10-28-2004

Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease

Moses Prabu-Jeyabalan
University of Massachusetts Medical School

Ellen A. Nalivaika
University of Massachusetts Medical School

Nancy M. King
University of Massachusetts Medical School

See next page for additional authors

Follow this and additional works at: http://escholarship.umassmed.edu/oapubs
Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

Repository Citation
http://escholarship.umassmed.edu/oapubs/1521

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease

Authors
Moses Prabu-Jeyabalan, Ellen A. Nalivaika, Nancy M. King, and Celia A. Schiffer
Structural Basis for Coevolution of a Human Immunodeficiency Virus Type 1 Nucleocapsid-p1 Cleavage Site with a V82A Drug-Resistant Mutation in Viral Protease

Moses Prabu-Jeyabalan, Ellen A. Nalivaika, Nancy M. King, and Celia A. Schiffer*

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts

Received 7 May 2004/Accepted 12 July 2004

Maturation of human immunodeficiency virus (HIV) depends on the processing of Gag and Pol polyproteins by the viral protease, making this enzyme a prime target for anti-HIV therapy. Among the protease substrates, the nucleocapsid-p1 (NC-p1) sequence is the least homologous, and its cleavage is the rate-determining step in viral maturation. In the other substrates of HIV-1 protease, P1 is usually either a hydrophobic or an aromatic residue, and P2 is usually a branched residue. NC-p1, however, contains Asn at P1 and Ala at P2. In response to the V82A drug-resistant protease mutation, the P2 alanine of NC-p1 mutates to valine (AP2V). To provide a structural rationale for HIV-1 protease binding to the NC-p1 cleavage site, we solved the crystal structures of inactive (D25N) WT and V82A HIV-1 proteases in complex with their respective WT and AP2V mutant NC-p1 substrates. Overall, the WT NC-p1 peptide binds HIV-1 protease less optimally than the AP2V mutant, as indicated by the presence of fewer hydrogen bonds and fewer van der Waals contacts. AlaP2 does not fill the P2 pocket completely; PheP1’ makes van der Waals interactions with Val82 that are lost with the V82A protease mutation. This loss is compensated by the AP2V mutation, which reorients the peptide to a conformation more similar to that observed in other substrate-protease complexes. Thus, the mutant substrate not only binds the mutant protease more optimally but also reveals the interdependency between the P1’ and P2 substrate sites. This structural interdependency results from coevolution of the substrate with the viral protease.

Human immunodeficiency virus type 1 (HIV-1) matures after the viral protease processes (35) the Gag and Pol polyproteins at 10 substrate locations (3, 15). Therefore, inhibition of HIV-1 protease represents an important avenue for antiviral therapy (13, 48). The substrate sequences cleaved by the protease are nonhomologous (3), with the sequence of the nucleocapsid-p1 (NC-p1) substrate being the most different. In spite of the poor sequence homology among the substrate sites, a series of substrate-protease crystal structures led us to hypothesize that substrate specificity in HIV-1 protease results from the enzyme’s recognizing an asymmetric shape (or envelope) rather than a particular amino acid sequence (40). This shape results from the conformation that a particular substrate sequence can adopt, implying that an interdependency necessarily exists among the different substrate residue sites.

All of the protease inhibitors whose designs were structure based bind competitively (8, 18, 28, 52, 53) at the active site. Since these inhibitors bind at the same site as the substrates, many protease residues contact both substrates and inhibitors. Drug resistance, which often develops in the presence of therapeutic protease inhibitors, results from high viral turnover, the infidelity of the viral reverse transcriptase (16, 42, 43), and selective pressure on the virus. With drug resistance, the protease no longer binds as tightly to inhibitors but retains the ability to recognize and cleave its substrate sites.

The first drug-resistant mutation that often occurs in patients, V82A (47), is frequently associated with indinavir or ritonavir therapy (5, 9, 11, 30). Our previous studies have shown that V82A is a prime site for drug resistance to occur, as the valine at residue 82 is not critical for substrate recognition but does extensively contact many of the commonly used inhibitors (39). However, the apparent coevolution of the NC-p1 cleavage site (AP2V) with the V82A mutation in the protease (1, 10, 20, 55) implies that Val82 plays an important role in the protease’s recognition of the NC-p1 substrate, a role that is lost when the V82A mutation occurs.

Within the viral context, the NC-p1 cleavage site is the slowest (50), final (37, 54), and, therefore, rate-determining site in HIV-1 Gag to be processed by the viral protease (6, 37). NC-p1 has a polar Asn at the P1 position, while a hydrophobic or aromatic residue is found at the same location in the other substrate sequences. An Ala occupies the P2 position of the NC-p1 substrate, while a branched side chain (14) exists at P2 in the other substrate sequences. The rate and order in which NC-p1 is cleaved may be dictated by its unusual sequence as well as by the location of this site within the Gag polyprotein.

In several studies of drug-experienced patient viral sequences, the NC-p1 substrate cleavage site has been seen to mutate (1, 6, 10, 24, 44, 55). The most frequently observed change occurs at P2, where the alanine mutates to a valine in viral sequences that also contain the V82A drug-resistant protease mutation. Selection for a valine at this site makes sense as valine is the wild-type (WT) residue at P2 in another substrate sequence, the capsid-p2 cleavage site. In fact, the WT HIV-1 protease cleaves AP2V in the NC-p1 cleavage site more efficiently than it cleaves the WT NC-p1 sequence (12).
The present study focuses on elucidating the structural rationale for HIV-1 protease recognition of the NC-p1 cleavage site. We determined the crystal structures of both the WT and V82A HIV-1 proteases in complex with their respective NC-p1 substrates. As seen in our previous studies (40), there is an interdependency between the substrate subsites throughout the cleavage site sequence that allows for the sequence to be recognized as a substrate sequence by HIV-1 protease.

**MATERIALS AND METHODS**

**Nomenclature.** HIV-1 protease variants (WT or V82A), the nucleocapsid-p1 (NC-p1) WT substrate, and its AP2V mutant substrate will be distinguished throughout this article by subscript and superscript acronyms. For example, WTNC-p1WT and AP2VNC-p1V82A denote the WT and AP2V mutant complexes in solution with the NC-p1 peptide, and AP2VNC-p1V82A denotes the AP2V mutant substrate in complex with the V82A mutant protease.

**Substrate peptides.** Two decameric peptides, WTNCP-p1 (P5 to P6: ERQAN* FLGKI) and AP2VNC-p1 (P4 to P6: RQVN*FLGKIN), with sequences corresponding to the NC-p1 substrate cleavage site within the Gag polyprotein, were chosen as previously described (40). The slight difference in the regions of the cleavage sites that were crystallized occurred because the AP2VNC-p1 peptide (P5 to P6: ERQVN*FLGKI) was not very soluble, whereas the decameric peptide (P4 to P6: RQVN*FLGKIN) was. Both peptides were purchased from Quality Controlled Biochemicals Inc., Hopkinton, Mass.

**Mutagenesis, protein purification, and crystallization.** The protease was expressed from a synthetic gene optimized for Escherichia coli cox4B1 using OQ7K to prevent autoproteolysis (45). This gene was used as the starting template for mutagenesis to introduce the D25N (to prevent the protease from cleaving its substrates) and V82A substitutions. Mutations were made using a QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, Calif.).

Protein expression, isolation, and purification were carried out as previously described (38, 39, 49). Stock solutions (25 mM) of substrate peptides used for crystallization were dissolved in refolding buffer rather than in dimethyl sulfoxide. The protein concentration was approximately 1 mg/ml. Small crystals started appearing after 6 to 8 weeks, with the longest length between 0.1 and 0.2 mm.

**Data collection.** Crystals chosen for data collection were flash frozen in a nitrogen stream. X-ray data on these cryocooled crystals were collected on our in-house Rigaku X-ray generator with an R-axis IV image plate system. Approximately 200 5-min frames were collected per crystal with 1-degree oscillations and no overlap between frames. The frames were later integrated and scaled by using the programs DENZO and ScalePack (29, 34). The WTNCP-p1WT and AP2VNC-p1V82A complexes diffracted up to 2.1 and 2.0 Å, respectively. Thus, the presence or absence of P5 or P6 in this crystallization condition was variable. The presence of P5 or P6 in this crystallization condition was variable.

### **TABLE 1. Crystallographic statistics for the two NC-p1 complexes**

<table>
<thead>
<tr>
<th>Variable</th>
<th>WTNCP-p1WT</th>
<th>AP2VNC-p1V82A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td>ERQAN*FLGKI</td>
<td>RQVN*FLGKIN</td>
</tr>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2_12_1</td>
<td>P2_12_1</td>
</tr>
<tr>
<td>a (Å)</td>
<td>51.1</td>
<td>50.8</td>
</tr>
<tr>
<td>b (Å)</td>
<td>57.7</td>
<td>57.3</td>
</tr>
<tr>
<td>c (Å)</td>
<td>61.5</td>
<td>60.9</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>−80</td>
<td>−80</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Total number of reflections</td>
<td>38149</td>
<td>80252</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>10490</td>
<td>12492</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>9.2</td>
<td>7.5</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.1</td>
<td>99.7</td>
</tr>
<tr>
<td>Rint (%)</td>
<td>4.5</td>
<td>9</td>
</tr>
<tr>
<td>Crystallographic refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rvalue (%)</td>
<td>20.6</td>
<td>19.5</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>23.3</td>
<td>23.1</td>
</tr>
<tr>
<td>Sigma cutoff</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RMSD* in:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.5</td>
<td>1.3</td>
</tr>
</tbody>
</table>

* RMSD, root mean square deviation.

### **RESULTS**

**Overall structure of the complexes.** Two decameric substrate peptides corresponding to the Gag polyprotein cleavage sites of NC-p1 (ERQAN*FLGKI) and its AP2V mutant (RQVN*FLGKIN) were crystallized in complex with inactive (D25N) variants of WT and drug-resistant HIV-1 protease having the V82A mutation (referred to as WTNCP-p1WT and AP2VNC-p1V82A complexes, respectively). Although the peptides used for crystallization span a slightly different region of the cleavage site (P5 to P6), these differences are likely to have minimal or no effect on the bound conformation of the substrate. From our previous experience with determining substrate complexes with HIV protease (38, 39, 49), the P5 and P6 residues usually have no contact with the protease and are often disordered in the electron density. Thus, the presence or absence of P5 or P6 in the crystallization condition is variable. The presence of P5 or P6 in this crystallization condition is variable. The presence of P5 or P6 in this crystallization condition is variable.

**Overall structure of the complexes.** Two decameric substrate peptides corresponding to the Gag polyprotein cleavage sites of NC-p1 (ERQAN*FLGKI) and its AP2V mutant (RQVN*FLGKIN) were crystallized in complex with inactive (D25N) variants of WT and drug-resistant HIV-1 protease having the V82A mutation (referred to as WTNCP-p1WT and AP2VNC-p1V82A complexes, respectively). Although the peptides used for crystallization span a slightly different region of the cleavage site (P5 to P6), these differences are likely to have minimal or no effect on the bound conformation of the substrate. From our previous experience with determining substrate complexes with HIV protease (38, 39, 49), the P5 and P6 residues usually have no contact with the protease and are often disordered in the electron density. Thus, the presence or absence of P5 or P6 in the crystallization condition is variable. The presence of P5 or P6 in this crystallization condition is variable. The presence of P5 or P6 in this crystallization condition is variable. The presence of P5 or P6 in this crystallization condition is variable.

As observed in other HIV-1 protease-substrate complexes (40, 51), the substrates exist in an extended β-strand conformation (Fig. 1). However, a difference between the two complexes is observed in the φ and ψ angles for the P1 and P1' residues that flank the scissile peptide bond (Fig. 1). The φ and ψ angles for the WTNCP-p1WT complex are −60° and 34°, re-
respectively, for P1 and $-50^\circ$ and $152^\circ$, respectively, for P1'. whereas for the $\text{AP}^{2V}\text{NC-}p1_{\text{V82A}}$ complex, they are $-72^\circ$ and $87^\circ$, respectively, for P1 and $-102^\circ$ and $143^\circ$, respectively, for P1'. These angles are still in the “generally allowed β-strand region” of the Ramachandran plot (41). Significant shifts are seen in the $\psi$ angle of P1 ($53^\circ$) and the $\phi$ angle of P1' ($52^\circ$), resulting in a large change in the orientation of the P1 carbonyl oxygen (Fig. 1). Double difference plots were calculated (39) in order to elucidate regions where changes have occurred between the

![FIG. 1. Conformation of the NC-p1 substrates. Stereo diagram of the superposition of the WT (cyan) and AP2V (magenta) variants of NC-p1 peptides. An arrow indicates the difference in conformation of the scissile bonds between the two structures. The superposition was based on structurally conserved parts of the respective protease structures. Disordered side chains, whose conformation could not be resolved in the electron density, have their names listed in parentheses.

FIG. 2. Double difference plots. Relative shifts in the V82A complexes in reference to the corresponding WT complexes with NC-p1 (a), CA-p2 (b), and p1-p6 (c). Double difference plots contour differences in internal $\text{C}^\alpha$-$\text{C}^\alpha$ distances between two complexes. Each contour line represents a deviation by 0.25 Å, Black, green, blue, and red distinguish the contour ranges $-1.0$ Å and below, $-1.0$ to $-0.5$ Å, $0.5$ to $1.0$ Å, and $1$ Å and above, respectively. In panel a, box A corresponds to conformational changes between the P1 loops (Gly78-Asn83) from each of the two monomers, and boxes B1 and B2 correspond to the conformational changes between the P1 loop with respect to the flap (Ile50) within each of the two monomers. All of the boxed regions emphasize internal regions of the protease whose relative conformation has changed by more than 1.0 Å between the two complexes.](image-url)
α-carbon backbones of the WTNC-p1WT and AP2VNC-p1V82A complexes (Fig. 2a). Double difference plots allow for comparison of two complexes without superimposing them and, thus, are unbiased by which residues are chosen for the superposition. To compute these plots, the distances between all the Cα atoms within each of the complexes are calculated, and the difference in these Cα-Cα distances between the two complexes is mapped as a contour plot. The contour plot indicates that the P1 loops (Gly78-Asn83) in the two monomers are more than 1Å farther apart in the WTNC-p1WT complex than in the AP2VNC-p1V82A complex (Fig. 2a, box A). Similarly, the distances between the P1 loops and the flaps within the monomers are larger in the WTNC-p1WT complex (Fig. 2a, boxes B1 and B2). These two observations suggest that the P1 loops in the AP2VNC-p1V82A complex have moved closer to each other in response to amino acid modifications at position 82 and at P2 in the substrate (Fig. 2a). Calculation of the shape complementarity (21) between the two substrates and their respective proteases shows that the complementarity is much better for the AP2VNC-p1V82A complex (0.71) than for the WTNC-p1WT complex (0.58). Thus, the AP2V substrate appears to fit more tightly within the V82A HIV-1 protease than the WT substrate within its protease.

We have previously determined the crystal structures of the WT and V82A protease variants with three other Gag polyprotein substrates: MA-CA, CA-p2, and p1-p6 (39, 40). Based on
their crystallographic statistics, the latter two were chosen for double difference distance analysis. Only minor changes were previously observed between these WT and V82A variant structures (Fig. 2b and c and Fig. 3b and c) (39). The previous study, which also analyzed WT and V82A variant protease complexes (40) (39) (Table 2). Only two substrate-protease complexes (40) (39) (Table 2). Only two substrate-protease complexes (40) and three substrate-protease V82A complexes (39).

### Hydrogen bonds.

All of the substrate complexes of HIV-1 protease show an extensive network of 12 to 18 hydrogen bonds in that structure relative to the WTNC-p1WT complex.

### van der Waals interactions.

Both V82A and AP2V mutations involve changes in the size of their hydrophobic side chains. A detailed examination of the substrate residues P2 to P2' (GlnP3 is disordered in both complexes and GlyP3 has no side chain) reveals several rearrangements in packing between the two complexes. At residue 82, the site of the protease mutation, packing between PheP1' and residue 82 is lost with the V82A mutation (Fig. 5a and Table 4). For each substrate residue, the closest van der Waals contact made by each atom with the protease atoms was measured, and an average contact distance was calculated for the residue. PheP1', which makes four additional van der Waals contacts to Gly27, Leu23, Pro81, and Val82 in the WTNC-p1WT complex, is also the case in the NC-p1 complexes, where only one substrate side chain is involved in hydrogen bonds. AsnP1 ND2 makes a weak hydrogen bond (3.5 Å) with Gly27 O in the AP2VNC-p1V82A complex. This is the only substrate complex structure where a hydrogen bond is observed between the protease and the side chain of P1. This hydrogen bond, however, is absent in the WTNC-p1WT structure. Although the side chains of ArgP4, GlnP3, and LysP4' could possibly be making hydrogen bonds, they are disordered in both complexes and so cannot be observed. Hence, apart from the lone and weak AsnP1 ND2-Gly27 O hydrogen bond, there appears to be no stable side chain hydrogen bond in either complex.

In addition to direct substrate-protease hydrogen bonds, water-mediated hydrogen bonds were also examined, as they can be relevant in rational drug design (22). Out of the five water sites conserved among HIV-1 protease complexes (23, 25, 40), only four are observed in the WTNC-p1WT complex (Fig. 4). However, all five of the water molecules are found in the AP2VNC-p1V82A complex. The water molecule that is absent in the WTNC-p1WT complex (Fig. 4, W3) directly mediates substrate backbone (GlnP3 O) with the protease (Gly27 O, AspP2 OD1, and ArgP4 NE). Absence of this water bridge, in the WTNC-p1WT complex may be a direct consequence of the poor substrate-protease hydrogen bonding.

In addition to the five conserved water molecules, three other water molecules stabilize the AP2VNC-p1V82A complex. The side chain of AsnP1 forms a hydrogen bond with W2 (Fig. 4 and Table 3), which itself forms a hydrogen bond with the conserved W3. W7 mediates interactions between the backbone atoms of Gly48' (N) and LysP4' (O) in the AP2VNC-p1V82A complex. The third water site, W8, bridges the side chain oxygens of AspP2 with IleP3'N. Thus, these three additional water molecules appear to further stabilize the AP2VNC-p1V82A complex whereas the absence of the W3 water site may be indicative of the higher level of disorder in the WTNC-p1WT complex.

### TABLE 2. Substrate-protease hydrogen bonds formed by the two NC-p1 substrates

<table>
<thead>
<tr>
<th>Substrate atom</th>
<th>Protein atom</th>
<th>Length (Å) of hydrogen bond</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlaP4 O</td>
<td>Gly48 N</td>
<td>3.2</td>
</tr>
<tr>
<td>AlaP4 N</td>
<td>Asp30 OD2</td>
<td>3.5</td>
</tr>
<tr>
<td>AlaP3 N</td>
<td>Asp92 OD2</td>
<td>3.6</td>
</tr>
<tr>
<td>AlaP3 O</td>
<td>Asp92 N</td>
<td>3.2</td>
</tr>
<tr>
<td>AlaValP2 N</td>
<td>Gly48 O</td>
<td>3.2</td>
</tr>
<tr>
<td>AsnP1 N</td>
<td>Gly27 O</td>
<td>3.2</td>
</tr>
<tr>
<td>AsnP1 ND2</td>
<td>Gly27 O</td>
<td>3.5</td>
</tr>
<tr>
<td>AsnP1 O</td>
<td>Asn25 ND2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

* a +, hydrogen bond is longer in the WTNC-p1WT complex than in the AP2VNC-p1V82A complex; –, hydrogen bond is shorter in the WTNC-p1WT complex than in the AP2VNC-p1V82A complex. Note that if no distance is listed for a particular interaction, it is because that distance is well beyond hydrogen bonding distance.

* b Average length of corresponding hydrogen bonds found in six substrate-protease complexes (40) (39) (Table 2). Only two substrate-protease complexes (40) and three substrate-protease V82A complexes (39).
and make 14, 7, and 10 additional van der Waals contacts, respectively. These additional contacts occur because of a change in the conformation of AsnP1 (Fig. 1) that enables it to fit within the pocket formed by Gly27, Pro81 and Asn25 (Fig. 5c). LeuP2 also changes conformation (Fig. 1) and forms closer van der Waals contacts with Gly27, Ala28, and Asp30 (Fig. 5c). The two added methyl groups in ValP2 allow it to make new contacts with Val32 CG2 and Ile50 CD. (However, in the WTNC-p1WT complex, the side chain of Ile50 is disordered in the electron density map.) However, no change is seen in the conformation of residues Val32, Ile47, and Ile84, which line the corresponding P2 pocket (Fig. 3a). Thus, except for the PheP1 Val82 interaction, the overall van der Waals packing in the AP2VNC-p1V82A complex is better than in the WTNC-p1WT complex. The loss of contact from the V82A mutation is compensated by the gain of contacts throughout the AP2VNC-p1V82A complex due to the AP2V substitution, implicating a subtle balance of interdependency between the different residues within the substrate. These results support the previous observation that the P2 substrate site for HIV-1 protease is preferentially a \(-\)branched amino acid (14) and, in particular, a valine (7).

**DISCUSSION**

The NC-p1 cleavage site is the rate-determining step in the processing of the Gag polyprotein (7, 37, 54). The amino acid sequence of this site, with an Asn at P1 and an Ala at P2, is also the least homologous compared with the other HIV-1 protease substrates sites, which have hydrophobic residues at P1 and branched residues at P2. If either P1 or P2 is replaced with Phe, Leu, Met, and Tyr or with Val, respectively, the site is cleaved more efficiently (36). Peptide library screens find that the P2 site within NC-p1 is preferentially a Val (7). However, these substitutions are not seen in WT virus, implying that a more efficient and presumably premature cleavage of this site may be detrimental to viral maturation (19, 26, 36).

The NC-p1 cleavage site has been seen to coevolve in response to the V82A drug-resistant mutation in HIV-1 protease, with the Ala residue at P2 mutating to a Val (1, 10, 20, 55). Previous studies on the kinetics of substrate cleavage have shown that this mutation results in the substrate's being more efficiently cleaved (36). However, these substitutions are not seen in WT virus, implying that a more efficient and presumably premature cleavage of this site may be detrimental to viral maturation (19, 26, 36).

**TABLE 3. Water-bridging network involving substrates and protease**

<table>
<thead>
<tr>
<th>Water site</th>
<th>Substrate atom(s)</th>
<th>Water</th>
<th>Protein atom(s)</th>
<th>Substrate atom(s)</th>
<th>Water</th>
<th>Protein atom(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>Ala P2 O Phe P1' O</td>
<td>Wat25</td>
<td>Ile50 N Ile50' N</td>
<td>ValP2 O PheP1' O</td>
<td>Wat33</td>
<td>Ile50 N Ile50' N</td>
</tr>
<tr>
<td>W2</td>
<td>Disordered</td>
<td></td>
<td>AsnP1 ND2</td>
<td>GlnP3 O^a</td>
<td>Wat27</td>
<td>Gly27 O Asp29 OD1</td>
</tr>
<tr>
<td>W3</td>
<td>Disordered</td>
<td></td>
<td></td>
<td>Wat93</td>
<td>Arg8' NE Wat27</td>
<td></td>
</tr>
<tr>
<td>W4</td>
<td>Wat6</td>
<td>Thr26 O Asp29 OD1 Arg87 NE</td>
<td>Wat4</td>
<td>Thr26 O Asp29 OD1 Arg87 NE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W5</td>
<td>LeuP2' O Wat9</td>
<td>Gly27' O Asp29' OD1 Wat1</td>
<td>LeuP2' O Wat55</td>
<td>Gly27' O Asp29' OD1 Arg87' NE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W6</td>
<td>Wat1</td>
<td>Thr26' O Arg87' NE Wat9</td>
<td>Wat9</td>
<td>Thr26' O Asp29' OD1 Arg87' NE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W7</td>
<td>LysP4' O^a</td>
<td>Wat112</td>
<td>Gly48' N</td>
<td>Wat50</td>
<td>Gly30' OD1 and OD2</td>
<td></td>
</tr>
<tr>
<td>W8</td>
<td>IleP5' N^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a These residues are currently modeled as alanines due to lack of electron density.

FIG. 4. Variations in the substrate-protease hydrogen bonds between the WTNC-p1WT and AP2VNC-p1V82A complexes. In the superimposed active site regions of the WTNC-p1WT (gray and cyan) and the AP2VNC-p1V82A (magenta and purple) complexes, only those side chains that form protease-substrate hydrogen bonds are explicitly shown (Asn25, 25'; Asp29, 29'; Asp30, 30'; and AsnP1). Water molecules are shown with van der Waals surfaces. The AP2VNC-p1V82A complex forms additional hydrogen bonds compared with the WTNC-p1WT complex. These are highlighted with dotted yellow lines (Tables 2 and 3).
cleaved 2.1 times more efficiently (ratios of $k_{cat}/K_m$ [mutant substrate/WT substrate]) for the V82A protease variant and 2.6 times more efficiently for WT HIV-1 protease (12). However, despite the enzymatic advantage, the AP2V mutation is not observed in sequences of the WT virus. In this study we have determined the crystal structures of two NC-p1 substrate complexes, the WTNC-p1WT complex and the AP2VNC-p1V82A complex, in order to understand the structural basis for this coevolution. Residue 82 is not in direct van der Waals contact with the P2 site of the substrate, so the coevolution is not as simple as a decrease in the size of one residue corresponding to an increase in the size of a neighboring residue. However, unlike what we have observed in other substrate complexes (CA-p2, MA-CA, and p1-p6), where the substrates make no or minimal contact with Val82 (39), in the WTNC-p1WT complex Val82 does make contact with both AsnP1 and PheP1'. Thus, in comparison with the other WT substrate complexes, residue 82 appears to be more important for recog-

TABLE 4. Total number of van der Waals contacts among substrate residues between P2-P2' and HIV-1 protease

<table>
<thead>
<tr>
<th>Substrate residue</th>
<th>WTNC-p1WT</th>
<th>AP2VNC-p1V82A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala/ValP2</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>AsnP1</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>PheP1'</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>LeuP2'</td>
<td>28</td>
<td>34</td>
</tr>
</tbody>
</table>

FIG. 5. Variations in the substrate-protease van der Waals interactions between the WTNC-p1WT and AP2VNC-p1V82A complexes. van der Waals surfaces of protease residue 82, viewed down the dimer twofold axis, and neighboring atoms of the corresponding NC-p1 substrate: (a) WTNC-p1WT complex (note that Val82 contacts PheP1') and (b) AP2VNC-p1V82A complex (note that no contact is made between Ala82 and the substrate). Substrates and proteases are distinguished in gray and black, respectively. A Co trace of the overlying flaps is also shown. (c) Active site residues in the WTNC-p1WT. The protease is shown in gray and the substrate in cyan. Protease residues within 4.2Å of the substrate are shown in ball-and-stick representation, and residue 82 is highlighted in yellow. van der Waals contacts that are lost compared with the AP2VNC-p1V82A complex are highlighted with dotted surfaces and the P2 site is circled. (d) Active site residues in the AP2VNC-p1V82A. The protease is shown in gray, and the substrate is shown in magenta; protease residues within 4.2Å of the substrate are shown in ball-and-stick representation, and residue 82 is highlighted in yellow. van der Waals contacts that are lost compared with the WTNC-p1WT complex are highlighted with dotted surfaces, and the P2 site is circled.
nition of the NC-p1 substrate cleavage site, and when the drug-resistant V82A mutation occurs, the interactions at P1 and P1′ are perturbed.

The question remains of why the loss of contact at P1 and P1′ would affect the residue at P2. Compared with the other substrate complexes where the P2 residue usually makes van der Waals contacts with protease residues in its surrounding pocket, in the WT NC-p1WT complex, the Ala at P2 is unable to do so. In addition, the hydrogen bonds in the WT NC-p1WT complex (Table 2) between the substrate and the protease are longer than those observed in other substrate complexes. The lack of interactions at the P2 site and the lengthening of the hydrogen bonds possibly indicate that NC-p1 binds in the active site of HIV-1 protease less optimally than in the other substrate cleavage sites.

In the AP2VNC-p1V82A complex, the PheP1′ and, to a lesser extent, the AsnP1 lose contact with residue 82. This loss of contact at residue 82 would probably cause the NC-p1 substrate to fit even more poorly within the active site of HIV-1 protease. However, the Val substitution at P2 fills the protease pocket, allowing van der Waals interactions to occur with the surrounding protease residues, thereby apparently stabilizing the conformation of the substrate within the active site. Additional evidence of stabilization is demonstrated in the hydrogen bonding pattern in the AP2VNC-p1V82A complex, which is more comparable, in terms of both the greater numbers of hydrogen bonds and shorter hydrogen bonding distances, to those seen in other substrate complexes (Table 2) than it is to the WT NC-p1WT complex. The ValP2, therefore, appears to do so. In addition, the hydrogen bonds in the WT NC-p1WT complex, the Ala at P2 is unable to do so. In addition, the hydrogen bonds in the WT NC-p1WT complex (Table 2) between the substrate and the protease are longer than those observed in other substrate complexes. The lack of interactions at the P2 site and the lengthening of the hydrogen bonds possibly indicate that NC-p1 binds in the active site of HIV-1 protease less optimally than in the other substrate cleavage sites.

ACKNOWLEDGMENTS

We thank Claire Baldwin for assisting with editorial work. We also acknowledge Balaji Bhryavbhatla and Luca Leon for technical support and Mohan Somasundaran and Jared Auclair for useful discussions. We thank Ronald Swanstrom for critical discussions.

This research was supported by the National Institutes of Health (grant R01-GM64347).

REFERENCES

