PROTEINS: Structure, Function, and Genetics 43:455–464 (2001)

Structural Implications of Drug-Resistant Mutants of HIV-1 Protease: High-Resolution Crystal Structures of the Mutant Protease/Substrate Analogue Complexes

Bhuvaneshwari Mahalingam,[1] John M. Louis,[3] Jason Hung,[3] Robert W. Harrison,[2] and Irene T. Weber[1]*
[1]*Department of Biology, Georgia State University, Atlanta, Georgia*
[2]*Department of Computer Science, Georgia State University, Atlanta, Georgia*
[3]*Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland*

ABSTRACT **Emergence of drug-resistant mutants of HIV-1 protease is an ongoing problem in the fight against AIDS. The mechanisms governing resistance are both complex and varied. We have determined crystal structures of HIV-1 protease mutants, D30N, K45I, N88D, and L90M complexed with peptide inhibitor analogues of CA–p2 and p2–NC cleavage sites in the Gag–pol precursor in order to study the structural mechanisms underlying resistance. The structures were determined at 1.55–1.9-Å resolution and compared with the wild-type structure. The conformational disorder seen for most of the hydrophobic side-chains around the inhibitor binding site indicates flexibility of binding. Eight water molecules are conserved in all 9 structures; their location suggests that they are important for catalysis as well as structural stability. Structural differences among the mutants were analyzed in relation to the observed changes in protease activity and stability. Mutant L90M shows steric contacts with the catalytic Asp25 that could destabilize the catalytic loop at the dimer interface, leading to its observed decreased dimer stability and activity. Mutant K45I reduces the mobility of the flap and the inhibitor and contributes to an enhancement in structural stability and activity. The side-chain variations at residue 30 relative to wild-type are the largest in D30N and the changes are consistent with the altered activity observed with peptide substrates. Polar interactions in D30N are maintained, in agreement with the observed urea sensitivity. The side-chains of D30N and N88D are linked through a water molecule suggesting correlated changes at the two sites, as seen with clinical inhibitors. Structural changes seen in N88D are small; however, water molecules that mediate interactions between Asn88 and Thr74/Thr31/Asp30 in other complexes are missing in N88D. Proteins 2001;43:455–464.**

Key words: aspartic protease; crystallography; inhibitors; conserved water; structural changes

INTRODUCTION

Human immunodeficiency virus (HIV) protease inhibitors have been successful in dramatically lowering the number of deaths due to acquired immunodeficiency syndrome (AIDS) during the past decade. They have been more potent than the reverse transcriptase inhibitors in suppressing HIV infection. However, the emergence of drug-resistant mutants has resulted in decreased effectiveness of protease inhibitor therapy. Clinical data show considerable cross-resistance among the five FDA-approved protease inhibitors used for therapy and preliminary data suggest that people infected with drug-resistant strains may not respond effectively to therapy.[1,2] It is therefore important to understand the mechanisms governing drug resistance in order to rationally design drug combinations that will maintain low viral levels.

Detailed structural information from the crystal structures of HIV protease inhibitor complexes has played a key role in structure-based design of more potent inhibitors.[3] Although a large number of crystal structures of HIV-1 protease complexed with nonpeptide inhibitors are available, there are relatively few reported structures with peptide analogues of the natural cleavage sites of the protease, or drug-resistant variants of these complexes. The mechanisms by which the mutants maintain sufficient enzymatic activity for viral replication in the presence of protease inhibitors are not fully understood. Understanding the structural mechanisms in the phenomenon of drug resistance in conjunction with clinical data will help in the design of better inhibitors of the protease. Toward this goal, we have undertaken an analysis of protease mutants that arise in response to protease inhibitor therapy. We have earlier reported kinetic data on 8 of these mutants, using peptide substrates and inhibitors representing the CA–p2, $p6^{pol}$–PR, and PR–RT cleavage sites in the Gag–pol polyprotein, and crystal structures of R8Q, K45I and L90M mutants with a CA–p2 analogue.[4] We present a detailed crystallographic analysis of D30N, K45I, N88D, and L90M mutant proteins in complexes with

Grant sponsor: U.S. Public Health Service; Grant number: AI41380; Grant number: GM62920; Grant sponsor: Intramural AIDS Targeted Antiviral Program of the Office of the Director of National Institutes of Health.

*Correspondence to: Irene T. Weber, Department of Biology, Georgia State University, Atlanta, GA 30303.
E-mail: iweber@gsu.edu

Received 31 August 2000; Accepted 2 February 2001

reduced peptide analogues of both the CA–p2 and p2–NC cleavage sites and the comparison with wild-type protease/p2–NC inhibitor complex.

HIV replication is highly error prone, and due to the high mutation rate, many strains of HIV have naturally occurring polymorphisms, which also contribute to drug resistance. However, polymorphisms are very rare in protease residues located at the active site or dimer interface, or both. Mutations D30N, N88D, and L90M have so far been seen primarily in response to drug selection pressure and N88D, L90M, and K45Q have been reported as polymorphisms only in very few (1%) of the isolates so far.[1] D30N and L90M are primary mutants, in that they cause resistance when present alone. N88D and K45I are mutants that contribute to resistance when present together with other mutations in the protease. D30N is mostly seen to arise in response to Nelfinavir and shows little cross-resistance to other protease inhibitors. L90M is a resistant mutant seen in Saquinavir and Nelfinavir treatment; it is cross-resistant to the other three protease inhibitors. N88D is resistant to Nelfinavir in combination with D30N and is cross-resistant to Indinavir, but it is also the only mutant reported so far to hypersensitize isolates to Amprenavir.[1] K45I has not been reported to be resistant to the currently used drugs, but arises in combination with mutations at L10 and I84 on exposure to XM323.[5]

The protease-mediated sequential processing of Gag–pol is not fully understood. However, the cleavage at the N-terminus of the protease is essential for the liberation of mature like catalytic activity and optimal ordered processing of Gag.[6,7] Initial cleavage in Gag occurs at the p2–NC site and is about 400-fold faster than the final slow cleavage at the CA–p2 site, which is thought to play a regulatory role.[8] The N-terminal peptide of p2 Gag has been shown to inhibit Gag processing in vitro[9] and the p2 peptide is important in assembly.[10] Our kinetic studies on inhibition of the protease mutants with the CA–p2 and p2–NC substrate analogue inhibitors showed the latter to be a weaker inhibitor in vitro.[4] While mutants D30N, K45I, and N88D are poorly inhibited by both analogues as compared with the wild-type, inhibition of L90M is similar to the wild-type, however, it is more sensitive to urea denaturation than other mutants. Sedimentation equilibrium studies have also shown that L90M has reduced dimer stability at pH 7.0, relative to the wild-type protease.[11]

MATERIALS AND METHODS

Expression and Purification of the HIV-1 Protease Mutants

The HIV-1 protease coding region containing the substitution mutations Q7K, L33I, L63I, C67A, C95A, designated wild-type, was expressed using pET11a vector and *Escherichia coli* BL21 (DE3) (Novagen, WI). These mutations restrict autoproteolysis as well as cysteine thiol oxidation; however, the kinetic parameters of this mutant are indistinguishable from those of the native enzyme.[4,7] Additional specific mutations corresponding to the sites of drug resistance in the protease domain were carried out using the Quick-Change mutagenesis protocol (Stratagene, CA). The nucleotide sequence of the cloned DNA was confirmed by sequencing. The molecular weight of the mutant protein was confirmed by mass spectroscopy.

Cells were grown at 37°C in Luria–Bertani medium and induced for expression according to described procedures.[12,13] Cells were suspended in 20 vol of buffer A [50 mM Tris-HCl, pH 8.2, 10 mM ethylenediaminetetraacetic acid (EDTA)] and lysed by sonication in the presence of 100 μg/ml lysozyme at 4°C. The insoluble recombinant protein (inclusion bodies) was washed by resuspension in buffer B (buffer A containing 2 M urea and 1% Triton X-100) and then in buffer A. In both cases, the inclusion bodies were pelleted by centrifugation at 20,000*g* for 30 min at 4°C. The final pellet was solubilized in 50 mM Tris-HCl, pH 8.0, 7.5 M guanidine-HCl, 5 mM EDTA and applied to a Superdex-75 column (HiLoad 2.6 cm × 60 cm, Amersham Pharmacia Biotech, NJ) equilibrated in 50 mM Tris-HCl, pH 8, 4 M guanidine-HCl, 5 mM EDTA and at a flow rate of 3 ml/min at ambient temperature. Peak fractions were pooled and subjected to reverse-phase high-performance liquid chromatography (HPLC) on a Poros 20 R2 column (Perceptive Biosystems, Framingham, MA). The protein was folded by dialysis into 0.05 M formic acid at pH 2.8, followed by 0.05 M sodium acetate buffer, pH 5.0. The protein was concentrated to the desired concentration using Centriprep10, followed by Centricon10 concentrators (Millipore, Bedford, MA) and stored at −20°C.

Determination of Crystal Structures

Protease crystals were grown at room temperature by vapor diffusion using the hanging drop method. The reservoir contained 0.25 citrate/0.5 M phosphate buffer, pH 5–6.5, 10% DMSO, 10 mM DTT and 20–50% saturated ammonium sulfate as precipitant. The protein (2–5 mg/ml) was preincubated with the inhibitor in a molar ratio of 1:10. Two peptidic inhibitors were used: the CA–p2 inhibitor (Arg—Val—Leu—*r*—Phe—Glu—Ala—Nle), where *r* represents the reduced peptide group, and the p2–NC inhibitor (Acetyl Thr—Ile—Nle—*r*—Nle—Gln—Arg). The crystallization drops had a 1:1 ratio by volume of reservoir solution and protein. The final drop was usually 2-μl, and crystals grew in 2–7 days.

X-ray diffraction data for L90M with CA–p2 inhibitor were collected on an RAXIS-II image plate system, using a RIGAKU rotating anode generator at room temperature. All other diffraction data were collected at the NSLS at Brookhaven. Crystals were frozen at 90K in a liquid nitrogen stream. Data were processed using the HKL suite.[14] The structures were solved by molecular replacement using AMoRe.[15] The new structures reported here used the K45I/CA–p2 protease coordinates[4] as the starting model. X-PLOR[16] was used for refinement and FRODO[17] was used for model building. The coordinates for all the structures have been deposited at RCSB and have entry codes 1daz, 1dw6, 1fej, 1ff0, 1ffi, 1fff, 1fg6, 1fg8, and 1fgc.

RESULTS AND DISCUSSION

We have determined crystal structures of HIV-1 protease mutants D30N, K45I, N88D and L90M, in complex

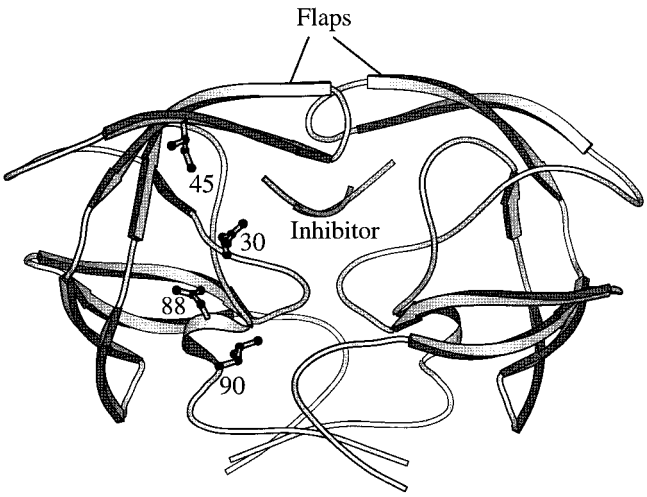

Fig. 1. Location of the mutated residues (30, 45, 88, 90) is shown in one subunit of the protease dimer. Residues 1–99 and 101–199 correspond to the two subunits in the dimer. The protease and inhibitor are in ribbon representation.

with reduced peptide analogues of the CA–p2 and p2–NC processing sites in the Gag precursor, in order to investigate any structural differences relative to the wild-type protease. The location of the mutations in the protease dimer is shown in Figure 1. All complexes crystallized in the space group $P2_12_12_1$ with the exception of D30N and N88D bound to the p2–NC inhibitor, which are in $P2_12_12$. All the crystals diffracted to ≥1.9-Å resolution, with K45I/CA–p2 data extending to 1.55 Å, the highest resolution reported so far for HIV protease crystals. All the structures were refined with one dimer in the crystallographic asymmetric unit. Table I shows the crystallographic details for these structures.

Overall Comparison of the Crystal Structures

The mutants were superimposed onto the wild-type/p2–NC complex, using ALIGN.[18] The root-mean-square deviation (RMSD) for the Cα atoms in the protease dimer is within 0.2–0.5 Å for all the structures. The largest differences in the main-chain with respect to wild-type are in the external loop of residues 38–41. The residues 22–33, which include the catalytic Asp—Thr—Gly at 25–27, have the least variability in the main-chain. The main-chain atoms of the hairpin loop (residues 49–52) that connects the two β-strands in the flap have the least variation among the flap residues in most of the complexes. Interestingly, residues 22–33 and 48–53 have also been shown to be sensitive to nonconservative substitutions.[19] Residues within the protein core have the same conformation for side-chains in all the superposed structures, with a few exceptions like Leu97 and Ile64, which have variable conformation for their side-chains. Differences at all other internal residues are either due to mutations and/or inhibitor binding. The larger variation for surface residues can be attributed to intermolecular crystal packing effects, interaction with solvent or inherent flexibility.

Features in the Electron-Density Map Around the Active Site

All the structures show disordered density for several hydrophobic residues in the inhibitor binding pocket, indicating the flexibility of HIV protease for binding different substrates or inhibitors (Fig. 2). The side-chains of Ile50/50′, Leu23/23′, and Val82/82′ have poorly defined density in all the crystal structures. Possible flipping of the carbonyl oxygen of 50/50′ is suggested by ambiguous density in some of the structures. Ile84/84′, which is a completely buried residue, has poor side-chain density in most of the CA–p2 complexes, while it is ordered in most of the p2–NC complexes.

In the K45I/CA–p2 complex, which is the highest resolution structure, the presence of more than one conformation is apparent beyond Cγ for Leu23. Cγ1 of Val32, which is in hydrophobic contact with the inhibitor, has weaker density compared with Cγ2. In addition, two residues that are not part of the binding site, Gln92 and Leu97, show two alternate conformations for atoms beyond Cγ. In one orientation of Gln92, Nε2 is within hydrogen bonding distance of carbonyl oxygen of Asn88, and in the other orientation it is close to P3Arg of a symmetry related inhibitor molecule. These side-chains could therefore participate indirectly in catalysis.

All the structures were refined with one inhibitor molecule bound to the protease dimer, except for K45I and N88D with CA–p2, which have two inhibitor molecules per dimer with about 40–60% relative occupancy. The two disordered inhibitors are bound in opposite orientations as described previously.[20] Side-chains in the p2–NC inhibitor are found to be more ordered compared with the CA–p2 analogue, where traces representing the second orientation at extremely low occupancies are visible at some side-chains. The distal side-chains of the inhibitor, P3Arg and P4′Nle in CA–p2 and P3′Arg in p2–NC, are usually disordered, except in the N88D/p2–NC complex, where the P3′Arg is ordered, and in K45I/CA–p2 complex, where the P4′Nle is ordered.

Structural Features of the Binding of Inhibitors

The inhibitor molecules in all complexes are bound to the protease in the same orientation, except for the N88D/CA–p2 complex. The set of main-chain hydrogen bonds through which the inhibitor binds to the enzyme is preserved in all the structures, as is the water at the active site (Water #7 in Fig 4). This water is connected to the amide N of 50/50′, the amide N and carbonyl O of P1′ residue and the carbonyl O of the P2 residue. The distances to this water in all structures are given in Table II. A weak hydrogen bond between the amide N of P1′residue and Asp25 Oδ2, with an average distance of 3.6 Å in CA–p2 and 3.2 Å in p2–NC complexes is also seen in all structures. Both inhibitors have a reduced peptide bond between their P1 and P1′ residues and hence are not cleaved by the protease. The superposed inhibitors are shown in Figure 3. The RMSD of the main-chain of P2–P2′ residues with respect to the wild-type/p2–NC complex is within the range of 0.1–0.3 Å for all the structures. The inhibitors

TABLE I. Crystallographic Details for the Complexes

Analogue	PR	Space group	Unit cell dimensions (Å) a, b, c	R_{merg} (%)	Resolution range for refinement (Å)	R_{work} (%)	R_{free} (%)	No. of waters	Completeness (%)
CA-p2	D30N	P_{212121}	50.9, 58.2, 61.2	7.5	8–1.9	21.2	24.8	105	97.8
	K45I	P_{212121}	51.2, 58.2, 61.2	5.4	8–1.55	21.3	25.6	142	99.6
	N88D	P_{212121}	51.2, 58.3, 61.0	6.4	8–1.85	22.3	26.7	81	98.6
	L90M	P_{212121}	52.0, 59.6, 62.0	6.4	8–1.88	21.8	24.9	36	94.8
p2-NC	WT	P_{212121}	50.7, 57.4, 61.1	7.5	8–1.9	21.2	26.8	103	98.3
	D30N	P_{21212}	57.8, 85.6, 46.3	5.9	8–1.7	21.1	25.2	117	98.4
	K45I	P_{212121}	51.2, 57.8, 61.5	6.1	8–1.85	20.1	23.2	122	98.6
	N88D	P_{21212}	57.3, 85.4, 46.2	6.8	8–1.8	21.9	24.8	111	94.5
	L90M	P_{212121}	51.2, 58.1, 61.1	6.7	8–1.7	21.5	28.5	148	85.0

PR, protease; WT, wild-type protease.

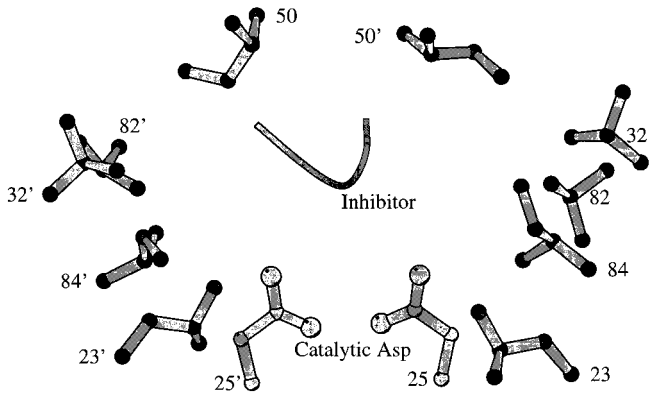

Fig. 2. Hydrophobic residues around the inhibitor with poorly defined electron density for the side-chains are shown in black. The catalytic aspartates are shown in gray.

make hydrophobic contact with the protein mainly through the P2–P2′ residues, as shown in Table III. There are small differences among the mutants in their interactions with the protein residues. The combined effect of these small changes may be responsible for some of the differences in the properties of the mutants.

The CA–p2 Analogue

The CA–p2 analogue (Arg—Val—Leu—*r*—Phe—Glu—Ala—Nle), binds to the protease through P3—P4′ residues. P3Arg is disordered in all the structures. P2Val is surrounded by hydrophobic residues of the protein, whose side-chains have similar conformation in all the superposed structures. P2Val is shielded from the solvent by Asp30, which is the only polar residue in contact with it. The pocket surrounding P1Leu/P1′Phe has a more hydrophilic character. Val82, which interacts with P1Leu/P1′Phe, has a variable conformation in all structures. This indicates that Val82 might modulate specificity at P1/P1′ through small conformational changes. In K45I and N88D, the inhibitor is bound in two orientations. In these two mutants, residues Val82′, Ile84/84′, Pro81′, and Ile50 have side-chain conformations such that they avoid unfavorable contacts with the inhibitor residues in the second orientation. Hence, these conformations are different from the other mutants. P2′Glu is pointed away from the hydrophobic interior of the protein toward the solvent and makes polar interactions with the main-chain and side-chain of residue 30′. The side-chain of P4′Nle is ordered only in the K45I structure where it makes van der Waals contact with Asp30/30′, Ile47/47′, Ile45 and Gln58; this difference arises because Cϵ of P4′Nle occupies the position corresponding to NZ of Lys45 in the wild-type protease.

p2–NC Analogue

The p2–NC inhibitor (acetyl Thr—Ile—Nle—*r*—Nle—Gln—Arg) has residues extending from the acetyl group at P4–P3′. Oγ1 of P3Thr has a water-mediated interaction with Oδ of residue 30 in all the complexes. This water may help stabilize the protease–inhibitor complex. The main-chain at P3Thr is flipped with respect to the corresponding position of P3Arg in CA–p2 (Fig. 3). As a result, the amide N of P3Thr forms a hydrogen bond with the carbonyl O of Gly48 in the flap, unlike the amide N of P3Arg in CA–p2 complexes, which interacts with Oδ2 of Asp29. Asp29 Oδ atoms in the p2–NC complexes are 0.5–0.7 Å farther from the inhibitor compared with their counterparts in CA–p2 complexes, whereby they avoid unfavorable contact with the Cβ atom of P3Thr. These differences may contribute to the weaker inhibition observed for the p2–NC analogue. Most of the protease side-chains in contact with P2Ile have similar conformations in all structures. Only the side-chain of Ile50′ has variable conformation. Cϵ atoms of P1/P1′Nle face the protease interior, indicating a preference for a hydrophobic environment (Fig. 3). The side-chain conformations of Val82/82′, which contact P1/P1′Nle are similar in all p2–NC complexes in contrast to the CA–p2 complexes. However, P2′Gln has protease interactions similar to those of P2′Glu in the CA–p2 analogue. P3′Arg is ordered only in N88D, where it makes contact with Arg8 and Glu21 through water molecules and is in hydrophobic contact with Val82.

Water-Mediated Interactions

Water molecules within the protein core play an important role in intra- and intersubunit interactions in HIV-1

TABLE II. Distances to the Conserved Water Molecule at the Active Site*

		Protease Residues		Inhibitor residues		
Inhibitor	Protease	I50 N	I50′ N	P2 O	P1′ O	P1′ N
CA-p2	D30N	2.8	2.8	2.8	2.7	3.4
	K45I	2.9	2.8	2.9/2.9	2.7/2.7	3.4/3.2
	N88D	3.0	2.8	2.9/2.8	2.6/2.6	3.2/3.2
	L90M	2.9	2.8	2.8	2.7	3.2
p2-NC	WT	2.8	2.9	2.8	2.6	—
	D30N	3.0	3.0	2.7	2.6	3.3
	K45I	2.9	2.9	2.8	2.7	—
	N88D	3.0	2.9	2.8	2.6	3.5
	L90M	3.1	2.9	2.7	2.6	3.6

WT, wild-type proteases.
*Distances in Ångstroms from the water at the active site, for amide nitrogens of Ile50/50′ of the protease, carbonyl oxygens of P2 and P1′, and amide nitrogens of P1′ residues of the inhibitors. CA–p2 complexes of K45I and N88D have two disordered inhibitor molecules.

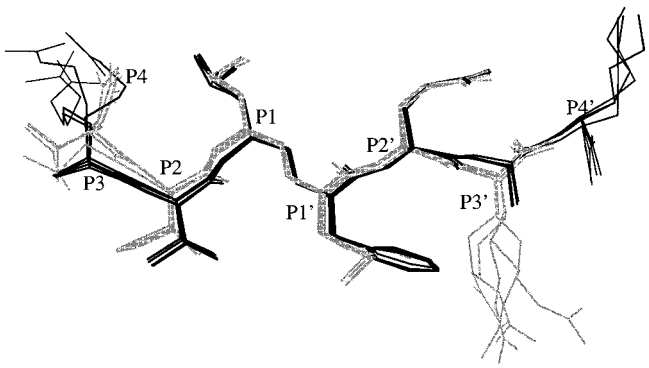

Fig. 3. Superposition of inhibitors. The CA–p2 inhibitors are in black and the p2–NC inhibitors are in gray. All mutants were superposed onto the wild-type/p2–NC complex.

protease (Fig. 4). Several intrasubunit water-mediated interactions are conserved in both monomers of the protease. These are indicated in one of the subunits by #1–6 in Figure 4. Water #1 links Oδ1 of Asp29, carbonyl O of Thr26, and Nε of Arg87, which are all conserved amino acids in the protease. The interactions of water #1 can stabilize the catalytic loop. Water #2 links carbonyl O of Ile63 to amide N of Gly16. Water #3 links the carbonyl oxygens of Lys7, Arg8, and Thr4, and imine N of Pro9, where Pro9 is conserved in all retroviral proteases. Water #4 links the amide N of Lys7, Oγ1/carbonyl O of Thr4, and amide N of Trp6. Interestingly, waters #2–4 are close to residues 7 and 63, which are sites of autoproteolysis in the native enzyme. Waters #5 and #6 link Nε2 of Asn88 to carbonyl oxygens/amide nitrogens of Thr31 and Thr74. These waters are absent, however, in the N88D mutant, as discussed later. Water molecules also contribute to intersubunit hydrogen-bonded contacts, where on average 6–7 water molecules are involved in bridging the two monomers. Conserved intersubunit waters observed in all our complexes include #7–10 (Fig. 4). Water #7 is conserved at the active site of most protease-inhibitor structures; it links amides of Ile50, Ile50′ with carbonyl oxygens of P2 and P1′ residues in the inhibitor. Water #8 links the carbonyl O of Gly27, Oδ1 of Asp29, and Nε of Arg8′, and also forms a hydrogen bond with the carbonyl oxygen of P2′ residue in the inhibitor, which could stabilize the protease–inhibitor complex. Water #9 links the carbonyl oxygen of Ala95 to amide N of Leu5′, which could stabilize the intersubunit β-sheet formed by the N- and C-terminal residues. Waters #8 and #9 are conserved in both subunits of the dimer. In contrast, water #10, which links the carbonyl oxygens of Asn98, Gly94′, and Oγ1 of Thr96′, is conserved only in one subunit. This water may be involved in the cleavage of the protease at the C terminus. Frequently, water molecules are associated with glycine residues, where they could stabilize the local secondary structure. Because many of the conserved waters are present in the vicinity of the autoproteolytic cleavage sites, they could play a role in the eventual degradation of the enzyme, in particular the cleavage that occurs between residues Leu5 and Trp6 in vivo.[7,21]

Dimer Interface

Dimer formation and stability is critical for catalysis in HIV-1 protease, since the two aspartates involved in the reaction are contributed from each of the monomers of the homodimer. Studies indicate that poor catalysis in some mutants is associated with reduced dimer stability.[4,11] About 70% of the hydrogen bond interactions at the dimer interface are contributed by residues 5, 26, 50, 87, 91, and 96–99, which comprise only 9% of the protein residues. Phe99/99′ also forms the maximum number of hydrophobic contacts at the dimer interface followed by Ile50/50′ and Gly51/51′. The HIV protease–inhibitor complex forms an extremely tight dimer interaction compared with other protein dimers.[22] This is apparent from the lower mean gap–volume index (GVI) (a measure of complementarity of two interacting surfaces) of 0.76 versus a mean of 2.2 in other homodimers. All our structures have similar GVI values of 0.74–0.77, except for the L90M/CA–p2 complex, which has a value of

TABLE III. Protease Residues That Have Hydrophobic Interactions With the Inhibitors*

Inhibitor	Inhibitor residues	Protease				
		WT	D30N	K45I	N88D	L90M
CA–p2	P3Arg					
	P2Val		28, 32,[a] 84[a]	28, 32,[a] 47,[a] 84,[a] 28′,[b] 32′,[a,b] 50′,[a,b] 84′,[a,b]	28, 32,[a] 50′,[a] 47,[a] 84,[a] 28′,[b] 32′,[a,b] 50,[a,b] 84′[a,b]	28, 32,[a] 84[a]
	P1Leu		23′, 81′, 82′,[a] 84′[a]	23′, 82′,[a] 84′,[a] 23,[b] 84[a,b]	23′, 82′,[a] 84′,[a] 50′,[a,b] 82,[a,b] 84[a,b]	23′, 81′, 82′[a]
	P1′Phe		23, 84[a]	23, 49′, 81, 82,[a] 84,[a] 23′,[b] 49,[b] 81′,[b] 82′[a,b]	23, 81, 82,[a] 84,[a] 23′,[b] 49,[b] 81′,[b] 82′[a,b]	23, 81
	P2′Glu		28′,[a] 30′,[a] 32′,[a] 47′,[a] 50,[a] 84′[a]	28, 30,[a] 28′,[b] 30′,[a,b]	28′, 30′,[a] 28,[b] 30,[b] 32[a,b]	28′, 30′,[a] 47′,[a] 50[a]
	P3′ Ala					
	P4′Nle		47′[a]	30,[a] 45, 47,[a] 30′,[a,b] 47′[a,b]		47′[a]
p2–NC	P4Acetyl	48,[a] 49, 81′		48,[a] 49	81′	48,[a] 49
	P3Thr	29	29	29		29
	P2Ile	28, 30,[a] 47,[a] 84[a]	30,[a] 47,[a] 84[a]	28, 47,[a] 84,[a] 50′[a]	28, 32,[a] 47,[a] 84[a]	28, 50,[a] 84[a]
	P1Nle	49, 50,[a] 81′, 84′[a]	23′, 49, 50,[a] 82′,[a] 84′[a]	50,[a] 81′, 82′,[a] 84′[a]	49, 81′, 82′,[a] 84′[a]	50,[a] 81′, 82,′[a] 84′,[a]
	P1′Nle	23, 50′,[a] 81, 84[a]	23, 49′, 81, 82,[a] 84[a]	23, 49′, 50′,[a] 81′, 84[a]	23, 49′, 81, 82,[a] 84[a]	23, 82,[a] 84[a]
	P2′Gln	28′, 50,[a] 84′[a]	28′, 84′[a]	28′, 30′,[a] 50,[a] 84′[a]	28′, 84′[a]	28′, 50,[a] 84′[a]
	P3′ Arg	29′	29′	29′	82	29′

WT, wild-type protease; Nle, norleucine.
*Hydrophobic atoms within a distance of 4 Å are considered in the interactions. The CA–p2 complexes of K45I and N88D have two disordered inhibitor molecules.
[a]Residues involved in resistance in the clinic.
[b]Residues in contact with the second orientation of the inhibitor.

0.83. In HIV-1 complexes with the nonpeptidic Food and Drug Administration (FDA) drugs, the GVI is within the same range observed for our complexes. However, these values are typically lower than those observed for other retroviral protease–inhibitor complexes, which show the higher values of 1.05 for HIV-2 protease (2mip), 1.5 for EIAV protease (2fmb), and 1.6 for FIV protease (1b11). Only SIV protease (1tcw) has a GVI of 0.84, comparable to those for HIV-1 protease. Even in the absence of inhibitor, the GVI has the lower value of 0.95 in unliganded HIV-1 protease (3hvp) compared with 1.29 in unliganded SIV protease (1az5). For the RSV S9 protease, which was engineered to recognize inhibitors of HIV-1 protease (23), the CA–p2 complex has a GVI of 1.23. Current experiments have suggested that mutation of binding site residues is not sufficient to completely alter the specificity of other retroviral proteases to that of HIV-1 protease, which has a broader specificity.[23,24]

The poorer subunit–subunit complementarity in other retroviral proteases might contribute to their narrower specificity compared with HIV-1 protease, and might also partly account for the poor inhibition of FIV and EIAV proteases by the FDA drugs.[25] The lower plasticity at the tight dimer interface in HIV-1 protease may also partly explain the observed intolerance to conservative changes like R87K.[26]

Structural Effects of Mutations

Among the mutants we have examined, K45I and D30N have side-chains that interact directly with the inhibitor, while side-chains of L90M and N88D have no direct

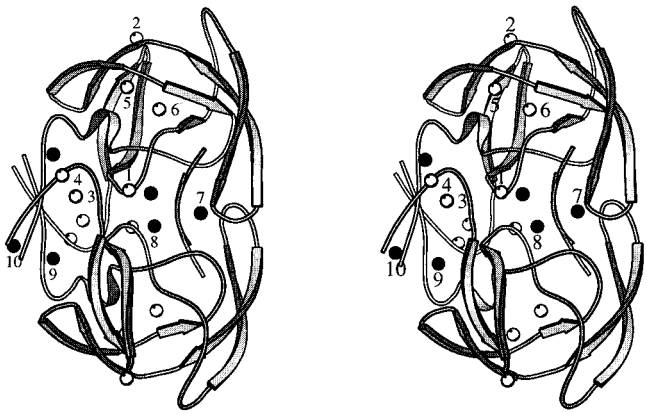

Fig. 4. Stereo drawing showing the location of conserved water molecules in the protease dimer. Waters in one of the subunits have been numbered and correspond to those in the text. Intersubunit waters are black, intrasubunit waters white.

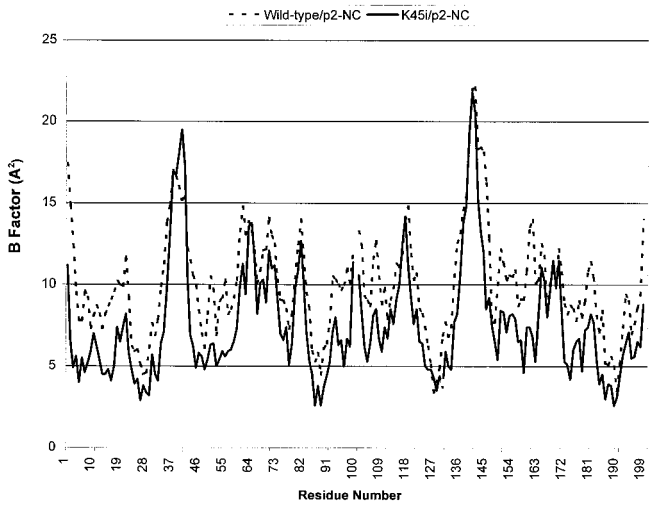

Fig. 5. Average B-factor for the main-chain atoms for K45I and wild-type with the p2–NC inhibitor. K45I is shown as continuous lines and wild-type as broken lines. Residues 1–99 and 101–199 correspond to the two subunits in the dimer.

contact with the inhibitor (Fig. 1). The structural changes will be described for individual mutations and correlated with the observed effects on protease activity and stability relative to wild-type enzyme.

Lys45 is situated in the flexible flap that is important for the binding of the substrate or the inhibitor. The flap consists of two β-strands connected by a short hairpin loop. K45 is located in the inner β-strand that is in direct contact with the inhibitor. The crystal structures suggest that mutation of lysine 45 to isoleucine has reduced the flexibility of the flap. Figure 5 shows the average main-chain B-factor for K45I and wild-type complexes with p2–NC inhibitor. The average B-factor was compared for all structures, except the L90M/CA–p2 complex, which was determined at room temperature. The average B-factor for all the main-chain atoms in the two K45I complexes (7.8 $Å^2$) is lower than in other structures (10.8 $Å^2$). More specifically, for main-chain atoms of the inner β-strand (residues 42–52 and 42′–52′) the average B values are 7.7 $Å^2$ and 10.7 $Å^2$ in K45I compared with 10.7 $Å^2$ and 14 $Å^2$ in other protease/p2–NC and CA–p2 complexes, respectively. Correspondingly, the main-chain atoms of the inhibitors have the lowest B-factors in K45I (8 and 9.5 $Å^2$) compared with an average of 12.6 and 13.1 $Å^2$ in other protease–p2–NC and CA–p2 complexes, respectively. The greater protease stability and activity of K45I[4] could be attributable to the observed lower mobility of the flaps and the inhibitor. When residue 45 is lysine, it is usually found to be ordered in one of the monomers and disordered in the other. NZ of Lys45 forms hydrogen bonds with Oδ1 of Asp30 or with a water molecule in the p2–NC complexes. Lys45′, when ordered, interacts with Oϵ1 of Gln58′. However, Ile45 and 45′ in K45I are both ordered and the Cδ1 makes hydrophobic contacts with Cδ and Cϵ of P4′Nle in the CA–p2 complex. Interactions at the S4 subsite have been seen in very few structures, as most of the studied inhibitors are shorter.[27,28] This could be one of the reasons K45 has not been seen as a resistant mutant in response to these drugs. The greater activity observed for K45I catalyzed cleavage of the PR/RT peptide[4] could result from favorable hydrophobic interactions of Ile45 with P3Leu and P4′Pro. Ile45 also contributes favorably to the hydrophobic pocket formed by side-chains of Ile47, Val 57, and Leu76. The lower mobility of the K45I main-chain as well as the favorable hydrophobic contribution of Ile45 could explain why K45I is more stable to urea denaturation than the wild-type.[4]

Leu90 is located in the short helix near the dimer interface. The side-chain of Leu90 lies in a hydrophobic pocket that contains Leu5, Thr26, Leu24, and Ala95. Substitution of the longer methionine for leucine has unfavorable steric effects. The side-chain of Met90 is seen to have two conformations in the L90M structures (Fig 6a). In one conformation, Cϵ of Met90 points toward Ala95. Residue 95 is a cysteine in the native protease and has been mutated to alanine in the protease clones we have studied. As seen from the comparison with the wild-type/CA–p2 complex,[20] the presence of cysteine would result in more constraints for this conformation in the L90M mutant. Although this conformation has fewer contacts due to the absence of cysteine, and would seem more favorable, a second conformation, where Cϵ of Met90′ points toward the carbonyl oxygen of Asp25′, is also seen in the L90M/CA–p2 complex. Here, Cϵ of Met90′ has unfavorable van der Waal contacts (3.2 Å) with the carbonyl oxygen of the catalytic Asp25′ (Fig 6a,b). The corresponding distance to Cδ1 of Leu90 is ≥3.6 Å in all other structures. Thus, the catalytic loop at the dimer interface is destabilized in the mutant. This observation correlates well with our studies in which the L90M dimer is seen to be most susceptible to urea denaturation and also has lowered activity relative to wild-type.[4] L90M also has poor replicative fitness in vivo[29] and reduced dimer stability.[11] L90M also shows small changes of 0.3–0.4 Å in Cδ atoms of Leu24′ and Leu5′ relative to wild-type protease. In the clinic, Leu24 contributes to resistance against Indinavir when present along with other mutants.[1] Since the dimer in HIV-1 protease is extremely tight, even small structural changes at this interface can greatly influence dimer stability as well as catalytic activity.

Asp30 lies in the short β-strand adjacent to the 79–loop and its side-chain atoms interact with the inhibitor and are also exposed to solvent and/or symmetry related molecules. The structural changes at Asp30 differ with respect to the interacting P2/P2′ residue of the inhibitor. In both the inhibitors, the P2 residue is hydrophobic (Val/Ile), whereas the P2′ residue (Glu/Gln) is polar. In CA–p2 complexes (orange in Fig. 7b), where Asp30 interacts with P2Val, the overall side-chain deviation for residue 30 with respect to wild-type is the largest (2 Å) in D30N compared with about 0.7 Å in all other mutants. The Asn30 side-chain has moved toward the inhibitor (Fig 7b, black dotted line), resulting in a shorter hydrogen bond of 3.1 Å between Asn30 Oδ2 and main-chain N of P3Arg in D30N, compared with 3.4 Å in other complexes. In the other subunit, Asp30′ interacts with P2′Glu of the inhibi-

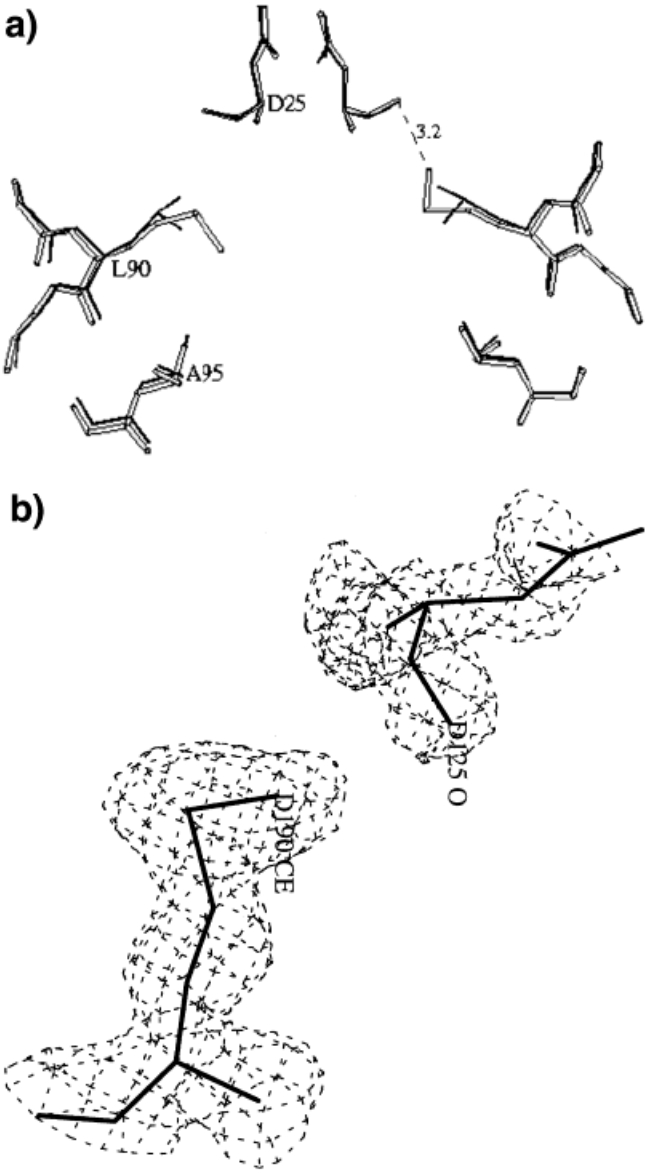

Fig. 6. **a:** The side-chain conformations of Met90 in the L90M/CA–p2 complex superposed on the wild-type/p2–NC complex. The wild-type atoms are black and L90M atoms are gray. The short van der Waal contact (3.2 Å) between the carbonyl O of Asp25′′ and Cϵ of Met90′ is shown. **b:** A 3Fo–2Fc SA omit map contoured at 1.6σ around residue Met90′. D125 O and D190 CE represent carbonyl oxygen of Asp25′ and Cϵ atom of Met90′ respectively in subunit D (residues 101–199). Residue numbers 1–99 and 101–199 correspond to first and second subunit of the dimer.

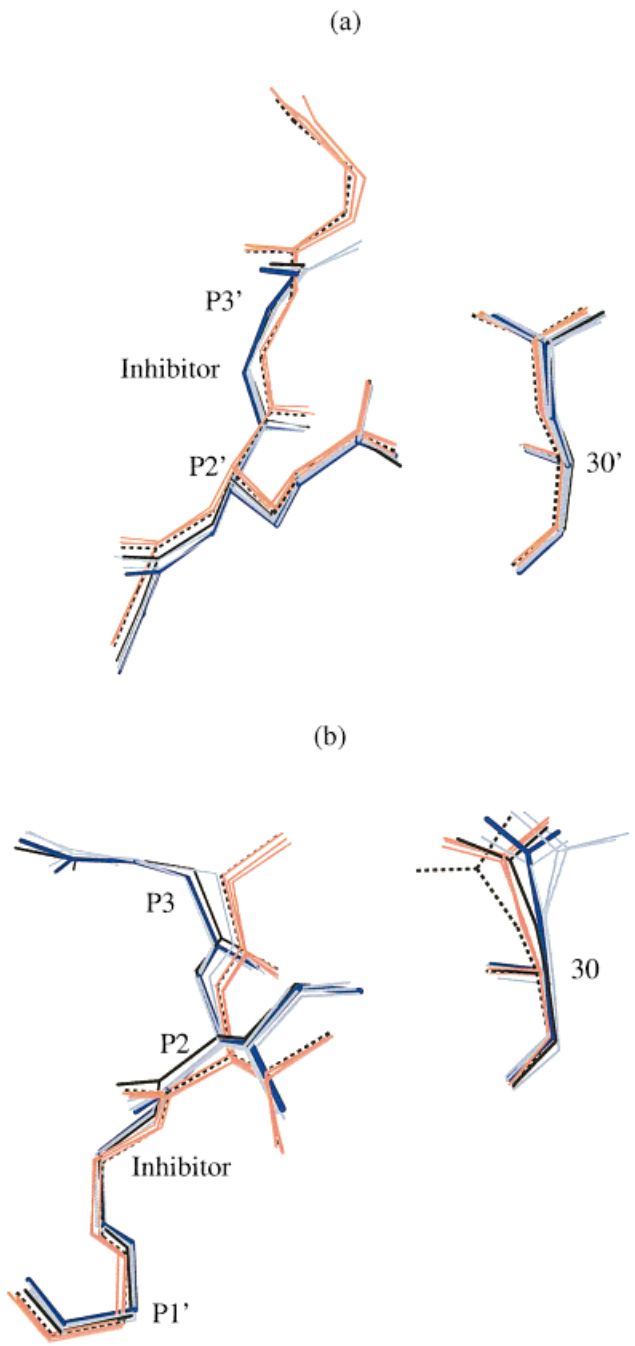

Fig. 7. Side-chain conformation of residues 30 and 30′ for mutants superimposed on wild-type/p2–NC. The inhibitor backbone and the side-chains of P2/P2′ residues are also shown. The CA–p2 complexes are indicated in orange and the p2–NC mutants in light blue lines. The wild-type/p2–NC complex is depicted by thicker dark blue lines. D30N/CA–p2 complex is indicated by dotted black lines and the D30N/p2–NC complex by continuous black lines. *Top*: Asp30′ and P2′Glu/Gln. *Bottom*: Asp30 and P2Val/Ile.

tor through a proton-mediated interaction, which has been observed in many retroviral complexes.[20,30,31] In D30N, the proton-mediated interaction with P2′Glu is replaced by a hydrogen bond with the corresponding Nδ2 of Asn30′. At Asp30′ side-chain deviations for residue 30′ with respect to wild-type are about 0.4 Å in all the complexes, including D30N (Fig. 7a, orange and black dotted line).

In the protease/p2–NC complexes (Fig. 7b, blue), Asp30 interacts with P2Ile of the inhibitor. Here, the side-chain deviation of residue 30 with respect to wild-type varies from 0.3–1.0 Å in other mutants, to the largest value of 1.6 Å in D30N (Fig 7b, continuous black line). However, the water-mediated interactions of Asp30 Oδ1 with Asn88 Nδ2 and P3Thr Oγ1 are maintained in D30N as well. The water-mediated interaction of Asp30 Oδ1 with carbonyl oxygen of Thr74, which is seen in two of the complexes, is

replaced by a weaker interaction through two waters in D30N. At Asp30′ which interacts with P2′Gln, the side-chain variations with respect to wild-type are about 0.2 Å in other mutants and rise to 1.2 Å in D30N (Fig 7a, continuous black line). The hydrogen bond between Asp30′Oδ2 and Nϵ2 of P2′Gln and the interaction of Asp30′Oδ1 with a water molecule are observed in D30N as well.

Thus, in both inhibitor complexes of the D30N mutant the side-chain of Asn30 shows the largest changes of all mutants relative to wild-type, suggesting altered interactions with inhibitor, specifically at position P2, consistent with the observed variation in activity on different peptide substrates.[4] Previous kinetic studies on the hydrolysis of Gag–pol peptides have shown that the D30N mutant cleaved the CA–p2 peptide 4–5 times more efficiently than it cleaved $p6^{pol}$–PR and PR–RT.[4] The CA–p2 peptide has an Arg at P3, unlike the hydrophobic residues in $p6^{pol}$–PR (Phe) and PR–RT (Leu) peptides. Since the environment around Asp30 is mostly polar, hydrophobic residues are less favorable and probably more so in D30N, where the side-chain has moved closer to the inhibitor. Hence, D30N hydrolyzes the CA–p2 peptide more efficiently. Many of the polar contacts of Asp30 are maintained through altered interactions in D30N, which suggests that they have an important role in structure or catalysis, or both. The maintenance of polar interactions in D30N is consistent with the observation that D30N and wild-type protease share similar sensitivity to urea.[4]

Residue Asn88 lies in the terminal helix and its side-chain atoms have proton and water-mediated interactions with Thr31 and Thr74 in both subunits of the dimer. There is little variation at Asn88 side-chains in N88D (RMSD with respect to wild-type/p2–NC is 0.2–0.5 Å). The hydrogen bond between carbonyl oxygen of Thr74 and Nδ2 of Asn88 is replaced by a weaker interaction of Thr74 with Oδ1 of Asp88 in N88D. In all the structures, Nδ2 of Asn88 is linked to main-chain atoms of Thr31 and Thr74 through two waters (waters #5 and #6 in Fig. 4), which can stabilize the secondary structure. Nδ2 of Asn88 is also linked to Asp30 Oδ through one of these waters in some complexes, while in others a different water molecule is involved. Hence it is possible to see correlated changes at N88D and D30N. It is not surprising therefore that, in patients treated with Nelfinavir, N88D is usually seen along with D30N. These waters could help stabilize the inhibitor complex by indirect interactions through polar residues like P2′Glu/Gln. However, in the N88D/CA–p2 complex, waters #5 and #6 are missing in both subunits. Interestingly, in this complex, the inhibitor is bound in two opposite orientations to the protease. In the N88D/p2–NC complex, these waters are retained in the subunit that interacts with the P1′–P3′ side of the inhibitor, where the environment is polar due to P2′Gln and P3′Arg. Although it is not clear from the structures of these complexes how the changes affect catalysis, they could alter/effect an intermediate step in catalysis or binding. Hence, it is possible that with a hydrophobic residue at P2 position Asn88 is not able to accommodate the water molecules involved at some hydrolysis step. Side-chains of Asp30, Asn88, and Lys45 are spatially adjacent and are in the vicinity of the substrate binding pocket. Since Asp30 is in direct contact with the inhibitor through hydrogen bonds or water molecules, mutation at any one of these residues is likely to affect the other two sites.

CONCLUSIONS

The mutants studied in this article represent both active-site mutations (D30N and K45I) as well as mutations in residues that are distant from the active site in the protease (N88D and L90M). Each of these mutations contributes differently to the properties of the protease. L90M alters van der Waals interactions in the hydrophobic interior at the dimer interface near the catalytic aspartates. Its effects on catalysis are correlated with the stability of the dimer. K45I replaces a charged residue with a hydrophobic residue in the flap. The mutation influences catalysis as well as structural stability by stabilizing the flaps. N88D and D30N represent mutations that alter polar residues and some of the polar contacts are maintained through altered interactions. This could suggest a possible role either in maintaining structural stability or in catalysis. D30N shows altered interactions with hydrophobic P2 side-chains, consistent with the observed variable specificity. N88D is missing water-mediated interactions with Thr31 and Thr74. Side-chains of Asn88 and Asp30 are linked through a water in some complexes, which could possibly explain why N88D and D30N are seen together in the clinic. At the tight dimer interface, relatively fewer polymorphisms as well as mutations are seen, since they can have unfavorable effects both on activity as well as stability. The implication is that inhibitors that target the dimer interface along with the active site could be more effective in avoiding mutations.

Water molecules are important for maintaining the structural stability as well as catalytic activity of the protease. The water-mediated interactions, especially at the dimer interface will be important in the design of dimerization inhibitors. Residues around the active site have conformational flexibility, which appears to be an important factor in adapting to different substrates. Subtle structural changes seen in these complexes emphasize the need for careful refinement with high-resolution X-ray data. Structural changes seen in the protease/inhibitor complex alone however, may not be sufficient to explain the effect of mutations, since there may be changes associated with the unliganded protease that are not seen in the final complex. The structural analysis of these mutants will be useful in the design of inhibitors that can be more effective in avoiding mutations in the protease.

ACKNOWLEDGMENTS

The authors thank Dr. Malcolm Capel and staff at beamline X12B at the National Synchrotron Light Source, Brookhaven, for data collection. We also thank Y.F. Wang, J. Petock, and Dr. S. Scott for help during data collection. We thank J. Krouse for purifying two of the mutants and C. Fischer for refolding the wild-type protease. This re-

search was supported in part by U.S. Public Health Service grants AI41380 and GM62920 (to I.T.W. and R.W.H) and by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of National Institutes of Health (to J.M.L).

REFERENCES

1. Shafer RW, Stevenson D, Chan B. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res 1999;27:348–352.
2. Hertogs K, Bloor S, Kemp SD, Van den Eynde C, Alcorn TM, Pauwels R, Van Houtte M, Staszewski S, Miller V, Larder BA. Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. AIDS 2000;14:1203–1210.
3. Vondrasek J, van Buskirk CP, Wlodawer A. Database of three-dimensional structures of HIV proteinases. Nature Struct Biol 1997;4:8.
4. Mahalingam B, Louis JM, Reed CC, Adomat JM, Krouse J, Wang YF, Harrison RW, Weber IT. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur J Biochem 1999;263: 238–245.
5. Schinazi RF, Larder BA, Mellors JW. Mutations in retroviral genes associated with drug resistance. Int Antiviral News 1997;5: 129–142.
6. Tessmer U, Krausslich HG. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. J Virol 1998;72:3459–3463.
7. Louis JM, Clore GM, Gronenborn AM. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. Nature Struct Biol 1999;6:868–875.
8. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol 1994;68: 8017–8027.
9. Misumi S, Kudo A, Azuma R, Tomonaga M, Furuishi K, Shoji S. The p2gag peptide, AEAMSQVTNTATIM, processed from HIV-1 Pr55gag was found to be a suicide inhibitor of HIV-1 protease. Biochem Biophys Res Commun 1997;241:275–280.
10. Accola MA, Hoglund S, Gottlinger HG. A putative alpha-helical structure which overlaps the capsid–p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. J Virol 1998;72:2072–2078.
11. Xie D, Gulnik S, Gustchina E, Yu B, Shao W, Qoronfleh W, Nathan A, Erickson JW. Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH. Protein Sci 1999;8:1702–1707.
12. Louis JM, Wondrak EM, Copeland TD, Smith CA, Mora PT, Oroszlan S. Chemical synthesis and expression of the HIV-1 protease gene in *E. coli*. Biochem Biophys Res Commun 1989;159: 87–94.
13. Wondrak EM, Louis JM. Influence of flanking sequences on the dimer stability of human immunodeficiency virus type 1 protease. Biochemistry 1996;35:12957–12962.
14. Otwinowski Z, Minor W. Processing of X-ray diffraction data in oscillation mode. Methods Enzymol 1997;276:307–326.
15. Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr D Biol Crystallogr 1994;D50:157–163.
16. Brunger AT. X-PLOR version 3.1: a system for X-ray crystallography and NMR. New Have, CT: Yale University Press; 1992.
17. Jones TA. A graphics model building and refinement system for macromolecules. J Appl Crystallogr 1978;11:268–272.
18. Cohen GE, ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. J Appl Crystallogr 1997; 30:1160–1161.
19. Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA III. Complete mutagenesis of the HIV-1 protease. Nature 1989;340:397–400.
20. Weber IT, Wu J, Adomat J, Harrison RW, Kimmel AR, Wondrak EM, Louis JM. Crystallographic analysis of human immunodeficiency virus 1 protease with an analogue of the conserved CA–p2 substrate—interactions with frequently occurring glutamic acid residue at P2′ position of substrates. Eur J Biochem 1997;249:523–530.
21. Rose JR, Babe LM, Craik CS. Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. J Virol 1995;69:2751–2758.
22. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
23. Wu J, Adomat JM, Ridky TW, Louis JM, Leis J, Harrison RW, Weber IT. Structural basis for specificity of retroviral proteases. Biochemistry 1998;37:4518–4526.
24. Lin YC, Beck Z, Lee T, Le VD, Morris GM, Olson AJ, Wong CH, Elder JH. Alteration of substrate and inhibitor specificity of feline immunodeficiency virus protease. J Virol 2000;74:4710–4720.
25. Kervinen J, Lubkowski J, Zdanov A, Bhatt D, Dunn BM, Hui KY, Powell DJ, Kay J, Wlodawer A, Gustchina A. Toward a universal inhibitor of retroviral proteases: comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. Protein Sci 1998;7:2314–2323.
26. Louis JM, Smith CA, Wondrak EM, Mora PT, Oroszlan S. Substitution mutations of the highly conserved arginine 87 of HIV-1 protease result in loss of proteolytic activity. Biochem Biophys Res Commun 1989;164:30–38.
27. Wlodawer A, Gustchina A. Structural and biochemical studies of retroviral proteases. Biochim Biophys Acta 2000;1477:16–34.
28. Tozser J, Gustchina A, Weber IT, Blaha I, Wondrak EM, Oroszlan S. Studies on the role of the S4 substrate binding site of HIV proteinases. FEBS Lett 1991;279:356–360.
29. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type1. J Virol 1999;73:3744–3752.
30. Louis JM, Dyda F, Nashed NT, Kimmel AR, Davies DR. Hydrophilic peptides derived from the transframe region of Gag–Pol inhibit the HIV-1 protease. Biochemistry 1998;37:2105–2110.
31. Rose RB, Rose JR, Salto R, Craik CS, Stroud RM. Structure of the protease from simian immunodeficiency virus: complex with an irreversible nonpeptide inhibitor. Biochemistry. 1993;32:12498–12507.