HIV-1 Protease and Substrate Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern

Ayşegül Özen[†], **Türkan Haliloğlu**[*,§], and **Celia A. Schiffer**[*,†]

[†]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605, United States

[§]Polymer Research Center, Bogazici University, Bebek, Istanbul, Turkey

Abstract

Drug resistance of HIV-1 protease alters the balance in the molecular recognition events in favor of substrate processing versus inhibitor binding. To develop robust inhibitors targeting ensembles of drug-resistant variants, the code of this balance needs to be cracked. For this purpose, the principles governing the substrate recognition are required to be revealed. Previous crystallographic studies on the WT protease–substrate complexes showed that the substrates have a conserved consensus volume in the protease active site despite their low sequence homology. This consensus volume is termed as the substrate envelope. The substrate envelope was recently reevaluated by taking the substrate dynamics into account, and the dynamic substrate envelope was reported to better define the substrate specificity for HIV-1 protease. Drug resistance occurs mostly through mutations in the protease, occasionally accompanied by cleavage site mutations. In this study, three coevolved protease–substrate complexes (^{AP2V}NC-p1$_{V82A}$, $^{LP1'F}$p1-p6$_{D30N/N88D}$, and $^{SP3'N}$p1-p6$_{D30N/N88D}$) were investigated for structural and dynamic properties by molecular modeling and dynamics simulations. The results show the substrate envelope is preserved by these cleavage site mutations in the presence of drug-resistance mutations in the protease, if not enhanced. This study on the conformational and mutational ensembles of protease–substrate complexes validates the substrate envelope as the substrate recognition motif for HIV-1 protease. The substrate envelope hypothesis allows for the elucidation of possible drug resistance mutation patterns in the polyprotein cleavage sites.

1. Introduction

Human immunodeficiency virus type-1 (HIV-1) protease (PR) is a key enzyme in the viral life cycle that processes the Gag and Gag-Pro-Pol viral polyproteins at 12 cleavage sites, yielding mature, infectious virions.[1] HIV-1 PR, a 99 residue, homodimeric aspartyl PR,[2,3] is essential for viral maturation[1] and hence is a main drug target. The United States Food and Drug Administration (FDA) has approved nine PR inhibitors (PIs) for clinical use. These PIs, used as part of highly active antiretroviral therapy, have significantly improved the management of disease condition by lengthening the life span and increasing the quality of life of HIV-infected patients.[4] However, the high rate of viral replication[5] combined with the lack of proofreading mechanism in viral reverse transcriptase[6] generate massive amounts of genetically distinct viral variants. Within these viral quasispecies, the selective pressure

[*]Corresponding Author: celia.schiffer@umassmed.edu; turkan@prc.boun.edu.tr.

Notes: The authors declare no competing financial interest.

of drug therapy populates the viral variants that are not completely inhibited by antiviral drugs targeting viral enzymes.

At a molecular level, drug resistance reflects a subtle change in the balance of enzyme recognition events, in favor of natural substrate processing versus inhibitor binding. Drug-resistant PR variants have mutations that significantly alter inhibitor binding but do not drastically change substrate processing. In addition, emergence of resistance to PR inhibitors does not always depend solely on PR mutational plasticity. The natural substrates can also mutate in association with drug therapy.[7–10] Two examples of this PR–substrate coevolution are NC-p1 and p1-p6, two cleavage sites on the Gag polyprotein that coevolve with the viral PR to confer resistance to PR inhibitors. In the NC-p1 cleavage site, Ala in the P2 position mutates to a Val in response to the V82A multidrug-resistance PR mutation (^{AP2V}NC-p1$_{V82A}$).[7,8,11,12] The p1-p6 cleavage site mutates predominantly at the P1′ or P3′ positions[13] associated with the PR double mutation D30N/N88D ($^{LP1'F}$p1-p6$_{D30N/N88D}$ and $^{SP3'N}$p1-p6$_{D30N/N88D}$), which is a signature of nelfinavir treatment.[14,15] Replicative capacity assays showed that the D30N/N88D viruses with the compensatory mutations in p1-p6 do not have improved fitness relative to viruses with mutations in the PR alone.[16] V82A virus has an even lower replicative capacity in combination with mutations at Gag 431, which corresponds to Ala-P2 of NC-p1 cleavage site, compared to those without this mutation.[16] However, these co-occurring PR—substrate mutations were shown to significantly affect the PR inhibitor susceptibilities.[16]

The structural basis for PR–substrate coevolution in ^{AP2V}NC-p1$_{V82A}$ variant came from analyzing the crystal structures of inactive D25N WT (WT) and V82A HIV-1 PR in complex with their respective WT and AP2V mutant NC-p1 substrates.[17] The crystal structures revealed that ^{WT}NC-p1 binds to HIV-1 PR less optimally than ^{AP2V}NC-p1 with fewer hydrogen bonds and fewer van der Waals (vdW) contacts. In addition, Ala-P2 was observed to incompletely fill the P2 pocket. For PR—substrate coevolution in $^{LP1'F}$p1-p6$_{D30N/N88D}$ or ^{SP3N}p1-p6$_{D30N/N88D}$ variants, however, no experimentally determined structures have been reported so far.

Completely understanding the molecular basis of substrate recognition is crucial to develop robust inhibitors that better compete with natural substrates of highly resistant PR variants. Co-evolution of PR and the natural substrates allows for the study of the interdependency between HIV-1 PR and its natural substrates, which facilitates the substrate recognition. The principles underlying substrate recognition by HIV-1 PR are not sequence specific because the amino acid sequences of the cleavage sites on the Gag and Gag-Pro-Pol polyproteins do not have an obvious sequence homology. These nonhomologous substrates, however, occupy a conserved consensus volume in the binding site of the PR in crystal structures.[18,19] This conserved three-dimensional shape, describing the "substrate envelope", was recently redefined by incorporating the substrate dynamics within the PR binding site.[20] In that previous study, the dynamic substrate envelope was shown to better define the substrate specificity for HIV-1 PR, compared to the static substrate envelope.

In the present study, the dynamics of three separate examples of HIV-1 PR—substrate coevolution were investigated, to elucidate the interdependence of substrate recognition with mutations in the PR. The molecular interactions of each set were analyzed, including with respect to the dynamic substrate envelope.[20] The compensatory mutations in the cleavage sites were shown to fit within the substrate envelope better than WT substrates. This was achieved by a variety of changes in interactions but not in one conserved manner for all three substrates. Thus, the substrate envelope is preserved by the cleavage site mutations in the presence of drug-resistance mutations in the PR, validating the substrate envelope as the substrate recognition pattern for HIV-1 PR, whether or not the substrate evolves. The

dynamic substrate envelope is potentially a powerful tool to predict the coevolution of the cleavage sites of HIV-1 PR.

2. Results and Discussion

The three separate cases of HIV-1 PR—substrate coevolution investigated in this study are listed in Table 1. The variants of the PR and the cleavage site are denoted by a subscript and a superscript, respectively. Details of the nomenclature used throughout the paper are described in the Methods Section.

2.1. Fit within the Dynamic Substrate Envelope

Substrate fit within the binding groove was compared over the course of molecular dynamics (MD) simulations with and without the effect of drug resistance by first calculating the substrate volume within and outside the dynamic substrate envelope and the overall vdW interactions of the substrate with the PR (Figure 1). Substrate volume in the binding groove can be analyzed in two components: the substrate volume protruding beyond the substrate envelope (V_{out}), and the substrate volume lying within the substrate envelope (V_{in}). A high degree of protrusion beyond the substrate envelope (quantified as high V_{out} values) was shown in our earlier work to indicate substrates susceptible to drug-resistance mutations in the PR, e.g., both p1-p6 and NC-p1 substrates have higher V_{out} values than expected based on their molecular volume.[20]

The current study shows that compensatory mutations in the cleavage sites optimize the portion of the substrate volume that stays within the substrate envelope. This observation is quantified in Figure 1A—C, where V_{in} is plotted for three cases of coevolution (^{AP2V}NC-p1$_{V82A}$, $^{LP1'F}$p1-p6$_{D30N/N88D}$, and $^{SP3'N}$p1-p6$_{D30N/N88D}$) with reference to their respective WT complex structure ($\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$). In each case, the substrate mutation compensates for the decrease in V_{in} as a result of primary PR mutation.

In coevolution of NC-p1 cleavage site, ΔV_{in} increases from —26.3 ± 1.9 Å^3 in ^{WT}NC-p1$_{V82A}$ complex to 4.2 ± 1.8 Å^3 in ^{AP2V}NC-p1$_{V82A}$ (Figure 1A), while in coevolution of p1-p6 cleavage site, ΔV_{in} increases from —122.2 ± 1.5 Å^3 in ^{WT}p1-p6$_{D30N/N88D}$ to 8.2 ± 1.3Å^3 in $^{LP1'F}$p1-p6$_{D30N/N88D}$ (Figure 1B) and to —48.5 ± 1.8 Å^3 in $^{SP3'N}$p1-p6$_{D30N/N88D}$ (Figure 1C).

^{AP2V}NC-p1 is slightly larger than ^{WT}NC-p1 because of two additional methyl groups on the side chain. Despite 2.5% larger volume, ^{AP2V}NC-p1 fills the dynamic substrate envelope better than ^{WT}NC-p1, i.e., the improvement in V_{in} (30.5 ± 2.6 Å^3, Figure1A) is more pronounced than the increase in V_{out} (6.9 ± 2.8 Å^3, Figure 1D) in ^{AP2V}NC-p1$_{V82A}$ compared to ^{WT}NC- p1$_{V82A}$. Similarly, in the presence of D30N/N88D mutations in the PR, LP1'F and SP3'N substitutions in p1-p6 improve V_{in} by 130.4 ± 2.0 and 73.7 ± 2.3 Å^3 (Figure 1B and C, respectively), making the substrate better fill the substrate envelope. On the other hand, the reduction in V_{out} is 96.2 ± 2.8 and 11.8 ± 3.3 Å^3 (Figure 1E and F, respectively) upon LP1'F and SP3'N occurs in the context of D30N/N88D PR variant. These compensatory mutations in p1-p6 optimize the substrate fit within the envelope by both increasing V_{in} and decreasing V_{out}.

Preservation of substrate envelope upon compensatory cleavage site mutations in drug-resistant PR variants is visualized in Figure 2 in the case of $^{LP1'F}$p1-p6$_{D30N/N88D}$. The altered fit of p1-p6 within the dynamic substrate envelope in ^{WT}p1-p6$_{D30N/N88D}$ and $^{LP1'F}$p1-p6$_{D30N/N88D}$ (Figure 2A and B, respectively) complexes is illustrated by mapping the ΔV_{in} grid matrix onto the cocrystal structure. The negative values of the difference matrix ΔV_{in} shown in cyan, correspond to the regions where the substrate in

mutant complex fits poorly within the dynamic substrate envelope compared to ^{WT}p1-p6$_{WT}$. The positive values of the difference matrix ΔV_{in}, shown in orange, correspond to the regions where the substrate in mutant complex fits better within the dynamic substrate envelope compared to ^{WT}p1-p6$_{WT}$. The improvement in ΔV_{in} upon the compensatory mutation in the p1-p6 cleavage site shown in Figure 1B is reflected by the complete loss of negative regions in Figure 2B.

2.2. Molecular Interactions between the Protease and Substrates: Analysis of vdW Contacts

Protease and/or cleavage-site mutations conferring drug resistance do not drastically alter the overall vdW contact potential between the PR and the substrates (Figure 1G—I). This result is consistent with our earlier observation[20] that a conserved consensus overall vdW interaction potential might be a prerequisite for a sequence to be recognized by the PR. This observation, combined with the ability of drug-resistant PR variants to still process cleavage sites, leads to the expectation that cleavage sites in the resistant viruses will naturally maintain this optimal overall interaction potential. In fact, this optimal vdW potential is preserved by all the drug-resistant PR—substrate complex variants investigated in this study, with overall ΔvdW on the order of $\approx$2.5 kcal/mol within the range of WT contact potential (Figure 1G—I, middle bars). For example, the PR— substrate vdW interaction potential of ^{WT}p1-p6$_{D30N/N88D}$ is 1.9 ± 0.2 kcal/mol (Figure 1H) less favorable than that of ^{WT}p1-p6$_{WT}$. The loss of this interaction due to D30N/N88D PR mutations is restored with LP1T and SP3 N substitutions in the cleavage site to a level of -1.7 ± 0.2 and -3.2 ± 0.2 kcal/mol (Figure 1H and I, respectively). In contrast, ^{WT}NC-p1$_{V82A}$ has -4.6 ± 0.2 kcal/mol more favorable vdW interactions compared to ^{WT}NC-p1$_{WT}$ (Figure 1G). NC-p1 is inherently a very flexible substrate due to the P3′-Gly and P4′-Lys residues. Substitution of Val-82 to an Ala results in a larger local volume for this flexible peptide to sample. These conformations may not be accessible to the peptide in ^{WT}NC-p1$_{WT}$ complex because of the additional methyl groups in Val-82. Providing extra volume in the variant allows the peptide to sample conformations that have more favorable interactions with the PR, as is evident by the change in vdW interaction potential.

AP2V mutation in NC-p1 substrate brings down vdW interactions by 1.1 ± 0.3 kcal/mol but still keeps it more favorable than the WT level. This increase in the overall vdW contact potential might contribute to ^{AP2V}NC-p1 being better recognized than ^{WT}NC-p1 by the WT PR.[10] However, the processing efficiency of each cleavage site should be optimal for viral fitness. A significantly lower efficiency leaves unprocessed viral polyproteins packaged in the virion, while a significantly higher efficiency might interfere with the accurate and precise regulation of Gag and Gag-Pro-Pol processing. NC-p1 is the rate-limiting cleavage site in HIV-1 Gag during processing by the viral PR.[21,22] Better than optimum cleavage efficiency for this site might introduce temporal imprecision in the regulation of its hydrolysis, leading to immature Gag processing. This may be the reason why (1) on average, ^{AP2V}NC-p1$_{V82A}$ variants have lower replicative capacity than ^{WT}NC-p1$_{V82A}$,[16] and (2) the more efficiently cleaved ^{AP2V}NC-p1 is not highly populated in the absence of drug resistance mutations in the PR.[17]

Although drug resistance does not severely impact the overall PR—substrate contact potential, the vdW contact potential profiles of substrate residues with any PR atom significantly deviate from the WT substrate behavior (Figure 3). The nature of this deviation, however, is not shared by all substrate variants. The loss of interactions on the primed side of the NC-p1 cleavage site is compensated by the gain of interactions on the unprimed side (Figure 3A). This compensation supports that the substrate amino acid residues are interdependent despite the very short substrate sequence. Consequently, this

interdependence among the amino acid residues helps maintain the overall vdW interaction potential at an optimal level within ≈ 3 kcal/mol of the WT range.[20]

In contrast, the compensatory mechanism is more symmetric in the p1-p6 cleavage site than in NC-p1 (Figure 3B and C). ^{WT}p1-p6$_{D30N/N88D}$ loses 4.3 ± 2.0 kcal/mol vdW interactions at the P4′ position and gains comparable interactions at the P2 position (-4.0 ± 0.9 kcal/mol, Figure 3B, blue bars). This severe deviation from the WT profile at these two positions is moderated with either LP1′F or SP3′N substitutions in p1-p6. $^{LP1'F}$p1-p6$_{D30N/N8D}$ has more favorable P4′ interactions than ^{WT}p1-p6$_{D30N/N8D}$, although the loss of interaction potential at this position is considerably higher than ^{WT}p1-p6$_{WT}$. Loss of interactions at P4′ position in $^{LP1'F}$p1-p6$_{D30N/N88D}$ with respect to ^{WT}p1-p6$_{WT}$ (2.3 ± 0.1 kcal/mol, Figure 3B, green bars) is compensated by the gain of interactions at two other substrate residues, P1 (-2.3 ± 0.1 kcal/mol) and P2′ (-1.1 ± 0.1 kcal/mol).

In addition, P2 in $^{LP1'F}$p1-p6$_{D30N/N88D}$ does not interact with the PR as favorably as P2 in ^{WT}p1-p6$_{D30N/N88D}$ (ΔvdW $= -4.0 \pm 0.1$ kcal/mol), even though P2 is invariant. The dramatic change in the vdW contacts of this invariant position is caused by the side-chain rearrangements within the substrate and the binding site upon LP1′F. The conformational flexibility of the substrate combined with the interdependency within the substrate sequence allows for the altered vdW interactions in an invariant position (Asp P2) as a result of a mutation in another position, LP1′F (Figure 4), within the cleavage site. Thus, a single mutation manages to alter the whole vdW interaction profile along the sequence.

In addition, SP3′N has a similar impact on the per residue vdW interaction profile of p1-p6. Compared to the ^{WT}p1-p6$_{WT}$ complex, P1 and P1′ in $^{SP3'N}$p1-p6$_{D30N/n88D}$ have slightly less favorable interactions with the PR, on the order of total ≈ 1 kcal/mol (Figure 3C). This decreased interaction, however, is compensated by the improved contact potentials of the surrounding residues, P3, P2, P2′ and P3′. Overall, the vdW contacts are preserved, however through a different interaction profile, likely facilitated by the interdependence within the substrate sequences.

2.3. Molecular Interactions between the Protease and Substrates: Analysis of Hydrogen Bonds

The hydrogen bonds formed between the PR and substrates were analyzed in two groups: hydrogen bonds formed by the substrate (1) backbone and (2) side chains. The backbone hydrogen bonds are conserved among various natural substrates, while no single side-chain hydrogen bond is shared by all or most of the substrates.[18,20] Therefore, hydrogen bonds between the side chains and any PR atoms are more likely than the backbone hydrogen bonds to contribute to substrate specificity.[20] The less specific backbone hydrogen bonds are not altered by drug-resistance mutations in the PR and/or substrate despite remarkable local conformational rearrangements (Figure 5A-D). The three complexes of ^{WT}p1p6$_{WT}$, ^{WT}p1-p6$_{D30N/N88D}$, and $^{LP1'F}$p1-p6$_{D30N/N88D}$ are superimposed to highlight these rearrangements (Figure 5A and E.). The percentage of time the hydrogen bonds existed throughout the simulations is listed in Table 2. The most stable NC-p1 backbone hydrogen bonds are formed between P2 and G48, P3′ and G48′, and P2′ and G27′, existing more than 90% of the time in most variants. These hydrogen bonds should contribute to stabilizing the substrate when bound to the PR active site. The hydrogen bonds formed by NC-p1 side-chains, on the other hand, are more variable (second main column, Table 2). The bonds formed in ^{WT}NC-p1$_{WT}$ between the side chain of P1 and I50, N25′ and G27 were not stable throughout the MD trajectories. However, both ^{AP2V}NC-p1$_{WT}$ and ^{AP2V}NC-p1$_{V82A}$ have more consistent hydrogen bonds formed by P1 and I50 and G27. These presumably stronger hydrogen bonds might be important for the increased substrate affinity in both ^{AP2V}NC-p1$_{WT}$ and ^{AP2V}NC-

p1$_{V82A}$ variants, consistent with ^{AP2V}NC-p1 being more efficiently cleaved than ^{WT}NC-p1 by the WT HIV-1 PR.[10]

^{WT}p1-p6$_{WT}$ forms substrate backbone hydrogen bonds at P1 and P3′ positions (Figure 5B). The majority of these bonds are preserved in drug-resistant ^{WT}p1-p6$_{D30N/N88D}$, although the bond between backbone nitrogen of P1 and backbone oxygen of G27 no longer exists in ^{WT}p1-p6$_{D30N/N88D}$ (Figure 5C). This loss of a hydrogen bond, however, is compensated by the hydrogen bonds newly formed by substrate P3 and P2′ positions. $^{LP1'F}$p1-p6$_{D30N/N88D}$ has almost the same pattern of hydrogen-bonding network as ^{WT}p1-p6$_{WT}$ (Figure 5D). In addition, the P1, P2′, and P3′ positions make backbone hydrogen bonds with the PR in $^{SP3'N}$p1-p6$_{D30N/N88D}$ (Table 2). In complex with the WT PR, these mutant substrates have more backbone hydrogen bonds with decreased percent time of existence. Drug resistance, conferred by mutations in the PR and/or the substrate, does not alter the backbone hydrogen-bonding pattern drastically. Even though the side chains vary by drug resistance, the substrates can still form these less specific hydrogen bonds with the PR, presumably stabilizing the substrates in the binding groove.

^{WT}p1-p6$_{D30N/N88D}$ has consistent PR–substrate side-chain hydrogen bonds formed by P2 and P2′ residues (Figure 5G). P2 side-chain hydrogen bond is not affected by drug resistance (Figure 5F and H). However, both SP3′N and LP1′F substitutions weaken P2 hydrogen bonds from ≈85 to 60% and less than 50%, respectively. These three P2 bonds (P2···N25, P2···D30, and P2···I85) do not occur in the WT complex, either. The viral advantage of weakening highly consistent hydrogen bonds formed in $^{LP1'F}$p1p6$_{D30N/N88D}$ could be to optimize substrate binding for optimal viral fitness. Stronger hydrogen bonding between the PR and the substrate could negatively impact substrate turnover, possibly resulting in product inhibition.

In ^{WT}p1-p6$_{WT}$, the substrate side chains have very consistent hydrogen bonds between P2′-D29′, P2′-D30′, and P4′-D30′ existing more than 90% of the time (Figure 5F). These hydrogen bonds are preserved in all mutant complexes from 75 to 96% of the time; hence, they should be crucial for the recognition of p1-p6 substrate (Figure 5G and H).

The natural substrates, in their bound conformation, have intramolecular hydrogen bonds that stabilize the preferred conserved consensus volume.[20] The drug-resistant PR– substrate complexes were evaluated for these intramolecular hydrogen bonds (Table 3). In ^{WT}NC-p1$_{WT}$, backbone nitrogen of the P1 position makes a hydrogen bond with the backbone oxygen of P3 position. This bond is lost upon V82A mutation in the PR but compensated by the secondary mutation AP2V in ^{AP2V}NC-p1$_{V82A}$. Similarly, in $^{LP1'F}$p1-p6$_{D30N/N88D}$, LP1′F mutation restores the hydrogen bond between P2 and P1′ positions of the substrate that is lost upon D30N/N88D PR mutations in ^{WT}p1-p6$_{D30N/N88D}$. In $^{SP3'N}$p1-p6$_{D30N/N88D}$ complex, on the other hand, no intramolecular hydrogen bond exists more than 50% of the time, suggesting that these hydrogen bonds are not necessary for this variant to adopt the required three-dimensional shape that fits the binding groove. These results point out that maintaining a stable intramolecular hydrogen-bonding network is only one of the various mechanisms through which the recognition motif is preserved by compensatory mutations.

2.4. Analysis of Atomic Fluctuations

The interdependency within the substrate residues was earlier shown[20] to play a key role in maintaining the balance between the conserved (i.e., their consensus volume and overall vdW contact potential with the PR) and varied (i.e., distribution of vdW contact potential across the substrate sequence and variations in the mean square fluctuations of the substrate residue side chains) properties of the substrates. The dynamic cooperativity within the substrate sequence as a way to maintain this balance was evaluated for drug-resistant

variants by computing the cross-correlation of the atomic positional fluctuations within the substrate. The heat maps for the NC-p1 and p1-p6 complexes are shown in Figures 6 and 7, respectively, where the correlation coefficients are color coded as a red to blue spectrum for full positive correlation to full negative correlation in the fluctuations. This analysis showed that ^{WT}NC-p1$_{V82A}$ is less cooperative compared to ^{WT}NC-p1$_{WT}$ (Figure 6A versus B). The cleavage site mutation AP2V appears to improve cooperativity (increased red regions in Figure 6C). The P4-P4′ residues of ^{AP2V}NC-p1$_{WT}$ are even more interdependent. This enhanced intrinsic cooperativity in ^{AP2V}NC-p1 should cause better communication within the substrate residues. The fact that NC-p1$_{AP2V}$ is a better processed substrate for even the WT PR supports that this interdependency within the substrate residues is a critical aspect of efficient substrate processing.

The p1-p6 substrate appears to be highly cooperative in both $^{LP1′F}$p1-p6$_{D30N/N88D}$ and $^{SP3′N}$p1-p6$_{D30N/N88D}$ complexes compared to ^{WT}p1-p6$_{WT}$ (Figure 7). In the ^{WT}p1-p6$_{WT}$ complex, the regions P4-P2′ and P3′-P4′ fluctuate almost like two independent domains, while the residues in each region are highly correlated with other residues in the same region (Figure 7A). The same substrate sequence behaves slightly differently in ^{WT}p1-p6$_{D30N/N88D}$ (Figure 7B), where the P4-P1 and P1′-P4′ residues define the two regions that fluctuate independently. Upon cleavage site mutations, these two distinct regions in the heat maps (Figure 7C and E) become less obvious. The cooperativity is evenly distributed across the substrate sequence through neighboring residues. These results show that correlated motion of the peptide sequence may be important to maintain the interdependent nature of the cleavage sites. In principle, both the substrate structures and PR–substrate interactions should have an impact on this interdependence. The alterations in the substrate conformations and PR–substrate interactions are not independent but rather highly interdependent.

3. Conclusions

Drug-resistant HIV-1 PR variants maintain substrate specificity, although they do not bind as tightly to PR inhibitors as the WT enzyme, which the inhibitors were designed to target. Resistance against these inhibitors is usually conferred by mutations in the PR gene. However, cleavage site mutations are also associated with resistance to certain PR inhibitors. Some of these substrate mutations co-occur with the primary drug resistance mutations in the PR. In this study, structural properties of two drug resistant PR variants (V82A and D30N/N88D) and the substrates that coevolved with these two variants (^{AP2V}NC-p1 and $^{LP1′F}$p1-p6/$^{SP3′N}$p1-p6, respectively) were investigated in conformational ensembles obtained from MD simulations. We found that the substrate envelope is preserved by the cleavage site mutations in the presence of primary drug-resistance mutations in the PR. These compensatory mutations make the substrate fit better within the envelope in the context of drug-resistant PR by increasing V_{in}. Consistency of these results ensures our understanding of the specificity determinants of substrate recognition. Protease– substrate coevolution validates the substrate envelope as the substrate recognition motif for HIV-1 PR.

The results of substrate coevolution and flexibility can potentially predict what other cleavage sites may be susceptible to coevolution. The increase in V_{in} is achieved by a variety of mechanisms: recovery of PR–substrate hydrogen bonding, vdW interactions, or the dynamic cooperativity within the substrate sequence. For each substrate we studied this increase in V_{in} is achieved by a different mechanism, but in every case upon mutation, the amino acid side chain becomes larger (Ala to Val in NC-p1 and Leu to Phe or Ser to Asp in p1-p6) resulting in substrate variants that better fill the substrate envelope. Thus we would expect that other compensatory mutations in the cleavage site would likely involve

mutations to larger residues. Our results also highlight the importance of the substrate conformational plasticity in the binding site. The mutant substrates that are intrinsically more flexible undergo local conformational rearrangements that assist a better fit within the substrate envelope. However this flexibility also can cause a larger V_{out}, making the sequence more susceptible to coevolution.

The exact effect of the substrate flexibility on the fit within the substrate envelope cannot be predicted solely on the cleavage site sequence. Our approach of combining the dynamic substrate envelope with the MD simulations of model structures of the mutant complexes can assist in this prediction. Theoretically, ΔV_{in} upon a single amino acid substitution could be estimated by combining *in silico* mutagenesis followed by energy minimization, MD simulations, and three-dimensional grid-based volume calculations. The substrate mutations resulting in higher ΔV_{in} values should be beneficial for the drug-resistant PR variants. Ultimately, the approach we present here could be useful in predicting the enzyme—substrate mutations that will more likely be tolerated in drug-resistant viral variants.

Being able to accurately predict the substrate coevolution is critical to avoid drug resistance in the drug design process. Our work highlights the importance of paying attention to the details of natural substrate recognition in designing inhibitors targeting resistant viral variants. We demonstrate, in this study, the impact of drug resistance mutations on substrate recognition in addition to inhibitor binding. Protease inhibitors, when designed to fit within the substrate envelope, are less likely to elicit PR mutations that can ultimately be compensated by mutations in the cleavage sites. In designing new inhibitors staying within the substrate envelope, the potency of the inhibitors must be maintained. Some inhibitors may protrude beyond the envelope to make favorable contacts with a PR residue to preserve affinity. In those cases, accurately predicting how likely for one or more substrates to coevolve with a potential resistance mutation in that particular PR residue is crucial. The availability of this probabilistic information can be used to decide at which location an inhibitor's protrusion beyond the envelope is more affordable in terms of avoiding the emergence of drug resistance via cleavage site mutations.

Researchers have so far used evolutionary information to predict the structural and functional features of proteins and their complexes. Sequence conservation allows for modeling the structure of a protein based on the experimentally determined structure of a homologous protein, while sequence variation, especially in the form of correlated mutations, is useful to infer functional features. This study, on the other hand, proposes the substrate envelope based on structural data as a tool to predict the evolution due to drug pressure for a PR–substrate system. Using the dynamic substrate envelope, the substrates can be assigned a probability of mutating in the presence of a particular PR mutation. To the best of our knowledge, this is the first structure-based approach to predict evolution under the selective pressure of drug therapy.

Using the Substrate Envelope in Prediction of Coevolution in Other Systems

In general, viral replication is a highly interdependent process that involves very complex interactions between separate viral macromolecules (proteins and polynucleotides) as well as host—virus interactions. This highly interdependent character of viral life cycle should make the coevolution of separate viral genes inevitable. In particular, HIV is a well-studied model system, with a massive amount of structural, biochemical, and sequence data on HIV available in the published literature and online databases,[23—26] allowing for indepth analysis to occur. In other viral systems far less data is available, but we expect the substrate envelope hypothesis to be applicable. For example, coevolution of individual sites within the hepatitis B viral genome has been reported to contribute to the rate of drug resistance.[27] In addition, correlated mutations in four of the influenza proteins have been reported to be

critical for host adaptation and pathogenicity,[28] demonstrating the interdependence of the viral genomic sites in evolution.

Hepatitis C virus (HCV) is also among these systems. The viral RNA-dependent RNA-polymerase, just like the HIV-1 reverse transcriptase, is inherently inaccurate accounting for a very high mutation rate.[29] A very recent Bayesian network analysis of viral polyprotein sequences has shown that the intrahost evolution of HCV is a complex process encoded in the interrelationships among many sites along the entire viral polyprotein.[30] Crystallographic studies from our group on the serine PR of this virus, NS3/4A, complexed with its natural substrates have shown that the substrate envelope hypothesis is valid also for HCV in explaining (1) the specificity determinants of the natural substrate recognition and (2) the emergence of primary drug resistance mutations.[31] The most severe resistance-conferring mutations occur where the inhibitors protrude from the NS3/4A substrate envelope, as these changes selectively weaken inhibitor binding without compromising the binding of substrates. The high mutation rate of HCV can also impact the sequence variability of the cleavage sites of NS3/4A, as a lesson learned from studying HIV-1 PR—substrate coevolution. No data on substrate coevolution have been reported so far. Nevertheless, the characteristics of the PR—substrate coevolution in HCV and other viral systems should be investigated. Our combined approach, involving molecular modeling, MD, and grid-based volume calculations, will better elucidate the interdependency of viral evolution by predicting the general sequence variability of each cleavage site and the particular cleavage site mutations that are likely to be selected in the context of a primary drug resistance PR mutation.

4. Materials and Methods

4.1. Nomenclature

HIV-1 PR (WT, V82A, or D30N/N88D) and the cleavage site (AP2V, LP1′F, or SP3′N) variants in a PR—substrate complex are designated by a subscript and a superscript to the name of the cleavage site. For example, $^{LP1'F}$p1-p6$_{D30N/N88D}$ denotes a complex of D30N/N88D PR variant with the LP1T mutant of the p1-p6 cleavage site, where LP1′F refers to a Leu-to-Phe mutation at P1′ position of the cleavage site.

4.2. Protease–Substrate Complex Structures

Two PR substrates and their mutant variants in complex with inactive, isosteric PR variants were investigated for structural and dynamic properties. All of these inactive PR variants have D25N mutation that has been shown to minimally alter the overall structure of the PR complexes.[32] This D25N mutation is not associated with drug-resistance; therefore, we refer to D25N as WT throughout the paper for simplicity in nomenclature. The crystal structures of ^{WT}p1-p6$_{WT}$[18] and ^{WT}NC-p1$_{WT}$[17] and ^{AP2V}NC-p1$_{V82A}$ variants were used as the starting structures in the MD simulations (PDB ID: 1KJF, 1TSU, 1TSQ). The other structures were modeled by side-chain mutations performed in PyMOL[33] by selecting the most probable rotamer from the rotamer library followed by energy minimization as described below. The PR—substrates complex variants are listed in Table 1.

4.3. Molecular Dynamics Simulations

The AMBER[34,35] simulation package (version 8) with the ff03 force field was used in all simulations. All structures were solvated explicitly in a truncated octahedron box using the TIP3P water model.[36] Overall charge of each system was neutralized by adding the appropriate number of Cl⁻ counterions using a Coulombic potential on a 1 Å grid with the preparatory program tleap of AMBER. The initial structures were first minimized at constant volume with convergence criteria of either maximum 90 000 cycles or a root-mean-

square deviation (rmsd) value of 0.01 Å by steepest-descent integration algorithm for 50 steps and then switched to conjugate gradient algorithm. Initial atom velocities corresponding to a temperature of 10 K were generated from a Maxwellian distribution, and the temperature was gradually raised to 300 K. The temperature was maintained at 300 K, and the pressure was maintained at 1 bar by the Berendsen weak-coupling approach.[37] Constant pressure periodic boundary conditions were used with isotropic position scaling. The particle mesh Ewald (PME) method[38] was used to calculate the full electrostatic energy of a periodic box, bypassing pair list creation and nonbonded force and energy evaluation by calling special PME functions to calculate the Lennard-Jones and electrostatic interactions with a cutoff distance of 9 Å. Covalent bonds involving hydrogen atoms were constrained by the SHAKE algorithm[39] with a relative geometrical tolerance of 10×10^{-5} Å. A time step of 2 fs was employed in the Leapfrog integrator. Various system properties (total, kinetic, and potential energies, temperature, pressure, density, and rmsd of the backbone atoms, and $V_{in/out}$) were monitored during equilibration to ensure the stability of the simulations. Coordinates and energies were written every 0.4 ps during the 11 ns production phase.

4.4. Calculation of V_{in} and V_{out}

The dynamic substrate envelope was previously modeled by a quantitative approach using the MD of seven PR–substrate WT complexes.[20] This approach was based on a three-dimensional grid with side length 10 Å and grid spacing 0.2 Å located in the binding site. The occupancy of each grid cell, g_{ijk}, was assigned an initial value of 0. Then the number of times that a grid cell was within the vdW volume of any peptide was counted. This value eventually became the overall occupancy of that grid cell. The total volume of the substrate envelope was essentially the summation over all grid cells with occupancy greater than 0 normalized by the total number of structures used. This dynamic substrate envelope was used in this study as a basis for the conserved consensus volume that is occupied by the substrates in the binding groove.

In this study, the substrate vdW volumes were mapped on the same grid to define probability distributions of the volume occupied by each substrate variant. Before this mapping, all frames from each trajectory were superimposed using the least flexible/mobile residues 24–26 and 85–90 on the $^{WT}CA-p2_{WT}$ complex structure using the rmsd trajectory tool of the molecular visualization and trajectory analysis software, VMD.[40] The effect of residue selection on the shape of the dynamic substrate envelope was assessed by aligning the trajectories using the C_α atoms of all PR residues, dimerization interface excluding the flaps, catalytic triad, and finally only the flaps. The results showed that the shape of the dynamic substrate envelope is not sensitive to the selection of the reference residues for structural alignment as long as these reference residues are not on a highly mobile region of the PR (Supporting Information Tables S1 and S2).

The substrates in various complexes were then compared to each other in terms of how well they fit within (V_{in}) and how much they protruded beyond the dynamic substrate envelope (V_{out}). For this comparison, the vdW volume of an individual substrate was computed from N_I time points in the MD trajectory of that substrate as described earlier.[20]

4.5. Estimation of vdW Potential

Protease–substrate vdW contacts were estimated by a simplified Lennard-Jones potential $V(r)$ using the relation $4\varepsilon [(\sigma/r)^{12} — (\sigma/r)^6]$, where r is the PR–substrate interatomic distance and ε and σ are the well depth and hard sphere diameter, respectively, for each PR–substrate atom pair. $V(r)$ for all possible PR–substrate atom pairs was computed within 5 Å, and when the distance between nonbonded pairs was less than ε, $V(r)$ was considered equal to ε. The rationale for this modification to the original 6-12 Lennard-Jones potential was

previously described in detail.[20] Using this simplified potential for each nonbonded PR–substrate pair, $\Sigma V(r)$ was then computed for each PR and each substrate residue.

4.6. Calculation of ΔV_{in}, ΔV_{out}, and ΔvdW

The property of interest (V_{in}, V_{out}, or vdW) was calculated for each frame taken from an MD trajectory as described above. These time series data were first plotted as a histogram to ensure Gaussian distribution, and the sample mean and standard error were calculated. The difference between two distributions (for example $\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$) was computed by subtracting the sample mean of one from the other. The corresponding error was estimated by combining the individual standard errors (SE) in quadrature assuming V_{in}^{WT} and V_{in}^{MUT} are independent variables.

$$SE_{\Delta\bar{x}} = \sqrt{SE_{\bar{x}_{MUT}}^2 + SE_{\bar{x}_{WT}}^2} \quad (1)$$

Finally, the calculated parameters were reported as $\Delta x \pm SE_{\Delta\bar{x}}$ in Figures 1 and 3, where x is V_{in}, V_{out}, or vdW. The details of the mathematical proof can be found elsewhere.[41]

4.7. Evaluation of Hydrogen Bonding

The trajectory analysis program ptraj was used to calculate the percentage of time a hydrogen bond existed during the simulated trajectories. A hydrogen bond was defined by a distance between the donor and acceptor of less than 3.5 A and a donor—hydrogen—acceptor angle of greater than 120°. Hydrogen bonds that existed more than 50% of the time were analyzed.

4.8. Fluctuation Dynamics

The normalized cross-correlations of residue pairs were defined as

$$CO_{i,j} = \frac{\langle \Delta R_i \Delta R_j \rangle}{\langle \Delta R_i^2 \rangle^{1/2} \langle \Delta R_j^2 \rangle^{1/2}} \quad (2)$$

where ΔR_i is the fluctuation in the position vector R of site i and ΔR_j is the fluctuation in the position vector R of site j. The brackets represent time averages over recorded snapshots. The cross-correlations vary in the range $[-1, 1]$, with the lower and upper limits indicating fully anticorrelated and correlated atomic fluctuations, respectively. $CO_{i,j} = 0$ gives uncorrelated atomic fluctuations.

The computer programs for calculating V_{in}, V_{out}, vdW contact potential, and the cross correlations of the atomic fluctuations were written in FORTRAN programming language. The vdW radii necessary for the volume and vdW contact potential calculations were taken from the amber03 force field.[42] The error analysis was performed using MATLAB.[43]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA. 1988; 85(13):4686–4690. [PubMed: 3290901]

2. Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature. 1989; 337(6208):615–620. [PubMed: 2645523]

3. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science. 1989; 245(4918):616–621. [PubMed: 2548279]

4. Flexner C. HIV drug development: the next 25 years. Nat Rev Drug Discovery. 2007; 6(12):959–966.

5. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science. 1995; 267(5197):483–489. [PubMed: 7824947]

6. Ji JP, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochemistry. 1992; 31(4):954–958. [PubMed: 1370910]

7. Bally F, Martinez R, Peters S, Sudre P, Telenti A. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. AIDS Res Hum Retroviruses. 2000; 16(13):1209–1213. [PubMed: 10957718]

8. Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L, Lamarre D. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. J Virol. 1996; 70(6):3763–3769. [PubMed: 8648711]

9. Mammano F, Petit C, Clavel F. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. J Virol. 1998; 72(9):7632–7637. [PubMed: 9696866]

10. Feher A, Weber IT, Bagossi P, Boross P, Mahalingam B, Louis JM, Copeland TD, Torshin IY, Harrison RW, Tozser J. Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. Eur J Biochem. 2002; 269(16):4114–4120. [PubMed: 12180988]

11. La Seta Catamancio S, De Pasquale MP, Citterio P, Kurtagic S, Galli M, Rusconi S. In vitro evolution of the human immunodeficiency virus type 1 gag-protease region and maintenance of reverse transcriptase resistance following prolonged drug exposure. J Clin Microbiol. 2001; 39(3):1124–1129. [PubMed: 11230439]

12. Zhang YM, Imamichi H, Imamichi T, Lane HC, Falloon J, Vasudevachari MB, Salzman NP. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. J Virol. 1997; 71(9):6662–6670. [PubMed: 9261388]

13. Kolli M, Lastere S, Schiffer CA. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. Virology. 2006; 347(2):405–409. [PubMed: 16430939]

14. Pai VB, Nahata MC. Nelfinavir mesylate: a protease inhibitor. Ann Pharmacother. 1999; 33(3):325–339. [PubMed: 10200859]

15. Patick AK, Duran M, Cao Y, Shugarts D, Keller MR, Mazabel E, Knowles M, Chapman S, Kuritzkes DR, Markowitz M. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. Antimicrob Agents Chemother. 1998; 42(10):2637–2644. [PubMed: 9756769]

16. Kolli M, Stawiski E, Chappey C, Schiffer CA. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. J Virol. 2009; 83(21):11027–11042. [PubMed: 19706699]

17. Prabu-Jeyabalan M, Nalivaika EA, King NM, Schiffer CA. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. J Virol. 2004; 78(22):12446–12454. [PubMed: 15507631]

18. 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(3):369–381. [PubMed: 12005435]

19. King NM, Prabu-Jeyabalan M, Nalivaika EA, Schiffer CA. Combating susceptibility to drug resistance: lessons from HIV-1 protease. Chem Biol. 2004; 11(10):1333–1338. [PubMed: 15489160]

20. Ozen A, Haliloglu T, Schiffer CA. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. J Mol Biol. 2011; 410(4):726–744. [PubMed: 21762811]

21. Cote HC, Brumme ZL, Harrigan PR. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. J Virol. 2001; 75(2):589–594. [PubMed: 11134271]

22. Pettit SC, Sheng N, Tritch R, Erickson-Viitanen S, Swanstrom R. The regulation of sequential processing of HIV-1 Gag by the viral protease. Adv Exp Med Biol. 1998; 436:15–25. [PubMed: 9561194]

23. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 2003; 31(1):298–303. [PubMed: 12520007]

24. Shafer RW. Rationale and uses of a public HIV drug-resistance database. J Infect Dis. 2006; 194(Suppl 1):S51–58. [PubMed: 16921473]

25. HIV Databases. http://www.hiv.lanl.gov

26. HIV-1, Human Protein Interaction Database. http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/

27. Khudyakov Y. Coevolution and HBV drug resistance. Antiviral Ther. 2010; 15(3 Pt B):505–515.

28. Hu W. Correlated mutations in the four influenza proteins essential for viral RNA synthesis, host adaptation, and virulence: NP, PA, PB1, and PB2. Nat Sci. 2010; 2(10):1138–1147.

29. Qureshi SA. Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon. Med Res Rev. 2007; 27(3):353–373. [PubMed: 16958135]

30. Lara J, Xia G, Purdy M, Khudyakov Y. Coevolution of the hepatitis C virus polyprotein sites in patients on combined pegylated interferon and ribavirin therapy. J Virol. 2011; 85(7):3649–3663. [PubMed: 21248044]

31. Romano KP, Ali A, Royer WE, Schiffer CA. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. Proc Natl Acad Sci USA. 2010; 107(49):20986–20991. [PubMed: 21084633]

32. Sayer JM, Liu F, Ishima R, Weber IT, Louis JM. Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. J Biol Chem. 2008; 283(19):13459–13470. [PubMed: 18281688]

33. The PyMOL Molecular Graphics System. Schrodinger, LLC; New York: 2011. version 1.4

34. Case, DA.; Cheatham, TE.; Darden, TA.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Mertz, KM.; Wang, B.; Pearlman, DA.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, P.; Caldwell, JW.; Ross, WS.; Kollman, PA. AMBER 8. University of California; San Francisco, CAM: 2004.

35. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem. 2005; 26(16):1668–1688. [PubMed: 16200636]

36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935.

37. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. J Chem Phys. 1984; 81(8):3684–3690.

38. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995; 103(19):8577–8593.

39. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys. 1977; 23(3): 327–341.

40. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graphics. 1996; 14(1):33–38. 27-38.

41. Miller, I.; Miller, M. Mathematical Expectation. In: Heath, A.; Grant, J., editors. John E Freund's Mathematical Statistics. 6th. Prentice-Hall; Upper Saddle River, NJ: 1999. p. 158-160.

42. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem. 2003; 24(16):1999–2012. [PubMed: 14531054]
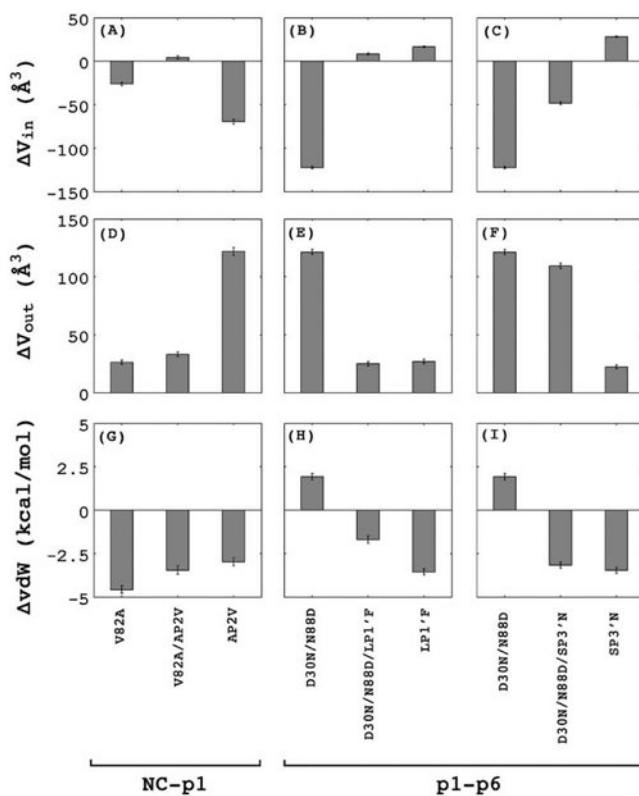
43. MATLAB, R2008b. The MathWorks, Inc; Natick, MA: 2008.

Figure 1.
Protease-substrate coevolution preserves the dynamic substrate envelope. ΔV_{in}, ΔV_{out}, and overall PR—substrate ΔvdW interactions are shown in panels A—C, D—F, and G—I, respectively. The difference in these three properties for each PR—substrate variant compared to the WT PR—substrate complex is plotted as bars. Labels of the x-axes correspond to the mutations in either the PR (V82A or D30N/N88D) or the cleavage site (AP2V, LP1′F, or SP3′N) or both.

$(\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$, $\Delta V_{out} = V_{out}^{MUT} - V_{out}^{WT}$, and $\Delta vdW = vdW^{MUT} - vdW^{WT})$.

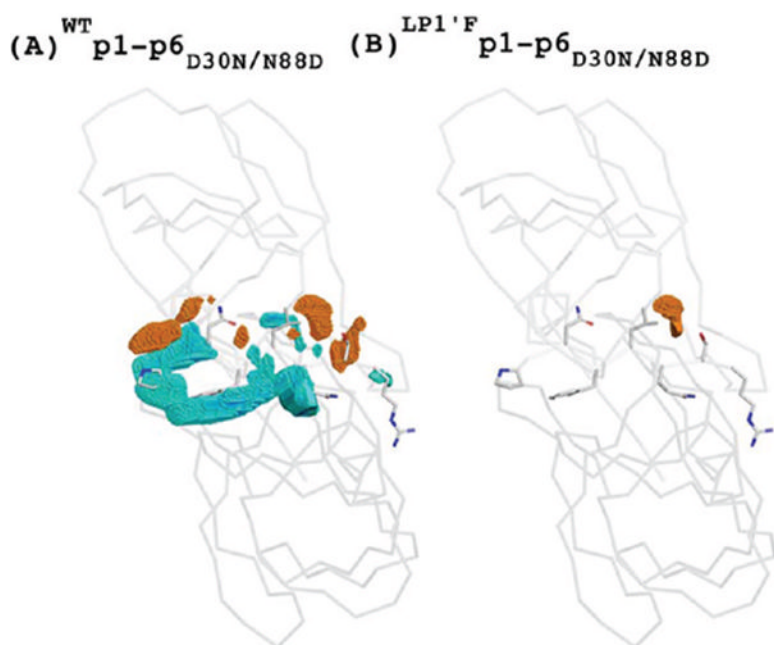(A) ^{WT}p1-p6$_{D30N/N88D}$ (B) $^{LP1'F}$p1-p6$_{D30N/N88D}$

Figure 2.
The compensatory mutation LP1′F in the p1-p6 cleavage site preserves the substrate envelope in drug-resistant D30N/N88D PR variant. δV_{in} is visualized for (A) ^{WT}p1-p6$_{D30N/N88D}$ and (B) $^{LP1'F}$p1-p6$_{D30N/N88D}$ by mapping the difference matrix (V^{MUT} — V^{WT}) onto the cocrystal structure. The negative values of the difference matrix ΔV_{in}, shown in cyan, correspond to the regions where the substrate in mutant complex fits poorly within the dynamic substrate envelope compared to ^{WT}p1-p6$_{WT}$. The positive values of the difference matrix ΔV_{in}, shown in orange, correspond to the regions where the substrate in mutant complex fits better within the dynamic substrate envelope compared to ^{WT}p1-p6$_{WT}$. To get clear images, only the grid cells with an occupancy factor of 0.5 or higher are visualized. The PR dimer is represented as light-gray ribbon, and p1-p6 side-chains are shown as sticks. In both panels, the substrate is oriented from N- to C-terminus (left to right). The figure was prepared with the molecular visualization software, PyMOL.[33]
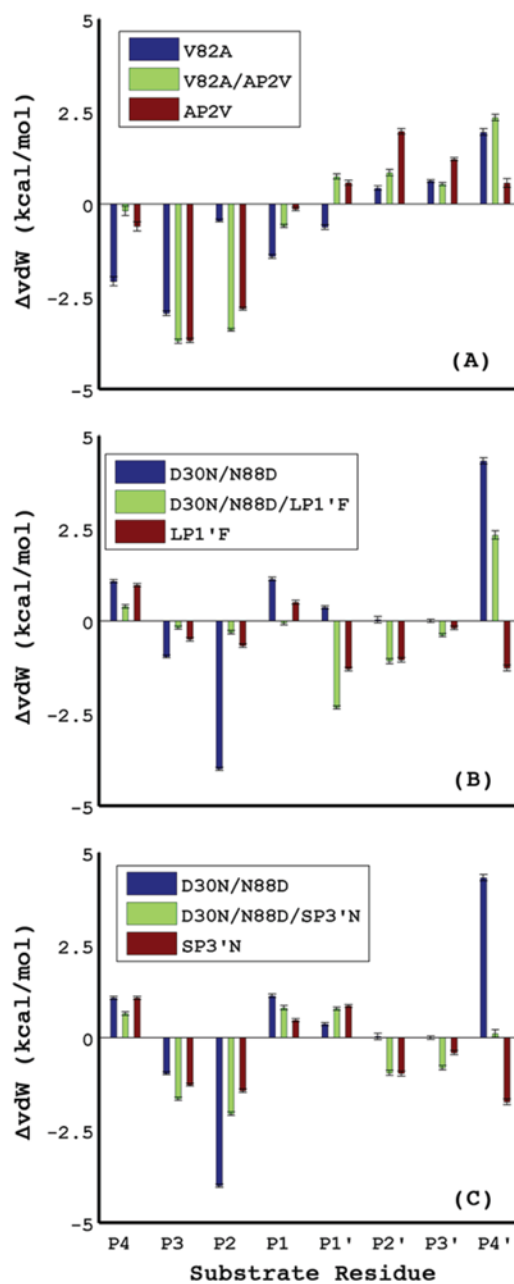
Figure 3.
Residue-based view of the effect of drug resistance on PR— substrate vdW interactions. Substrate residual δvdW interactions with the PR for (A) NC-p1 variants (^{WT}NC-p1$_{V82A}$, ^{AP2V}NC-p1$_{V82A}$, and ^{WT}NC-p1$_{V82A}$); (B) the first set of p1-p6 variants (^{WT}p1-p6$_{D30N/N88D}$, $^{LP1'F}$p1- p6$_{D30N/N88D}$, and $^{LP1'F}$p1-p6$_{WT}$); and (C) the second set of p1-p6 variants (^{WT}p1-p6$_{D30N/N88D}$, $^{SP3'N}$p1-p6$_{D30N/N88D}$, and $^{SP3'N}$p1-p6$_{WT}$) with reference to their respective WT complexes (ΔvdW = vdW^{MUT} — vdW^{WT}).
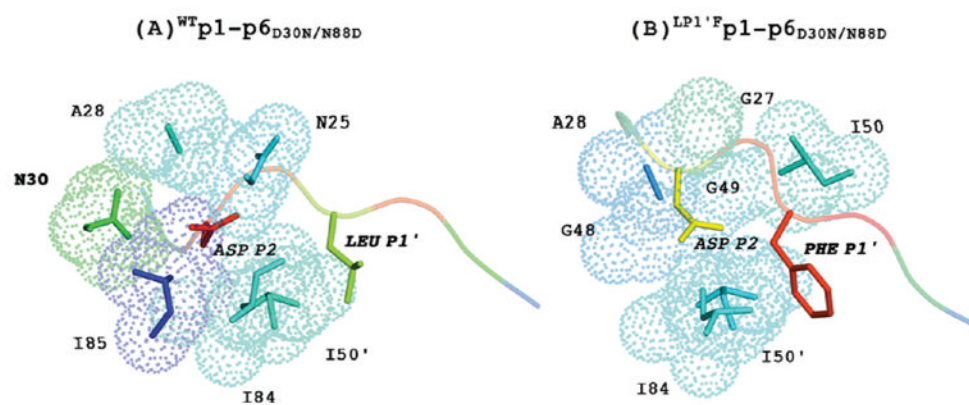
Figure 4.
The compensatory mutation in the p1-p6 cleavage site, LP1′F, causes major structural rearrangements reducing vdW interaction potential of the invariant Asp P2. The side chains of the PR residues that interact with Asp P2 at 10 ns are displayed as dotted spheres for (A) ^{WT}p1-p6$_{d30N/N88D}$ and (B) $^{LP1′F}$p1-p6$_{D30N/N88D}$. The substrate residues were colored based on the vdW interaction potential with any PR atom, and the PR residues were colored based on the vdW interaction potential with any substrate atom. The substrate residues are shown in italic and the mutated residues in bold.
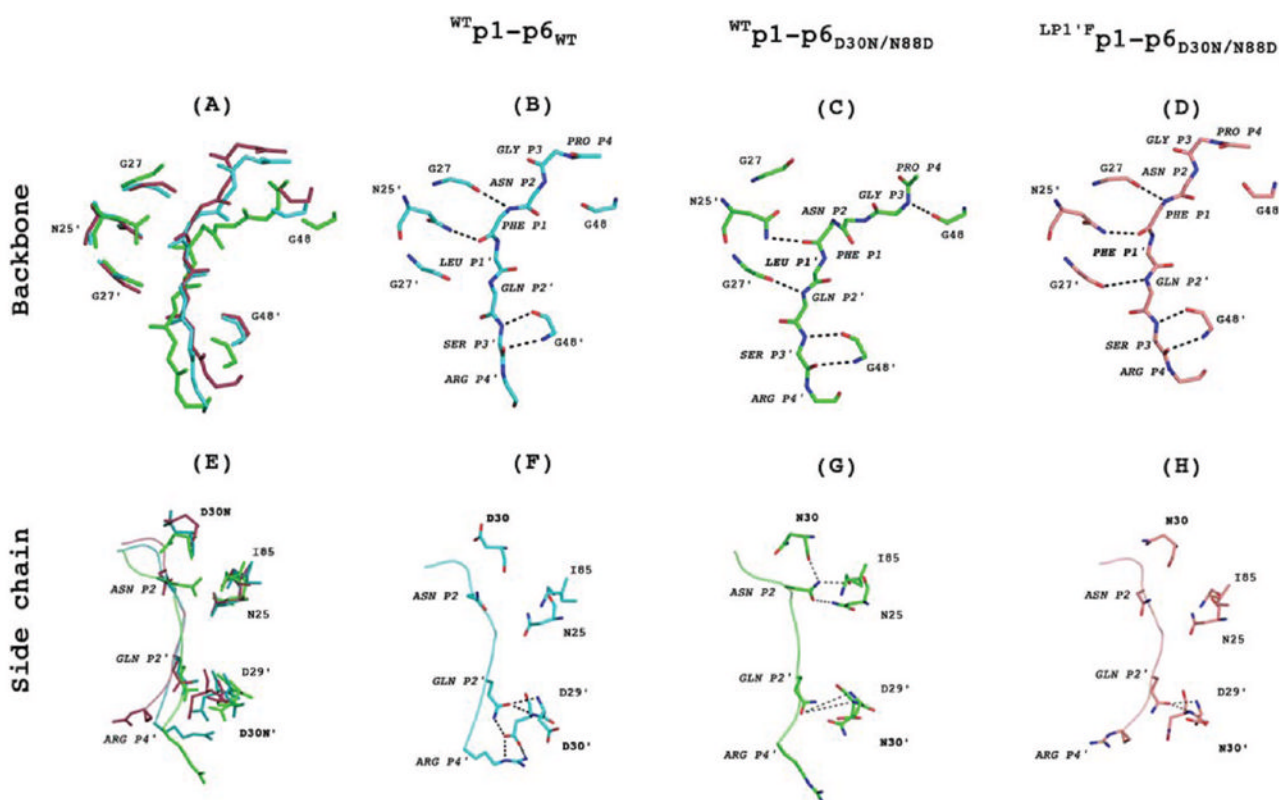
Figure 5.
Substrate backbone hydrogen bonds are robust against the local conformational rearrangements due to drug-resistance mutations. The side-chain hydrogen-bonding pattern, however, is altered by PR mutations but restored by cleavage site mutation. Coordinates of ^{WT}p1-p6$_{WT}$, ^{WT}p1p6$_{D30N/N88D}$, and $^{LP1'F}$p1p6$_{D30N/N88D}$ after 10 ns were superimposed to reveal the local conformational rearrangements due to the drug resistance mutations (A, E). Substrate backbone (A) and side chain (E) atoms are displayed as sticks. The backbone and side-chain hydrogen bonds are shown as dashed lines for (B and F) ^{WT}p1-p6$_{WT}$, (C and G) ^{WT}p1p6$_{D30N/N88D}$, and (D and H) $^{LP1'F}$p1p6$_{D30N/N88D}$. The substrate residues are labeled in italic and mutated residues in bold.

Figure 6.
Effect of drug resistance on intrinsic dynamic cooperativity of NC-p1 substrate variants. The cross-correlations of the atomic positional fluctuations are color-coded (red: positively correlated and blue: negatively correlated).
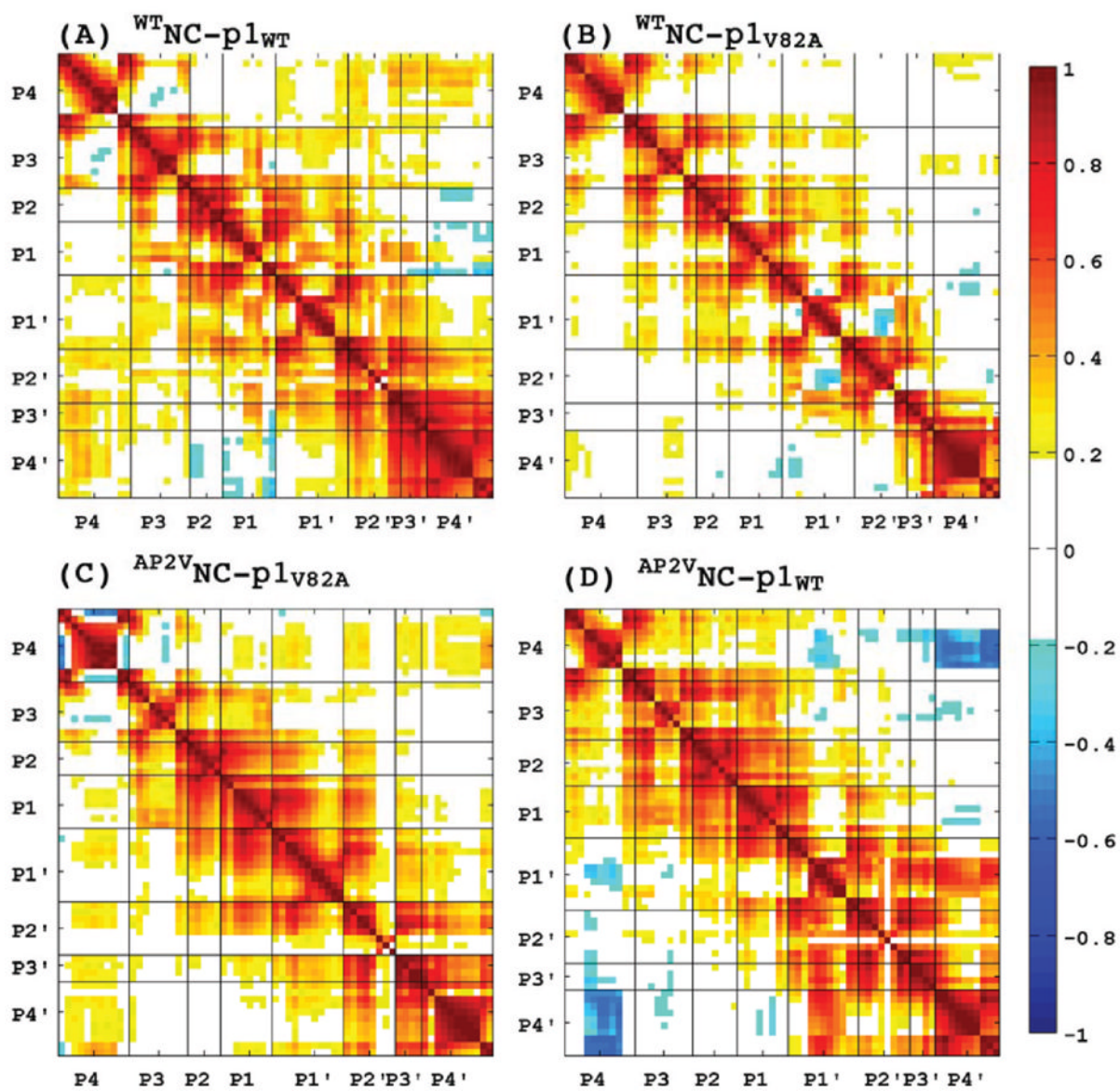
Figure 7.
Effect of drug resistance on intrinsic dynamic cooperativity of p1-p6 substrate variants. The cross-correlations of the atomic positional fluctuations are color-coded (red: positively correlated and blue: negatively correlated).
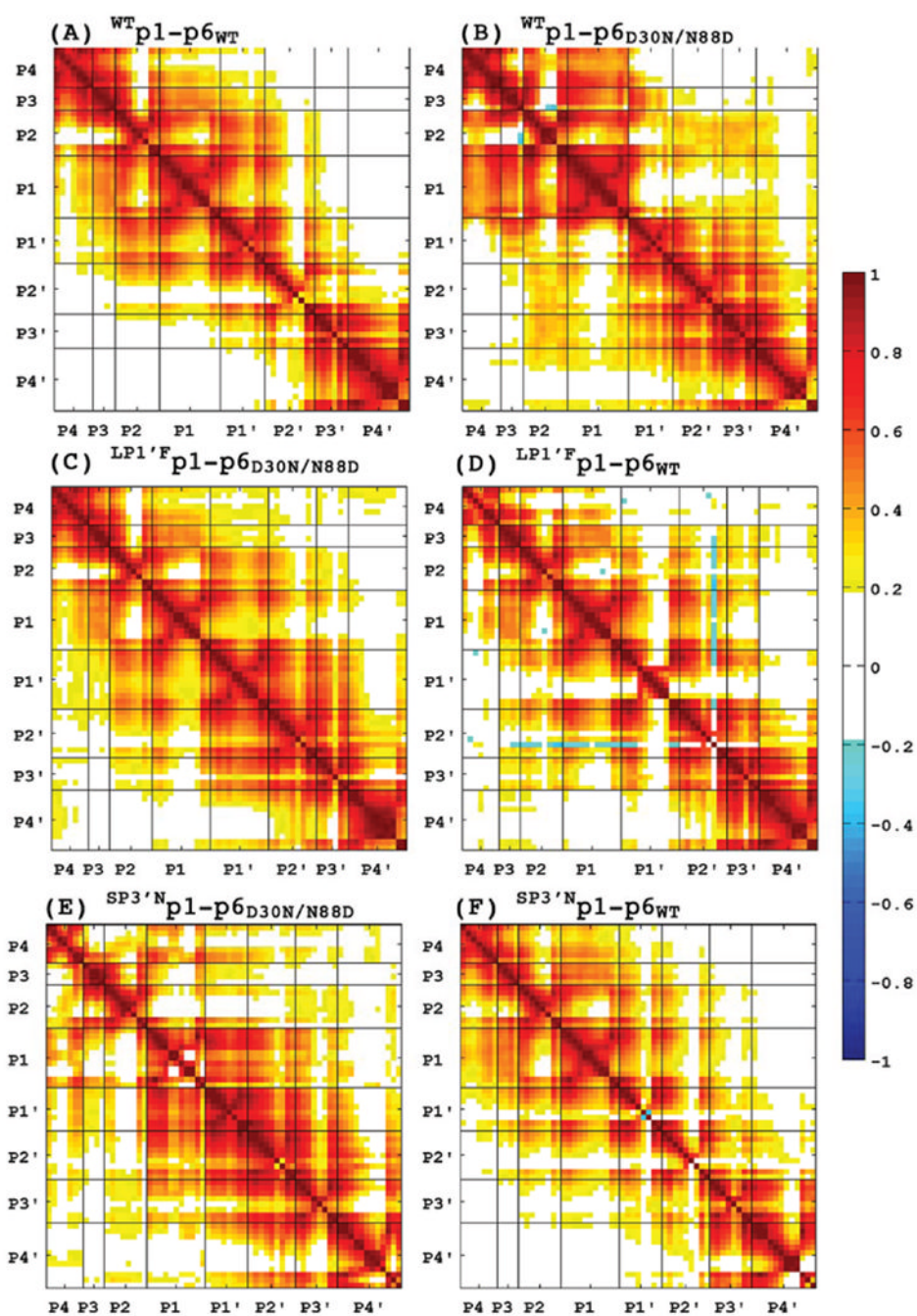
Table 1
Protease—Substrate Complex Variants

| protease—substrate complex variant | drug-resistance mutations | | starting structure (PDB code) |
	protease (D25N)[a]	substrate	
^{WT}p1-p6$_{WT}$	—	—	^{WT}p1-p6$_{WT}$ (1KJF)
$^{LP1'F}$p1-p6$_{WT}$	—	LP1′F	^{WT}p1-p6$_{WT}$ (1KJF)
$^{SP3'N}$p1-p6$_{WT}$	—	SP3N	^{WT}p1-p6$_{WT}$ (1KJF)
^{WT}p1-p6$_{D30N/N88D}$	D30N/N88D	—	^{WT}p1-p6$_{WT}$ (1KJF)
$^{LP1'F}$p1-p6$_{D30N/N88D}$	D30N/N88D	LP1′F	^{WT}p1-p6$_{WT}$ (1KJF)
$^{SP3'N}$p1-p6$_{D30N/N88D}$	D30N/N88D	SP3N	^{WT}p1-p6$_{WT}$ (1KJF)
^{WT}NC-P1$_{WT}$	—	—	^{WT}NC-P1$_{WT}$ (1TSU)
^{AP2V}NC-p1$_{WT}$	—	AP2V	^{WT}NC-P1$_{WT}$ (1TSU)
^{WT}NC-p1V82A	V82A	—	^{WT}NC-P1$_{WT}$ (1TSU)
^{AP2V}NC-p1V82A	V82A	AP2V	^{AP2V}NC-p1$_{V82A}$ (1TSQ)

[a]All 10 PR variants have D25N mutation. This mutation inactivates the PR allowing for natural substrate cocrystal structure determination and is known to hardly alter the structure.32 D25N is not a drug-resistance mutation; hence, the PR with only D25N is referred WT to simplify the notation.

Table 2

Effect of Drug Resistance on Intermolecular Hydrogen-Bonding Pattern[a]

Backbone — NC-p1

	WTNC-p1WT	WTNC-p1V82A	AP2VNC-p1V82A	AP2VNC-p1WT
P3...D29	<50	92.3	99.8	99.1
P2...G48	97.6	80.1	98.7	100.0
P1...N25	67.5	83.7	59.6	54.4
P1...N25'	<50	79.1	<50	<50
P2'...G27'	86.0	91.5	53.8	<50
P2'...D29'	58.0	<50	<50	<50
P3'...G48'	96.9	92.1	78.2	66.2
P4'...G48'	<50	<50	62.0	<50

Side-chain — NC-p1

	WTNC-p1WT	WTNC-p1V82A	AP2VNC-p1V82A	AP2VNC-p1WT
P4...D30	62.7	73.8	52.4	79.8
P1...150	<50	<50	98.4	95.5
P1...N25'	63.1	<50	<50	<50
P1...G27	<50	<50	77.8	83.6

Backbone — p1-p6 (LP1'F)

	WTp1-p6WT	WTp1-p6D30N/N88D	LP1'Fp1-p6D30N/N88D	LP1'Fp1-p6WT
P4...R8'	<50	<50	<50	<50
P3...G48	<50	90.7	<50	55.5
P3...D29	<50	<50	<50	<50
P2...G48	<50	<50	<50	72.4
P1...G27	97.3	<50	98.8	97.1
P1...N25'	72.9	95.9	81.2	64.25
P2'...G27'	<50	91.5	76.5	94.9
P3'...G48'	99.8	98.3	99.7	99.3
P4'...D29'	<50	<50	<50	<50

Side-chain — p1-p6 (LP1'F)

	WTp1-p6WT	WTp1-p6D30N/N88D	LP1'Fp1-p6D30N/N88D	LP1'Fp1-p6WT
P2...N25	<50	84.7	<50	<50
P2...D30	<50	83.1	<50	<50
P2...I85	<50	52.7	<50	<50
P2'...D29'	91.6	81.3	74.7	86.8
P2'...D30'	89.1	87.6	86.8	81.8
P4'...D30'	96.0	<50	<50	91.1

Backbone — p1-p6 (SP3'N)

	WTp1-p6WT	WTp1-p6D30N/N88D	SP3'Np1-p6D30N/N88D	SP3'Np1-p6WT
P4...R8'	<50	<50	<50	60.7
P3...G48	<50	90.7	<50	70.0
P3...D29	<50	<50	87.3	59.1
P2...G48	<50	<50	<50	83.3
P1...G27	97.3	<50	<50	98.2
P1...N25'	72.9	95.9	92.4	89.1

Side-chain — p1-p6 (SP3'N)

	WTp1-p6WT	WTp1-p6D30N/N88D	SP3'Np1-p6D30N/N88D	SP3'Np1-p6WT
P2...N25	<50	84.7	57.8	<50
P2...D30	<50	83.1	60.7	<50
P2...I85	<50	52.7	<50	<50
P2'...D29'	91.6	81.3	87.1	58.0
P2'...D30'	89.1	87.6	96.2	61.1
P4'...D30'	96.0	<50	<50	98.2

	Backbone				Side-chain			
	WTNC-p1WT	WTNC-p1V82A	AP2VNC-p1V82A	AP2VNC-p1WT	WTNC-p1WT	WTNC-p1V82A	AP2VNC-p1V82A	AP2VNC-p1WT
P2′…G27′	<50	91.5	98.4	97.5				
P3′…G48′	99.8	98.3	93.1	88.7				
P4′…D29′	<50	<50	63.3	50.7				

[a] Percent time that hydrogen bonds existed between HIV-1 PR and variants of NC-p1 and p1-p6 substrates are shown separately for substrate backbone and side chains. The table is color-coded based on the frequency of hydrogen bonds as a spectrum; red being highest frequency and green being lowest frequency hydrogen bonds. Only hydrogen bonds that existed more than 50% of the time were analyzed.

Table 3

Effect of Drug Resistance on Intramolecular Substrate Hydrogen-Bonding Pattern[a]

	WTNC-p1WT	WTNC-p1V82A	AP2VNC-p1V82A	AP2VNC-p1WT
P1(N)…P3(O)	61.8	<50	<50	<50
P1(ND2)…P3(OE1)	<50	<50	78.0	81.6

	WTp1-p6WT	WTp1-p6D30N/N88D	LP1′Fp1-p6D30N/N88D	LP1′Fp1-p6WT
P2(O)…P1′ (N)	<50	79.9	<50	<50
P2(ODl)…P1′ (N)	87.3	<50	83.7	69.0
P3′ (O)…P2′ (NE2)	<50	61.6	73.7	<50

	WTp1-p6WT	WTp1-p6D30N/N88D	SP3′Np1-p6D30N/N88D	SP3′Np1-p6WT
P2(O)…P1′ (N)	<50	79.9	<50	<50
P2(ODl)…Pl′ (N)	87.3	<50	<50	71.6
P3′ (O)…P2′ (NE2)	<50	61.6	<50	<50

[a]Percent time that substrate intramolecular hydrogen bonds existed is shown for each PR—substrate complex. The table is color-coded based on the frequency of hydrogen bonds as a spectrum; red being highest frequency and green being lowest frequency hydrogen bonds. Only hydrogen bonds that existed more than 50% of the time were analyzed.