at the chosen density threshold, 202 sites had occupancy values between 0.4 and 0.6 over the simulations. The distribution is skewed to the left, with 136 sites having occupancy over 0.6 and 8 sites over 0.9, while only 9 sites had occupancy less than 0.4.

Hydrogen Bonds Formed by Waters Occupying the Hydration Sites—The total number of hydrogen bonds the water molecules made while occupying a given hydration site was determined, and the average value calculated over the simulations. The number of hydrogen bonds had a broad distribution with a mean around 3.4, lower than 3.6 of bulk water (Figure 2B). The value we calculated for bulk water agrees with the reported 3.5 hydrogen bonds formed by TIP3P in bulk under similar conditions of 1 atm pressure and 298 K.²³ The lower number of hydrogen bonds for water molecules occupying hydration sites compared to bulk water can be attributed to the chemically heterogeneous composition of the protein surface that limits the number of formed bonds. The fraction of hydrogen bonds with either the protein or the ligand was quantified in contrast with hydrogen bonds formed between water molecules. The fraction of hydrogen bonds formed with the protein–ligand complex had a mean of 0.25, indicating that on average one of the four possible hydrogen bonds is formed with a protein residue (Figure 2C). Of the 347 hydration sites, 313 formed less than 50% of their hydrogen bonds with the protein–ligand complex. 61 sites interacted almost exclusively with other water molecules, forming less than 10% of their hydrogen bonds with the protein–ligand complex. In contrast, the number of hydration sites interacting predominantly with the solute is significantly lower, with 6 sites forming over 70% and only 3 sites over 90% of their hydrogen bonds with the protein–ligand. The fraction of hydrogen bonds between a water molecule present at a hydration site and the protein provides a measurement of how strongly the site interacts with the protein. Nevertheless, the correlation between occupancy and fraction of hydrogen bonds with protein–ligand complex was weak (Pearson correlation = 0.31; Figure 3 and S4).

Neighboring Water Molecules Around Hydration Sites—Next, solvent accessibility of hydration sites was assessed by determining the number of water molecules around the water occupying the hydration sites. The oxygen atoms within 4 Å of the central water molecule occupying the site were counted and averaged over the trajectories (Figure 2D). Under the conditions of the simulations (300 K, 1 atm, $\rho_{\text{bulk}} = 0.033 \text{ Å}^{-3}$), on average a sphere of radius 4 Å, corresponding to a volume of 268 Å^3 , would contain 8.8 water molecules in bulk, or 7.8 water molecules surrounding a central water molecule.

The average number of neighboring water molecules was 5.3 for HIV-1 protease hydration sites, significantly lower than the bulk water value (7.8), as expected due to exclusion by the protein–ligand complex. More than half of the sites (60%) had between 4 and 7 water neighboring water molecules and the decrease in coordinated waters correlated with hydrogen bonds with the protein complex (Figure 3 and S4). The number of neighboring waters had a one-sided distribution with a sharp increase toward the mean 5.2 neighboring water molecules and a relatively abrupt decay once the peak value has been reached. The later observation can be readily explained through the rapid increase in repulsive non-bonded forces in a crowded environment. While the number of surrounding water molecules is strongly anti-correlated with the number of protein hydrogen bonds ($r = -0.79$ Figure 3 and S4) it is worth mentioning that the majority of hydration sites surrounded by more than 7 water molecules were forming at least one hydrogen bond with either the highly hydrophilic side-chains of aspartic acid and glutamic acid or the negatively charged carboxy group of the C termini. This relative enrichment of ordered solvation sites around negatively charged groups has recently been pointed out in crystallographic studies.^{24, 25} Only 3 sites were completely buried and had on the average less than 1 water molecule, which were the same 3 sites mentioned above to have over 90% of their hydrogen bonds with the protein–ligand complex. Thus, the neighboring water molecules around hydration sites negatively correlated with the extent of interactions with the protein–ligand complex.

Dynamics of Water Molecules at Hydration Sites—The overall dynamics of water molecules occupying hydration sites were examined by calculating the root-mean-square fluctuations (RMSF) of water oxygen atoms around their mean position. The RMSF values have a narrow distribution around 1.0 Å, with the majority of hydration sites being occupied by water molecules with dampened dynamics relative to bulk water (Figure 2E). 241 hydration sites had RMSF values between 1.0 and 1.2 Å, while the remaining 106 sites had RMSF values below 1.0 Å. The distribution had a long tail toward the minimum observed fluctuation of 0.56 Å for the highly buried sites.

The dynamics of water molecules occupying hydration sites as probed by RMSF values were highly correlated with the number of coordinated water molecules around the site (Pearson correlation $r = 0.80$; Figure 3 and S4). In reverse, sites with more hydrogen bonds with the protein–ligand complex had smaller RMSF scores ($r = -0.79$), reflecting dampening of water fluctuations due to the protein surface. There was a weaker negative correlation between water dynamics and occupancy ($r = -0.43$). For sites that were occupied around the mean occupancy of 0.59, the RMSF values were also around the mean, whereas sites that deviated from the mean occupancy were in general less dynamic.

To reveal whether there was any directionality to the dynamics of the water molecules imposed by the protein complex, next the anisotropy of water molecule fluctuations at hydration sites were calculated. For bulk water, as fluctuations in all directions are equally probable, the anisotropy of RMSF is exactly 1. For the hydration sites, the mean anisotropy was 1.35, with almost all (344 out of 347) hydration sites having anisotropy above 1.1. For 53 sites, the average anisotropy was above 1.5, and of those 7 had an average anisotropy above 2.0 indicating highly directional dynamics. All sites with higher than mean anisotropy values were located proximal to the protein surface forming 2 or more hydrogen bonds with the protein–ligand complex.

Overall, in comparison to the bulk water, dynamics of water molecules in the principal hydration layer were moderately perturbed by the presence of the protein–ligand complex, with lower fluctuations and a slight anisotropy induced onto these fluctuations. However, with the exception of some highly buried sites, water molecules proximal to the protein surface were still highly dynamic.

Symmetry of HIV-1 Protease Hydration Sites

HIV-1 protease forms a C_2 symmetric homodimer; however, this symmetry is perturbed when an asymmetric ligand, either substrate or inhibitor, is bound at the active site. Our previous crystal structure and MD simulation studies showed that binding of either the substrate or the peptidomimetic inhibitor DRV causes differences in structure and dynamics of the otherwise equivalent monomers, both through structural rearrangements and the differential protonation of the catalytic dyad.^{20, 26, 27} Since the structure of the hydration layer is dictated by the structure of the encased solute, we expected the hydration sites to reflect the pseudosymmetry of the inhibitor-bound enzyme. Consequently, unique hydration sites that do not have symmetric partners can be directly linked to the binding of the inhibitor whereas symmetric sites would be due to the overall symmetry of the system.

To evaluate the symmetry of the hydration sites, first, the distance between each site and the protein Ca atoms in the homodimer was determined defining the position of the hydration site relative to the protease. Subsequently the similarity of two given sites was evaluated by calculating the correlation of the distance vectors of the sites by swapping the protease monomer identities in the distance vector for one of the sites (see Methods). A pair of hydration sites was considered to be symmetric if their correlation was above the threshold value of 0.95 and the highest of all pairwise combinations that included one of the sites. Using this method, 145 symmetric pairs were identified in addition to 2 sites that were located on the symmetry axis, indicating that 84% of the hydration sites were conserved in both monomers (Figure 4). When superimposed across the symmetry axis, the root-mean-square deviation (rmsd) of the symmetric hydration sites was 0.98 Å, consistent with the 0.58 Å RMSD of the Ca atoms and the resolution of the grid used to identify the hydration sites. In addition, the average occupancy of symmetric hydration sites proximal to both monomers was the same (about 0.6), except for 9 sites around the active site which had differences in occupancy greater than 0.15. Hence, most of the hydration sites were symmetric with similar occupancy and position relative to the protease, reflecting the overall symmetry in the HIV-1 protease structure.

Apart from the symmetric sites, there were 55 hydration sites that did not have a symmetry partner. A hydration site containing the flap water in the center of the active site had almost equivalent distance vectors to both monomers and thus was located on the symmetry axis. This hydration site is situated to form a total of four hydrogen bonds with the inhibitor and residue 50 of the protease flaps, interactions that are essential for substrate and peptidomimetic inhibitor binding.²⁸ A second site was located on the symmetry axis on top of the protease flaps. Of the remaining asymmetric sites, 21 were located proximal to monomer A, which is the monomer that interacts with the large bis-THF moiety of DRV, whereas 34 were located proximal to monomer B (designated as the prime monomer, with residue numbers indicated with the prime symbol; Figure 5). Asymmetric hydration sites were distributed across the protease surface, indicating that the structural rearrangements induced by inhibitor binding propagated through the protein.

Highly occupied asymmetric sites were however concentrated around the active site. A closer inspection of the DRV-bound HIV-1 protease reveals asymmetry at the active site surface, with a concave topology at the aniline side and convex topology at the bis-THF side of the inhibitor (Figure 5A,B). This difference in surface topology significantly altered the symmetry of hydration structure around the active site. More water molecules could be accommodated at the aniline side of DRV due to the concave shape and the additional hydrogen bond donor in the form of the aniline amino group, resulting in a larger number of high occupancy (> 0.6) hydration sites. (Figure 5C,D). The larger solvent exposed surface area of the aniline moiety also reduced the spatial restraints imposed on the hydration sites leading on average to a less ordered hydration structure. On the other side, the bulky bis-THF and phenyl groups formed more extensive contacts with the protease monomers compacting the active site surface resulting in no high-occupancy asymmetric hydration sites at this site (Figure 5E,F).

Hydration Site Clusters and Interactions with Protease Active Site

With large asymmetry in hydration structure being confined to the region surrounding the inhibitor binding site, our subsequent analysis focused on the role of hydration sites surrounding the active site. At the bis-THF side of DRV, a central and tightly packed cluster was formed by 4 high occupancy hydration sites (HS-3A HS-4A HS-5A and HS-6A; Table S1 and Figure 6A). Water molecules occupying two of these sites (HS-4A and HS-5A) hydrogen bonded with the side chains of D29 and R8', which form a stable salt bridge connecting protease monomers and pack tightly against the bis-THF and phenyl moieties of DRV. Water in the other two hydration sites of the cluster (HS-3A and HS-6A) hydrogen bonded with the backbone atoms of G48 located in the flap. Due to their spatial proximity, the 4 hydration sites frequently formed a tetracyclic network of hydrogen bonds connecting D29, R8' and G48 during the MD simulations (Figure 6A). On the aniline side of DRV, the symmetry-related 4 sites (HS-3B to 6B) also interacted with the corresponding residues of the other monomer (Table S1). However, the occupancy of these sites was significantly lower and water molecules arranged in a wedge rather than a teracycle. These differences can be attributed to the divergent conformation of the D29'–R8 salt bridge on this side, which did not pack against the concave aniline face of DRV but instead faced away from the protein–inhibitor interface. As a result, sites HS-4B and HS-5B, which were interacting with

D29 and R8, were located more than 5 Å away from HS-3B and HS-6B and thus did not form a cluster, instead forming a diffusive network of hydrogen bonds with two other asymmetric hydration sites (HS-101 and HS-102) that occupied the space between the D29'–R8 salt bridge and the aniline face of DRV (Figure 6B).

Close to the central water cluster, a single symmetric hydration site (HS-1A; Table S1 and Figure 6C) was located in the interstitial cavity formed by G27, A28, D29, R87, R8', P9' and L23'. Being fully shielded from bulk solvent, HS-1A was occupied permanently by a single water molecule in all ten simulations, while water occupying the symmetry-related HS-1B was observed to exchange with the water molecule occupying the asymmetric hydration site HS-101 (Figure 6D, Table S1). Both HS-1A and HS-1B waters formed hydrogen bonds with the side chains of residue D29 and R87 as well as the backbone oxygen of G27. HS-1B alternated between forming a hydrogen bond with the side chain of D29' as depicted for HS-1A in Figure 6C, and forming a hydrogen bond with the backbone oxygen of T26' (Figure 6D). Despite being localized at the dimer interface and being able to accept an additional hydrogen bond, neither HS-1B nor HS-1A mediated interactions between the two monomers as neither the side chain of the proximal L23' side chain nor the backbone of P9' are able to provide a hydrogen for the formation of such an interaction. Instead, this hydration site likely acts by stabilizing the G27 A28 D29 loop as well as mediating interactions between D29 and R87. The exact orientation of the D29 carboxy group has been suggested to be essential for the formation of the active site dimerization interface^{29–31}. The structural importance of this hydration site in stabilizing R87–D29 interactions is further substantiated by the fact that comparable crystallographic water molecules can be found in crystal structures of many retroviral protease in both ligand bound and apo form. (Table S2)

On the other side of the active site, two symmetric hydration sites, (HS-2A and HS-2B; Figure 6E and F) were located in the cleft between the protease flaps, the anti-parallel beta-sheet formed by residues 43 to 58, and the loop formed by residues 78' to 82' (hereafter referred to as the 80s loop). The HS-2A hydration site was confined by the side chains of P79, P81, I54 and I50', while HS-2B with the corresponding residues in the other protease monomer. While HS-2A and HS-2B were symmetric in regards to their position relative to the protease, they had significant differences in terms of occupancy, with HS-2A being occupied 72% and HS-2B 93% of the time. Water molecules occupying both HS2A and HS2B formed two hydrogen bonds with the protein, connecting the flaps with the loop formed by residues 78–82. Compared to HS-2A, hydrogen bonds formed between HS-2B and the P79 oxygen in the 80s loop were significantly stronger. HS-2B was further stabilized by two asymmetric hydration sites, HS-201 and HS-202, with HS-201 forming a hydrogen bond with the backbone oxygen of G49 and HS-202 interacting with a number of medium occupancy sites located around the protease flaps (Figure 6F). HS-2A is situated above the phenyl moiety, and HS-2B is located above the smaller isopropyl moiety, which allows tighter packing of the 80s loop against the protease flaps. Despite not having any direct interaction with the inhibitor (both HS-2A and HS-2B were more than 8 Å away from any DRV heavy atoms), both hydration sites were positioned to bridge the protease flaps with the 80s loop of the active site to stabilize the protease in the closed inhibited conformation.

Changes in hydration structure due to drug resistance mutations

Analysis on wild-type HIV-1 protease revealed specific hydration sites clustered around the active site interfacing the inhibitor with the protease. Many of the hydrophobic residues that confine these hydration site clusters at the inhibitor interface are known to mutate and cause drug resistance by modulating the shape of the active site.^{32, 33} While a large portion of the effects of these mutations was attributed to the loss of van der Waals contacts between the protease and competitive inhibitors, it is likely that these changes also alter the cluster of hydration sites around the active site and the network of hydrogen bonds that complements and stabilizes the closed inhibited conformation of HIV-1 protease. To investigate the effect of resistance mutations, we applied the same methodology to a well-characterized drug resistant variant of HIV-protease.^{34–37} The resistant variant Flap+ contains 4 mutations relative to wild-type HIV-1 protease (L10I, G48V, I54V, V82A), 2 of which are located at the active site (Figure 7). DRV loses 5.8-fold potency against the Flap+ variant compared to wild-type protease, with unusual entropy-enthalpy compensation in binding thermodynamics resulting in a complete inversion of the enthalpic and entropic components driving binding.³⁴ While DRV binding to wildtype protease is enthalpically driven, binding to Flap+ was endothermic with a favorable entropic term. This dramatic inversion in binding thermodynamics has been suggested to be related to the changes in solvation.³⁴

To investigate the effect of mutations, hydration structure of darunavir-bound Flap+ variant was analyzed. Hydration sites were calculated following the same protocol applied to the wildtype complex, and 353 hydration sites were resolved for the Flap+ protease (Figure 7). Overall, the hydration structure and location of identified hydration sites were conserved relative to wild-type protease, confirming the reproducibility of the protocol. Most hydration sites, including the symmetric hydration sites around the inhibitor binding site including HS-1B did not show any significant difference in occupancy (0.97, 0.98 in wildtype and Flap+ respectively). HS-2A, which is located proximal to the I54V mutation, had a minor increase in occupancy from 0.72 in wildtype to 0.8 in the Flap+ mutant.

More detailed analysis revealed changes in hydration, especially in occupancies and asymmetric hydration sites around the inhibitor binding site (Figure 7C). The hydration site corresponding to the flap water decreased from 90% to 82% occupancy. The flap water is key to inhibitor binding and coordinates intramolecular hydrogen bonds between inhibitor and residue I50 in the flaps. The occupancy of HS-1A which coordinates the backbone of G27 with the and the sidechain of D29 and R87 decreased from almost 100% in wildtype to 89% in the Flap+ Mutant. In wildtype protease, the sidechain of residue R8' packs tightly against the P1 phenyl and shields HS-1A (Figure 5E, Figure 6C). The mean distance between the central carbon of the R8 guanidinium and the P1 phenyl was 5.6 ± 0.8 Å. In the Flap+ mutant, this distance increased to 7.5 ± 1.4 Å. The larger distance between R8' and DRV allowed for exchange of HS-1A hydration site with other water molecules resulting in an overall lower occupancy. Thus, even among the most occupied hydration sites alterations were observed in their occupancy and positioning.

Considerable changes were observed in key hydration sites that were identified to coordinate the inhibited conformation in the analysis of wild-type protease. The cluster formed by hydration sites 3–6 (Figure 6A,B), which mediates intramolecular interactions between the

D29, R8 salt bridge and G48 of the protease flap, were destabilized in the Flap+ variant (Figure 7B,C). In the mutant variant, G48V introduced a nonpolar sidechain in the vicinity of this water cluster, which clashes with HS-4A, HS-3A and HS-5B. Thus, these hydration sites were completely lost in Flap+ protease and occupancy of remaining sites within the cluster were decrease. Preventing, in this drug-resistant variant, the formation of the hydration cluster that stabilizes the inhibitor through both direct hydrogen and an ordered water shell, the inhibitor bound to the wildtype enzyme.

Discussion

The hydration structure around HIV-1 protease and the asymmetry induced by binding of the inhibitor DRV have been characterized by identifying and characterizing hydration sites based on MD simulations. MD simulations give both spatial distribution and dynamics of water molecules, enabling a detailed and comprehensive analysis. A total simulation time of 1 μ s ensured convergence, and identification of both high and low occupancy hydration sites with confidence. We found that the surface topology was a major determinant in the location of high-density peaks, which defined the hydration sites. This result is in agreement with previous findings and recent reports investigating the molecular mechanism behind change in water dynamics close to protein surfaces.^{38–40} Overall, the protease–inhibitor surface dampened the water dynamics, and induced anisotropic fluctuations unlike for bulk water molecules.

Binding of the asymmetric inhibitor to the otherwise symmetric homodimeric protease resulted in asymmetric hydration sites, especially around the active site. There were more of these relatively lower occupancy asymmetric sites around the aniline moiety of the inhibitor. Compared to the bis-THF side, the aniline side of the inhibitor displays dynamics more weakly coupled to the protease dynamics of DRV and the concave shape of the solvent impose inhibitor imposes less spatial restrains on ordering of solvent molecules, leading to a more dynamic and extensive hydration network.²⁷ Aside from differences at the protease–DRV interface, the symmetry of the inhibitor-bound HIV-1 protease system was largely conserved in the hydration shell as well. The ability to recapture the overall symmetry of the system confirms the convergence of the methodology in identification and analysis of the hydration sites.

Since large changes in hydration site symmetry were largely confined to sites proximal to the active site, our subsequent analysis focused on the role of hydration sites at the active site. A pair of symmetric, buried hydration site has been found to stabilize the electrostatic interactions between G27, D29 loop and R87. The conserved hydrophilic residues, D29, D30, R87 and R8' are known to play an important role in active dimer formation and substrate recognition. The D29–R87 and subsequent D29–R8' salt bridge formation is a necessary first step in the formation of the active dimer.³¹ This salt bridge is conserved in a number of retroviral proteases, and the permanent occupancy of the hydration site stabilizing these interactions underpins its structural significance (Table S2). We further identify a central hydration site cluster spanning the protease interface coordinates interactions between D29, R8 and the flap residue G48. The tetracyclic arrangement of this water cluster at the bis-THF side of DRV is reminiscent of the low-lying minima of small water clusters

identified in ab-initio calculations.⁴¹ These calculations suggest these clusters interact via non-additive many-body terms not considered explicitly in MD force fields, but would further stabilize the hydration site cluster. While this central water cluster stabilizes the closed conformation of HIV-1 protease when bound to the inhibitor, binding to the substrate polypeptide would displace the water cluster. Thus, alterations in the hydration structure around the active site can have a large impact on binding to inhibitor while not affecting substrate binding.

The inhibitor–protease interface consists of a cluster of interconnected hydrophilic residues, circumscribed by small and hydrophobic residues (Figure 5C,E). The local topology promoted the formation of highly ordered hydration sites at this interface providing both strong electrostatic interaction partners and spatial restraints due to hydrophobic sidechains. In contrast to the highly conserved hydrophilic residues R8 and D29, the hydrophobic residues surrounding the protease–inhibitor interface is not conserved and act as hotspots for drug resistance mutations.

To compare the impact on drug resistance, we compared the results of our simulations on the DRV complex with a multi-drug resistant variant known as Flap+. ³⁴ This variant had a number of mutations in the hydrophobic residues surrounding the active site, I10L, G48V and V82A. Binding of DRV to Flap+ variant was accompanied by a remarkable inversion of enthalpic and entropic terms of the Gibbs free energy of binding. ³⁴ We found that well-defined hydration sites distal from the active site were largely conserved as in wild-type protease, but identified key changes at the protein-inhibitor interface (Figure 7A,B). In the Flap+ variant, the resistance mutations caused both a loss of hydration sites as well as an overall loss of occupancy at hydration sites that stabilize the inhibited form of protease. G48V mutation in Flap+ variant had deleterious effects on the hydration cluster connecting the R8/D29 sidechains with the protease flaps. Steric hindrance caused both a loss of hydration sites and an overall reduction of hydration site occupancy at the protein inhibitor interface (Figure 7B,C). In the Flap+ variant, the isopropyl sidechain of V48 acts as a molecular stirrer preventing the formation of well-ordered networks of water that would stabilize the inhibitor bound closed conformation of HIV-1 protease. These alterations in hydration structure may contribute to the loss in inhibitor binding affinity and the experimentally observed changes in binding thermodynamics.

The hydration site analysis we present here enabled the identification of key water–protein interactions that stabilize the closed conformation of HIV-1 protease in its inhibited form. While the hydration sites described largely corresponded to water molecules present in crystal structures, quantitatively evaluating the hydration structure and stability is challenging, if not impossible, from crystallographic data alone. Our protocol enabled identifying key changes in hydration structure of a drug resistant variant of HIV-1 protease, which correlate with experimental binding thermodynamics. In the resistant variant well-ordered hydration sites, which stabilize the inhibited conformation of HIV-1 protease, were disrupted. This does not only offer an explanation to the observed loss in potency but also explains the large change in enthalpic and entropic contributions to DRV binding. Previously we reported how subtle changes to the DRV P1' moiety can alter the occupancy of hydration sites surrounding the active site of HIV-1 protease,²⁷ indicating that, in principle,

changes to the inhibitor may restore stable hydration clusters. Consequently inhibitor modifications that incorporate hydrophilic moieties that stabilize the interfacial water cluster (Figure 6A,B) may improve activity against drug resistant protease variants. Recent reports have demonstrated the potential utility and successful application of targeting water molecules in structure-based drug design.^{42, 43} To develop robust small molecule inhibitors targeting rapidly evolving viral proteins, it is paramount to not only consider the immediate energetic gains from replacing a water molecules, but also the conservation of that water molecule in the evolutionary context. Consequently targeting water molecules surrounding conserved structural motifs, such as the R8'-D29-R87 dimerization interface (Figure 6C,D), will be more viable in the long run since those water molecules are less likely to be altered by resistance mutations. Future structure-based drug design efforts would benefit from incorporating knowledge of the hydration structure of both wildtype and resistant variants into the design of inhibitors targeting HIV-1 protease. Furthermore the methodology presented can be used to analyze the water structure of other systems, and thereby help predict changes in water structure due to either binding of ligands or mutations in the protein, with implications in a wide variety of biological processes.

Methods

System Preparation

The crystal structures of DRV bound to wildtype HIV-1 protease, bearing the sequence of isolate ARV2/SF2, and DRV bound to Flap+ variant whose structures we previously determined were taken from the Protein Database (PDB IDs: 1t3r and 3ekt).^{34, 44} All co-crystallized phosphates were removed and water molecules were kept prior to using Protein Preparation Wizard from the Schrodinger Suite.⁴⁵ Missing atoms were added using Prime.⁴⁶ The protonation state of the protein side chains at pH 7.0 was determined using PROPKA.^{47, 48} Particular care was taken to ensure the correct protonation state assignment for the catalytic aspartic acids, which is critical for binding. At physiological pH the catalytic dyad exists in a mono-protonated:unprotonated form.^{26, 49} Based on the pKa predicted by PROPKA, the aspartic acid proximal to the inhibitor's P1' and P2 moieties was chosen to be mono-protonated (pKa: 8.66). The hydrogen bond network was optimized by exhaustive sampling of the orientation of crystallographic waters and side chains of polar amino acids¹². Finally, the structure was subjected to gradient minimization with convergence criterion 0.5 Å using Impref.⁵⁰

Molecular Dynamics Simulations

The protein-inhibitor complex was placed in a cubic solvent box maintaining at least 1.5 nm spacing between any solute atom and the box boundaries. The system's net charge was neutralized by adding chloride ions; additionally, sodium and chloride ions were added to a total salt concentration of 0.15 M.

MD simulations were carried out using the Desmond software suite.^{51, 52} Protein and ligand were parameterized using the OPLS3 force field.⁵³ For the water molecules, TIP3P force field parameters were used.⁵⁴ During simulations, long-range electrostatic forces were calculated using the Particle Mesh Ewald method.⁵⁵ Short-range non-bonded forces were

truncated smoothly at 1.2 nm. The RESPA integrator was used with a 2 fs time step for bonded and short-range non-bonded forces and a 6 fs time step for long range electrostatic forces.⁵⁶

Before running MD simulations, the solvated system was energy minimized using a stepwise protocol. In the first iteration, all solvent molecules were minimized using 10 steepest decent steps followed by up to 5000 L-BSFG minimization steps. The convergence criterion was an energy gradient of $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ while applying solute heavy atoms a force constant of $1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. In the second iteration only the backbone restraints were kept and the system was subjected to the same minimization procedure as in the first iteration. In the third step, the restraints on the backbone heavy atoms were lowered to $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. For the final minimization step, all restraints were removed and the system was minimized using the L-BFGS method until an energy gradient of $0.05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ was reached.

Following minimization, a number of short MD simulations were performed to equilibrate the system. An initial 12 ps simulation was performed at constant temperature (10 K) and volume (NVT) using the Berendsen thermostat. The backbone position was restrained using a force constant of $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. This was followed by a 24 ps simulation at constant pressure (1 bar) and a constant temperature of 10 K (NPT), maintaining the restraints on the protein backbone. Subsequently a 50 ps unrestrained NPT simulation was run during which the temperature was increased from 10K to the target temperature of 300K. This was followed by a 500 ps NPT simulation at 300 K allowing thermal equilibration. The final production stage consisted of a 100 ns simulation in the NPT ensemble at 300 K and 1 bar. Atomic coordinates were recorded every 5 ps. A total of 10 MD simulations were run for wild-type HIV-1 protease using the protocol described above, where starting velocities were randomized for each production run. Simulations of DRV-bound Flap+ variant were performed following the same protocol, generating a total of 3 MD simulations of 100 ns each.

Identification of Hydration Sites

The solvent density distribution was calculated on a rectangular grid with a uniform spacing of $0.5 \text{ \AA}$ between each grid point. Snapshots from ten 100 ns simulations with randomized starting velocities were taken at 5 ps intervals, and the rotational and translational motions were minimized by superimposing each snapshot on the crystal structure by minimizing the distance between the C α atoms. The coordinates of all water oxygen atoms were subsequently mapped onto the grid by assigning each atom to the nearest grid point. The ten individual grids were subsequently pooled and the averaged number density was calculated by dividing each grid point by the total number of snapshots.

The hydration sites were defined as peaks in the water density distribution with a density of at least twice the density of bulk water (1 kg/L equivalent $0.033 \text{ \AA}^{-3}$). All grid points that passed the density threshold were compared with the 6 adjacent points to determine whether they qualify as local maxima. For all peaks that were within $1.5 \text{ \AA}$ of each other, the density-weighted centroid of these maxima was used.

Hydration Site Properties

Calculation of hydration site properties was carried out using the ten aligned 100 ns trajectories. The occupancy of a hydration site was defined as the fraction of snapshots in which a water oxygen was located within 1.5 Å of a given hydration site. Hydrogen bonds made by the occupying water molecule were determined using a geometric criterion. A hydrogen atom bonded to an eligible donor heavy atom was determined to form a hydrogen bond with a qualifying acceptor if the distance between the hydrogen atom and the acceptor atom was no more than 2.5 Å and the angle between donor-hydrogen-acceptor atoms was no less than 120°. This criterion gave average of 3.6 hydrogen bonds formed by TIP3P in bulk-solution, which is comparable to the 3.5 hydrogen bonds reported earlier using an energy criterion.²³

The root mean squared fluctuations (RMSF) of water molecules occupying a hydration site was calculated using $\sqrt{\frac{\sum_{i=1}^N (r_i - \langle r \rangle)^2}{N}}$ where N is the number of occupied frames for a given hydration site, r is the coordinate vector of the occupying water molecule oxygen, and the angular brackets denote mean. The anisotropy of the fluctuations was defined as $\frac{2\langle \Delta r_i^2 \rangle}{\langle \Delta r_j^2 \rangle + \langle \Delta r_k^2 \rangle}$ where $\langle \Delta r_i^2 \rangle$, $\langle \Delta r_j^2 \rangle$, $\langle \Delta r_k^2 \rangle$ are the mean displacement vectors along the principal axes, with $\langle \Delta r_i^2 \rangle$ being the vector of maximum displacement.

Hydration Site Symmetry

Evaluation of the hydration site symmetry was done by calculating the average reciprocal distance between each hydration site and the Cα atoms of the two protein monomers. This gave the distance vector AB defining the position of a hydration site relative to the protein Cα atoms, going from monomer A to monomer B. Since the two monomers are equivalent in length and amino acid composition, the distance vector BA going from monomer B to monomer A describes the mirror image of that site across the symmetry axis. The similarity of two sites x and y was thus determined by calculating the Pearson correlation between the

two vectors ABx and BAy: $\frac{\sum_{i=1}^n (AB_i^x - \langle AB^x \rangle)(BA_i^y - \langle BA^y \rangle)}{\sqrt{\sum_{i=1}^n (AB_i^x - \langle AB^x \rangle)^2} \sqrt{\sum_{i=1}^n (BA_i^y - \langle BA^y \rangle)^2}}$. The two sites were

considered to be symmetric if the correlation was above the threshold value of 0.95 and the highest among all pairwise combinations for both hydration sites.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIH P01 GM109767.

References

1. Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct.* 2006; 35:389–415. [PubMed: 16689642]
2. Maruyama Y, Harano Y. Does water drive protein folding? *Chem Phys Lett.* 2013; 581:85–90.
3. Auffinger P, Hashem Y. Nucleic acid solvation: from outside to insight. *Curr Opin Struct Biol.* 2007; 17:325–333. [PubMed: 17574833]
4. Otting G, Liepinsh E, Wuthrich K. Protein hydration in aqueous solution. *Science.* 1991; 254:974–980. [PubMed: 1948083]
5. Biedermannova L, Schneider B. Hydration of proteins and nucleic acids: Advances in experiment and theory. A review. *Biochimica et Biophysica Acta (BBA) - General Subjects.* 2016; 1860:1821–1835. [PubMed: 27241846]
6. Sterpone F, Stirnemann G, Laage D. Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *J Am Chem Soc.* 2012; 134:4116–4119. [PubMed: 22335572]
7. Mattea C, Qvist J, Halle B. Dynamics at the protein-water interface from ^{17}O spin relaxation in deeply supercooled solutions. *Biophys J.* 2008; 95:2951–2963. [PubMed: 18586840]
8. Heyden M, Tobias DJ. Spatial dependence of protein-water collective hydrogen-bond dynamics. *Phys Rev Lett.* 2013; 111:218101. [PubMed: 24313531]
9. Fogarty AC, Laage D. Water dynamics in protein hydration shells: The molecular origins of the dynamical perturbation. *The Journal of Physical Chemistry B.* 2014; 118:7715–7729. [PubMed: 24479585]
10. Zia SR, Gaspari R, Decherchi S, Rocchia W. Probing hydration patterns in class-A GPCRs via biased MD: the A2A receptor. *J Chem Theory Comput.* 2016; 12:6049–6061. [PubMed: 27951680]
11. Imai T, Kovalenko A, Hirata F. Solvation thermodynamics of protein studied by the 3D-RISM theory. *Chem Phys Lett.* 2004; 395:1–6.
12. Abel R, Young T, Farid R, Berne BJ, Friesner RA. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc.* 2008; 130:2817–2831. [PubMed: 18266362]
13. Sun H, Zhao L, Peng S, Huang N. Incorporating replacement free energy of binding-site waters in molecular docking. *Proteins: Struct Funct Bioinform.* 2014; 82:1765–1776.
14. Denisov VP, Halle B. Protein hydration dynamics in aqueous solution. *Faraday Discuss.* 1996; 103:227–244.
15. Makarov VA, Andrews BK, Smith PE, Pettitt BM. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys J.* 2000; 79:2966–2974. [PubMed: 11106604]
16. Scott WR, Schiffer CA. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure.* 2000; 8:1259–1265. [PubMed: 11188690]
17. Huang X, de Vera IMS, Veloro AM, Blackburn ME, Kear JL, Carter JD, Rocca JR, Simmerling C, Dunn BM, Fanucci GE. Inhibitor-Induced Conformational Shifts and Ligand-Exchange Dynamics for HIV-1 Protease Measured by Pulsed EPR and NMR Spectroscopy. *The Journal of Physical Chemistry B.* 2012; 116:14235–14244. [PubMed: 23167829]
18. Zhang Y, Chang YCE, Louis JM, Wang YF, Harrison RW, Weber IT. Structures of darunavir-resistant HIV-1 protease mutant reveal atypical binding of darunavir to wide open flaps. *ACS Chem Biol.* 2014; 9:1351–1358. [PubMed: 24738918]
19. Wang YX, Freedberg DI, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Bhat TN, Erickson JW, Torchia DA. Bound Water Molecules at the Interface between the HIV-1 Protease and a Potent Inhibitor, KNI-272, Determined by NMR. *J Am Chem Soc.* 1996; 118:12287–12290.
20. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure.* 2002; 10:369–381. [PubMed: 12005435]
21. Liu Z, Wang Y, Yedidi RS, Dewdney TG, Reiter SJ, Brunzelle JS, Kovari IA, Kovari LC. Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes. *Biochem Biophys Res Commun.* 2013; 430:1022–1027. [PubMed: 23261453]

22. Fogarty AC, Duboué-Dijon E, Sterpone F, Hynes JT, Laage D. Biomolecular hydration dynamics: a jump model perspective. *Chem Soc Rev.* 2013; 42:5672–5672. [PubMed: 23612685]
23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983; 79:926–926.
24. Biedermannova L, Schneider B. Structure of the ordered hydration of amino acids in proteins: analysis of crystal structures. *Acta Cryst D Biol Crystallogr.* 2015; 71:2192–2202. [PubMed: 26527137]
25. Cerny J, Schneider B, Biedermannova L. WatAA: Atlas of Protein Hydration. Exploring synergies between data mining and ab initio calculations. *PCCP.* 2017
26. McGee TD Jr, Edwards J, Roitberg AE. pH-REMD simulations indicate that the catalytic aspartates of HIV-1 protease exist primarily in a monoprotated state. *The Journal of Physical Chemistry B.* 2014; 118:12577–12585. [PubMed: 25340507]
27. Paulsen JL, Leidner F, Ragland DA, Kurt Yilmaz N, Schiffer CA. Interdependence of Inhibitor Recognition in HIV-1 Protease. *J Chem Theory Comput.* 2017; 13:2300–2309. [PubMed: 28358514]
28. Li Z, Lazaridis T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *J Am Chem Soc.* 2003; 125:6636–6637. [PubMed: 12769565]
29. Ishima R, Ghirlando R, Tözsér J, Gronenborn AM, Torchia DA, Louis JM. Folded monomer of HIV-1 protease. *J Biol Chem.* 2001; 276:49110–49116. [PubMed: 11598128]
30. Ishima R, Gong Q, Tie Y, Weber IT, Louis JM. Highly conserved glycine 86 and arginine 87 residues contribute differently to the structure and activity of the mature HIV-1 protease. *Proteins: Struct Funct Bioinform.* 2009; 78:1015–1025.
31. Hayashi H, Takamune N, Nirasawa T, Aoki M, Morishita Y, Das D, Koh Y, Ghosh AK, Misumi S, Mitsuya H. Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir. *Proc Natl Acad Sci USA.* 2014; 111:12234–12239. [PubMed: 25092296]
32. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW, Richman DD. 2015 update of the drug resistance mutations in HIV-1. *Top Antivir Med.* 2015; 23:132–141. [PubMed: 26713503]
33. Ragland DA, Nalivaika EA, Nalam MNL, Prachanronarong KL, Cao H, Bandaranayake RM, Cai Y, Kurt-Yilmaz N, Schiffer CA. Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J Am Chem Soc.* 2014; 136:11956–11963. [PubMed: 25091085]
34. King NM, Prabu-Jeyabalan M, Bandaranayake RM, Nalam MN, Nalivaika EA, Özen Ael, Halilo lu Tr, Yilmaz NeK, Schiffer CA. Extreme Entropy–Enthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease. *ACS Chem Biol.* 2012; 7:1536–1546. [PubMed: 22712830]
35. Cai Y, Myint W, Paulsen JL, Schiffer CA, Ishima R, Kurt Yilmaz N. Drug resistance mutations alter dynamics of inhibitor-bound HIV-1 protease. *J Chem Theory Comput.* 2014; 10:3438–3448. [PubMed: 25136270]
36. Foulkes-Murzycki JE, Rosi C, Kurt Yilmaz N, Shafer RW, Schiffer CA. Cooperative effects of drug-resistance mutations in the flap region of HIV-1 protease. *ACS Chem Biol.* 2012; 8:513–518. [PubMed: 23252515]
37. Cai Y, Kurt Yilmaz N, Myint W, Ishima R, Schiffer CA. Differential flap dynamics in wild-type and a drug resistant variant of HIV-1 protease revealed by molecular dynamics and NMR relaxation. *J Chem Theory Comput.* 2012; 8:3452–3462. [PubMed: 23144597]
38. Brovchenko I, Oleinikova A. Which Properties of a Spanning Network of Hydration Water Enable Biological Functions? *Chemphyschem.* 2008; 9:2695–2702. [PubMed: 19035367]
39. Laage D. A Molecular Jump Mechanism of Water Reorientation. *Science.* 2006; 311:832–835. [PubMed: 16439623]
40. Laage D, Elsaesser T, Hynes JT. Water Dynamics in the Hydration Shells of Biomolecules. *Chem Rev.* 2017
41. Xantheas SS. Cooperativity and hydrogen bonding network in water clusters. *Chem Phys.* 2000; 258:225–231.

42. Bucher D, Stouten P, Triballeau N. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J Chem Inf Model*. 2018
43. Lee KL, Ambler CM, Anderson DR, Boscoe BP, Bree AG, Brodfuehrer JI, Chang JS, Choi C, Chung S, Curran KJ. Discovery of Clinical Candidate 1-[(2 S, 3 S, 4 S)-3-Ethyl-4-fluoro-5-oxopyrrolidin-2-yl] methoxy-7-methoxyisoquinoline-6-carboxamide (PF-06650833), a Potent, Selective Inhibitor of Interleukin-1 Receptor Associated Kinase 4 (IRAK4), by Fragment-Based Drug Design. *J Med Chem*. 2017; 60:5521–5542. [PubMed: 28498658]
44. Surleraux DLNG, Tahri A, Verschueren WG, Pille GME, de Kock HA, Jonckers THM, Peeters A, De Meyer S, Azijn H, Pauwels R, de Bethune MP, King NM, Prabu-Jeyabalan M, Schiffer CA, Wigerinck PBTP. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J Med Chem*. 2005; 48:1813–1822. [PubMed: 15771427]
45. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput-Aided Mol Des*. 2013; 27:221–234. [PubMed: 23579614]
46. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol*. 2002; 320:597–608. [PubMed: 12096912]
47. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *J Chem Theory Comput*. 2011; 7:525–537. [PubMed: 26596171]
48. Søndergaard CR, Olsson MH, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK (a) values. *J Chem Theory Comput*. 2011; 7:2284–2295. [PubMed: 26606496]
49. Piana S, Sebastiani D, Carloni P, Parrinello M. Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J Am Chem Soc*. 2001; 123:8730–8737. [PubMed: 11535077]
50. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR. Integrated modeling program, applied chemical theory (IMPACT). *J Comput Chem*. 2005; 26:1752–1780. [PubMed: 16211539]
51. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters; SC 2006 Conference, Proceedings of the ACM/IEEE; 2006/november; 43–43.
52. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G. Probing the alpha-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. *Chem Biol Drug Des*. 2010; 75:348–359. [PubMed: 20331649]
53. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput*. 2015; 12:281–296. [PubMed: 26584231]
54. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983; 79:926–935.
55. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995; 103:8577–8577.
56. Tuckerman MBBJM, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. *J Chem Phys*. 1992; 97:1990–2001.

Figure 1. HIV-1 protease structure and hydration sites

(A) Crystal structure of HIV-1 protease in ribbon representation, bound to inhibitor darunavir (DRV, teal sticks) at the active site. The two monomers are depicted in blue and orange. (B) Distribution of grid points with density greater than twice ρ_{bulk} around HIV-1 protease (blue and orange surface). Grid points are shown as green fishnet. (C) Hydration sites identified using the water density distribution, and displayed as purple spheres. (D,E) Cross-section of the water density through the protease active site. Grid cells colored according to density values. Warmer colors represent higher density, cooler colors represent lower density. Voxels only occupied by HIV-1 Protease and Darunavir are colored in grey.

Figure 2. Characterization of hydration sites around HIV-1 protease

Histograms of (A) occupancy, (B) mean number of hydrogen-bonds, (C) fraction of hydrogen bonds made with protein residues, (D) number of water oxygens within 4 Angstrom, (E) root-mean-square fluctuation (RMSF) of the occupying water oxygen, and (F) anisotropy of the RMSF, for the identified hydration sites around HIV-1 protease. The mean values are indicated as black dashed lines, the mean values for bulk water are depicted as red dashed lines when applicable.

Figure 3. Correlation between hydration site characteristics

The matrix displays Pearson correlation coefficient between features of hydration sites displayed as histograms in Figure 2 (occupancy, RMSF of waters, anisotropy of RMSF, number of neighboring hydration sites and number of protease hydrogen bonds), and colored red to blue for positive to negative correlations, with gray depicting no correlation between the two features. The scatter plots for all 15 correlations are in Supplementary Figure S2.

Figure 4. Symmetry of above mean occupancy hydration sites

Symmetric hydration sites common in both protease monomers are shown as purple and green spheres, while the asymmetric sites with no matching counterpart associated with the other protease monomer are in red. The “flap water” located at the symmetry axis is depicted as a yellow sphere. Only the hydration sites with mean occupancy greater than mean occupancy (0.6) are depicted.

Figure 5. Ligand induced Asymmetry A) Chemical structure of Darunavir P1–P2 and P1'–P2' subsites are labeled. B) DRV bound in the active site of HIV protease, asymmetry indicated in yellow outline. Flaps residues 45–55 hidden. C) Close-up views of hydration sites around the protease active site facing the aniline moiety of DRV. Active site residues are color coded as yellow: apolar, blue: polar, red: charged. D) Mean occupancies. E) Close-up views of hydration sites around the protease active site facing the bis-THF moiety of DRV. Active site residues are color coded as yellow: apolar, blue: polar, red: charged. F) Mean occupancies.

Figure 6. High occupancy hydration sites around the Protease-Darunavir interface

A) View of the central water cluster around the bis-THF interface of DRV. B) View of the central water cluster and asymmetric sites facing the aniline site of the Inhibitor. C) View of the buried water molecule HS-1A coordinating interactions between the G27-D29 Loop and R87, facing the bis-THF moiety of DRV. D) View of the buried water site HS-1B and HS-101 at the aniline face of DRV. E) View of the interstitial hydration site HS-2A bridging interactions between the “flaps” residue G51 and P79’ of the 80s loop. F) View of the interstitial water molecule HS-2B and coordinated hydration sites HS-201 and HS-202, mediating interactions between the I50’ if the closed flaps and residue P79.

Figure 7. Comparison of High occupancy hydration sites in the WT-DRV and WT-Flap+ Complex

A) Hydration sites with a mean occupancy of over 65% show on the surface of the Flap+ Darunavir complex. Mutations are highlighted in tan. Wildtype hydration sites are shown as transparent cyan spheres, hydration sites identified in the drug resistant variant are shown as solid red spheres. B) High occupancy hydration sites in the active site of HIV-1 Protease. Wildtype hydration sites shown as transparent cyan spheres, Flap+ sites shown as solid red spheres. Hydration sites correspond to sites introduced in Figure 5,6. C) Table of high occupancy waters in the active site of Darunavir bound to Wildtype and Flaps+ HIV-1 Protease variants. Nomenclature adopted from Figure 6. Errors were calculated based on 10 replicates for WT complex and 3 replicates for Flap+ Complex. Hydration sites that could not be resolved marked as NR.