Structure, Vol. 10, 369–381, March, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00720-7

Substrate Shape Determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes

Moses Prabu-Jeyabalan, Ellen Nalivaika, and Celia A. Schiffer[1]
Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA 01655

Summary

The homodimeric HIV-1 protease is the target of some of the most effective antiviral AIDS therapy, as it facilitates viral maturation by cleaving ten asymmetric and nonhomologous sequences in the Gag and Pol polyproteins. Since the specificity of this enzyme is not easily determined from the sequences of these cleavage sites alone, we solved the crystal structures of complexes of an inactive variant (D25N) of HIV-1 protease with six peptides that correspond to the natural substrate cleavage sites. When the protease binds to its substrate and buries nearly 1000 Å^2 of surface area, the symmetry of the protease is broken, yet most internal hydrogen bonds and waters are conserved. However, no substrate side chain hydrogen bond is conserved. Specificity of HIV-1 protease appears to be determined by an asymmetric shape rather than a particular amino acid sequence.

Introduction

As the worldwide AIDS epidemic continues into its third decade, a cure for HIV-1 still eludes the medical community [1]. In the absence of a cure for HIV-1 pathogenesis, suppressing viral replication and maintaining it at low to undetectable levels have become critical goals in the field of HIV-1 research [2–5]. To this end, highly active antiretroviral therapy (HAART) has become a successful strategy in providing long, quality lives for infected individuals [6, 7]. Many patients have had complete response to HAART. Reports of failure, partial response, and/or breakthrough with antiretroviral treatment, as measured by viral load, however, have compromised the future of HIV-1 treatment [8, 9]. Viral resistance has been recognized as one of the most important factors involved in therapeutic failure [10, 11]. A comprehensive understanding of the development of HIV-1 resistance to antiretroviral agents is critical to improving therapeutic management [12–15]. Protease inhibitors are essential components of most HAART therapies [16, 17]. Six FDA-approved HIV-1 protease inhibitors, all of which are competitive inhibitors binding at the active site, are presently on the market. These drugs are often the first lines of treatment for infected patients, as they are well tolerated. All these drugs are peptidomimetics that resulted from structure-based drug design efforts of the pharmaceutical industry. All of them have large, generally hydrophobic moieties that interact with the mainly hydrophobic P2–P2′ pockets in the active site [18]. Drug resistance in HIV-1 protease presents a new challenge, however, to future structure-based drug design efforts. Although drug resistance by HIV has been well characterized [19–25], understanding the subtle balance of molecular recognition events that confer drug resistance in HIV-1 is crucial to the development of second generation drugs in the treatment of HIV-1 infection.

HIV protease is the aspartyl protease that processes the Gag and Pol polyproteins and allows for the maturation of the immature HIV virion, thus allowing the spread of the virus. Remarkably, the precise physical parameters that govern how HIV-1 protease binds to its ten natural, nonhomologous substrates [26–31] (Table 1) remain poorly understood. The active site of the homodimeric protease is at the dimer interface [18, 32]. Despite the symmetry conferred on its active site because it is a homodimer, the enzyme recognizes asymmetric substrate sites within the Gag and Pol polyproteins. The amino acid sequences of these substrates are asymmetric around the cleavage sites in both size and charge distribution. In addition, these sites share little sequence homology. How then does the protease recognize a particular peptide sequence as being a substrate? There must be a breakdown in the symmetry within the individual protease dimer when it binds to its substrates. This breakdown has often been difficult to characterize, however, since many of the complexes of HIV protease bound to asymmetric ligands do not uniquely orient the protease dimer in the crystal cell. This lack of unique orientation resulted in protease-substrate structures with 50% of the ligand oriented in one direction and 50% in the other, thus averaging out the asymmetry within the protease. To elucidate how HIV-1 protease recognizes its substrates, we determined the crystal structures of six complexes of HIV-1 protease with decameric peptides that correspond to six substrate cleavage sites within the Gag and Pol polyproteins (Figure 1A; Table 2). Analysis of these structures shows that the protease recognizes an asymmetric shape adopted by the substrate peptides rather than a particular amino acid sequence. As HIV-1 protease also binds inhibitors in the same active site, understanding how the enzyme recognizes substrates may be critical for the next generation of inhibitor design.

Results

Determination of Crystal Structures of Inactive HIV-1 Protease-Substrate Complexes

An inactive variant (D25N) of HIV-1 protease [33, 34] was complexed with peptides corresponding to six of the ten substrate sequences in the Gag and Pol polyproteins, and their crystal structures were determined (Figure 1A; Table 2). Four of these peptides correspond to

[1]Correspondence: celia.schiffer@umassmed.edu

Key words: crystal structure; HIV-1 protease; specificity; substrate recognition; toroidal shape; water structure

Table 1. Sequences of the Ten Sites within the HIV-1 Gag and Pol Polyproteins that are Cleaved by HIV-1 Protease

Substrate Cleavage Site	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′
MA-CA	**Val**	**Ser**	**Gln**	**Asn**	**Tyr**	**Pro**	**Ile**	**Val**	**Gln**	**Asn**
CA-p2	**Lys**	**Ala**	**Arg**	**Val**	**Leu**	**Ala**	**Glu**	**Ala**	**Met**	**Ser**
p2-NC	**Pro**	**Ala**	**Thr**	**Ile**	**Met**	**Met**	**Gln**	**Arg**	**Gly**	**Asn**
NC-p1	Glu	Arg	Gln	Ala	Asn	Phe	Leu	Gly	Lys	Ile
p1-p6	**Arg**	**Pro**	**Gly**	**Asn**	**Phe**	**Leu**	**Gln**	**Ser**	**Arg**	**Pro**
TF-PR	Val	Ser	Phe	Asn	Phe	Pro	Gln	Ile	Thr	Leu
AutoP	Pro	Gln	Ile	Thr	Leu	Trp	Lys	Arg	Pro	Leu
PR-RT	Cys	Thr	Leu	Asn	Phe	Pro	Ile	Ser	Pro	Ile
RT-RH	**Gly**	**Ala**	**Glu**	**Thr**	**Phe**	**Tyr**	**Val**	**Asp**	**Gly**	**Ala**
RH-IN	**Ile**	**Arg**	**Lys**	**Ile**	**Leu**	**Phe**	**Leu**	**Asp**	**Gly**	**Ile**

The crystal structures of the six sequences in bold were determined in complex with an inactive HIV-1 protease variant. The cleavage sites are identified by the proteins released once the site is cleaved: matrix (MA), capsid (CA), nucleocapsid (NC), trans frame peptide (TF), protease (PR), auto proteolysis site (AutoP), reverse transcriptase (RT), RNAse H (RH), and integrase (IN).

substrate sequences within the Gag polyprotein (matrix-capsid [MA-CA], capsid-p2 [CA-p2], p2-nucleocapsid [p2-NC], and p1-p6) and two correspond to substrate sequences within the Pol polyprotein (reverse transcriptase-RNaseH [RT-RH] and RNaseH-integrase [RH-IN]). The MA-CA complex crystallized in the space group I222 with two dimers in the asymmetric unit and diffracted to a resolution of only 2.9 Å. The remaining complexes crystallized in space group $P2_12_12_1$ with similar cell dimensions and diffracted to 2.0 Å resolution with many water sites resolved in the electron density. Decameric peptides were used in the crystallization. Although all ten residues in the peptide were not always resolved, in every case, the longer peptide uniquely oriented the protease dimer in the crystal cell. The direction of the peptide through the dimer in the asymmetric unit depended, however, on which peptide was bound. Of the five complexes having the same space group, two had the peptide oriented in one direction, and three had the peptide oriented in the opposite direction. Having these six substrate protease complexes to analyze thus reduced the possibility that the changes observed from one monomer to the other were primarily due to crystal packing.

All six substrate peptides were bound in an extended conformation with an inactive variant (D25N) of HIV-1 protease. In all the complexes, the residues that should correspond to the P1 and P1′ sites within the peptides, as previously identified from enzymatic studies [35–40], bridged the asparagines in the mutated "active site." The substrates formed hydrogen bonds with the protease flaps, making a parallel β sheet with one flap on the unprimed side and an antiparallel β sheet on the primed side, as has been seen previously in complexes of peptidomimetics [31, 41–43]. A sharp deviation in the backbone torsion occurred at the cleavage site (P1–P1′; $\phi \approx -85°$, $\psi \approx 50°$) about the scissile peptide bond. This alteration enables the carboxyl oxygen of P1 to make an invariant hydrogen bond with the "catalytic" Asn25′ ND2.

Analysis of the Complexes: Binding of Substrate Peptides

Conserved Hydrogen Bonds

Because the substrates bind primarily in β strand conformations, most of the conserved hydrogen bonds occur primarily between the backbone of the protease and the backbone of the substrates (Figure 2). Eight completely conserved hydrogen bonds involve substrate residues, P4–P4′, and protease residues Asn25′, Gly27/27′, Asp29/29′, and Gly48/48′. In addition, there are six partially conserved hydrogen bonds: four involving the backbone atoms and two involving a substrate side chain and the amino nitrogen atom of Asp29/29′ and Asp30/30′. In p1-p6 and RH-IN, where the backbone conformations are altered at P3 and P4, several otherwise conserved backbone hydrogen bonds are disrupted (Table 3A). In RH-IN, RT-RH, and p2-NC, a conserved glycine residue at P4′ forms a hydrogen bond between P4′N and Asp29′ OD2. Thus, many hydrogen bonds formed between the substrate peptides and the inactive protease are not sequence specific in the complex, as they are with the backbone atoms of the peptides.

Substrate-Specific Hydrogen Bonds

Specificity is often conferred in protein-ligand interactions by hydrogen bonds between side chains. This is especially the case for proteases that often need to cleave at a particular site. In the six complexes studied, each substrate peptide formed between one and three side chain hydrogen bonds with the protease (Table 3B). Although these hydrogen bonds did not form between any specific residue or site, they often involve the P2/P2′ or P3/P3′ residues of the peptide. Only four protease side chains (Arg8/8′, Asp29, Asp30/30′, and Gln58′) formed direct side chain hydrogen bonds with the substrate peptides. When the P2/P2′ residue is either Glu or Gln, as in CA-p2, p2-NC, and p1-p6, the side chain makes a hydrogen bond with Asp30′ OD2 as well as with the amide nitrogens of aspartic acid residues 29 and 30. The latter hydrogen bonds are also formed when P2/P2′ is either Asp or Asn. The absence of a polar side chain at the P2/P2′ position in the structures MA-CA, RT-RH, and RH-IN prevents them from forming this kind of hydrogen bond. Two other side chain hydrogen bonds are formed, however, in MA-CA: one between SerP4 OG and Asp30 OD2 and one between GlnP3 OE1 and Arg8′ NH2. In addition, in the RT-RH and RH-IN complexes, AspP3′ forms a salt bridge with Arg8. The RT-RH complex also makes a hydrogen bond between TyrP1′ OH and Arg8 NH1, the only instance at either P1 or P1′

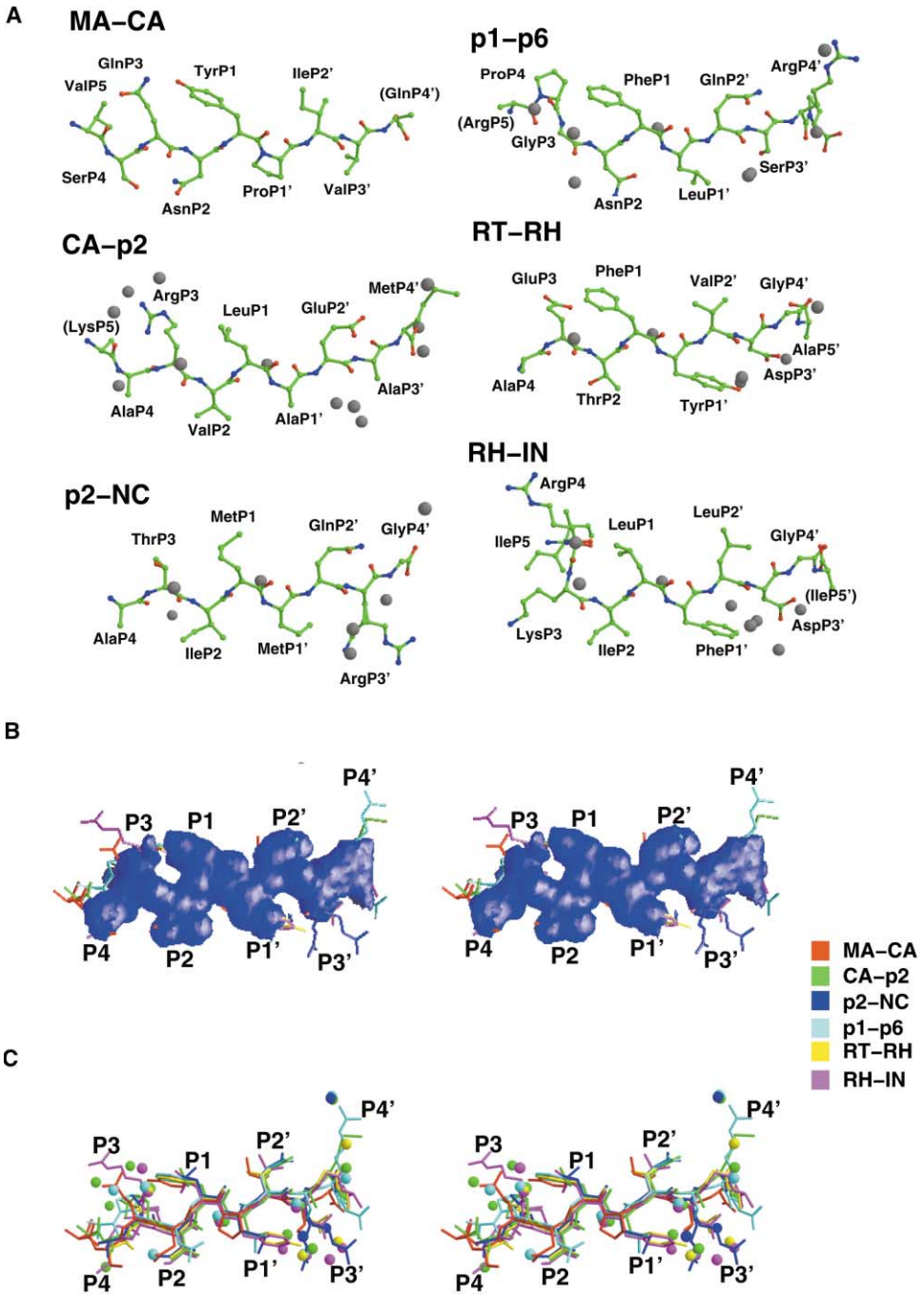

Figure 1. Conformation of the Six Substrate Peptides

(A) The conformation of six substrate peptides as observed in complexes with an inactive yet isosteric variant of HIV-1 protease, D25N. The peptides are colored by atom type, and crystallographic water molecules within 4.5 Å of the peptide are shown as small gray spheres.

(B) Consensus volume (shown in stereo) occupied by the substrate peptides when they are bound to the inactive D25N HIV-1 protease. The structures were superimposed using the α-carbon atoms of residues 1–9 and 86–99 in both monomers. The blue surface indicates the region that is within the van der Waals radii simultaneously of at least four of the superimposed peptides.

(C) The same superposition showing the peptides and water molecules within 4.5 Å of the peptide.

where this kind of hydrogen bond is formed. Thus, no single substrate side chain hydrogen bond is conserved between all six substrate complexes.

Shape Complementarity

Less specific, yet still important to binding, are the changes in solvent-accessible surface area [44] when substrate and protease bind and the extent to which their shapes are complementary. A total of 900–1000 Å^2 of the accessible substrate surface is buried by the protease upon substrate binding (Table 2). These extensive buried surface areas are on par with many protein-protein interfaces. The shape complementarity between

Table 2. Crystallographic Statistics for the Six Complexes

Parameter	Protease-Substrate Complex					
	MA-CA	CA-p2[a]	p2-NC	p1-p6	RT-RH	RH-IN
Protein Data Bank ID code	1KJ4	1F7A[a]	1KJ7	1KJF	1KJG	1KJH
Data Collection						
a (Å)	91.65	51.62	51.51	51.29	51.30	51.73
b (Å)	93.81	59.04	59.43	59.07	58.65	59.08
c (Å)	118.17	61.35	61.74	61.81	62.10	61.26
Space group	I222	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Z	16	4	4	4	4	4
Resolution (Å)	2.9	2.0	2.0	2.0	2.0	2.0
Total number of reflections	41,786	49,525	68,643	41,786	60,432	74,481
Number of unique reflections	12,376	12,474	12,913	12,376	13,019	12,905
R_{merge} (%)[b]	10.7	4.8	4.7	6.7	7.0	6.9
Completeness (%)	95.5	94.3	96.7	93.4	98.6	97.5
Refinement						
R value (%)	19.7	19.7	20.6	20.3	18.4	18.8
R_{free} (%)	24.8	23.3	24.9	25.1	22.6	22.7
Rmsd						
Bond length (Å)	0.008	0.005	0.006	0.006	0.006	0.006
Bond angles (°)	1.5	1.3	1.4	1.3	1.3	2.7
Dihedral angles (°)	26.8	26.2	26.8	26.1	26.0	26.6
Improper angles (°)	1.0	0.9	1.2	0.9	0.8	0.9
Number of acetate ions	12	6	5	4	2	2
Protein Hydration						
Number of water molecules	43	96	73	101	89	87
Interactions involving only backbone	—	111	106	126	126	123
Interactions involving only side chain	—	55	55	62	62	59
Buried Surface Area						
Total area buried on substrate binding (Å²)	1007	1038	990	1090	1045	1182

[a] Adaped from [30].
[b] $R_{merge} = \Sigma|I - \bar{I}|^2 / |I|^2$.

these substrates and the protease is between 0.65 and 0.70 (values above 0.5 indicating high complementarity) [45]. With nearly 1000 Å² buried and high shape complementarity on the substrate-protease interface, the shapes of the substrates must play a critical role in their recognition. To determine which parts of the peptide

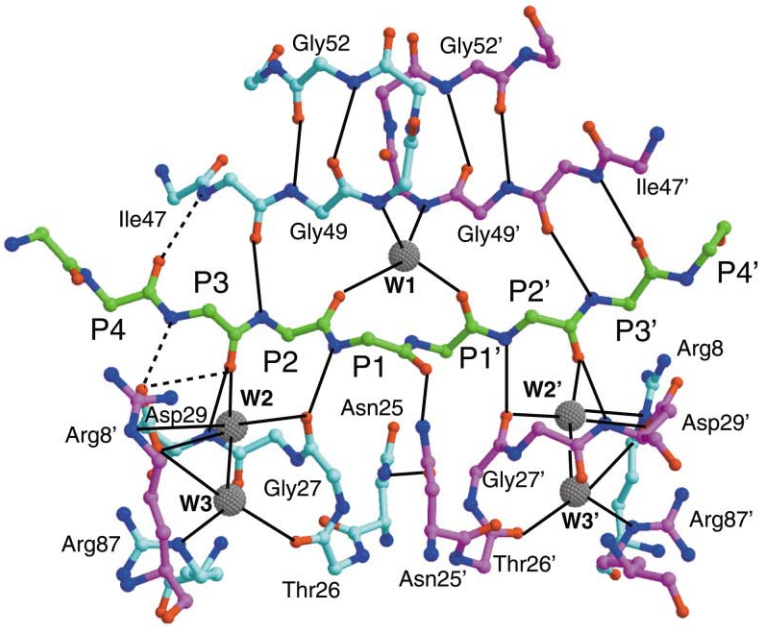

Figure 2. Conserved Substrate-Protease Hydrogen Bonds

Conserved hydrogen bonds are shown between the peptide in green, and the two monomers of the protease are distinguished in cyan and magenta. The hydrogen bonds are primarily backbone, with only the side chains Asp29 and Asn25 shown as conserved peptide hydrogen bonds. Solid lines indicate hydrogen bonds that are conserved in all six complexes. The dashed lines indicate hydrogen bonds that are conserved in all complexes except RH-IN and p1-p6. Five water molecules are also conserved and bridge the peptide to the protease.

Table 3A. Conserved Hydrogen Bonds and Their Relative Distances in Angstroms (Å)

Substrate Atom	Protein Atom	MA-CA Dimer 1[a]	MA-CA Dimer 2[a]	CA-p2	p2-NC	p1-p6	RH-IN	RT-RH
P4 O	Gly48 N	2.7	2.8	3.2				3.2
P3 O	Asp29 OD2	2.7	2.8	3.5	3.3	3.5		3.5
P3 O	Asp29 N	3.0	2.8	2.9	3.3	2.8	2.8	3.3
P3 N	Asp29 OD2	3.0	2.8	2.9	2.7			3.1
P2 N	Gly48 O	2.7	2.9	3.0	2.9	3.1	2.9	2.9
P1 N	Gly27 O	2.8	2.6	3.0	2.9	2.9	2.9	2.8
P1 O	Asn25′ ND2	2.8	2.8	2.7	2.5	2.7	2.7	2.7
P2′ N	Gly27′ O	3.2	3.0	3.0	2.9	3.0	3.1	3.0
P2′ O	Asp29′ N	3.4	3.2	3.1	3.1	3.1	3.0	3.2
P3′ N	Gly48′ O	2.5	2.8	3.1	2.9	3.0	2.7	2.7
P3′ O	Gly48′ N	2.8	3.1	2.9	3.3	2.8	3.0	3.2
P4′ N	Asp29′ OD2 (only when P4′ is Gly)				3.0		2.8	2.9

[a] The two dimers found in the crystals of MA-CA were analyzed separately.

Table 3B. Observed Side Chain to Side Chain and Side Chain to Backbone Hydrogen Bonds

	Side Chain to Side Chain Interaction			Side Chain to Backbone Interaction		
Cleavage Site	Substrate Atom	Protein Atom	Distance	Substrate Atom	Protein Atom	Distance
MA-CA (dimer 1)	SerP4 OG	Asp30 OD2	2.4	AsnP2 OD1	Asp29 N	2.9
	GlnP3 OE1	Arg8′ NH2	3.0	AsnP2 OD1	Asp30 N	2.8
MA-CA (dimer 2)	SerP4 OG	Asp30 OD2	2.5	AsnP2 OD1	Asp29 N	2.7
	GlnP3 OE1	Arg8′ NH2	3.1	AsnP2 OD1	Asp30 N	2.9
Ca-p2	GluP2′ OE1	Asp30′ OD2	2.5	GluP2′ OE2	Asp29 N	2.9
				GluP2′ OE2	Asp30 N	3.3
p2-NC	GlnP2′ NE2	Asp30′ OD2	2.9	GlnP2′ OE1	Asp29 N	3.1
	ThrP3 OG1	Asp29 OD2	3.4	GlnP2′ OE1	Asp30 N	2.7
		Arg8′ NH2	2.8			
	ArgP3′[b] NE	Arg8 NH1	3.0			
		Arg8 NH2	3.2			
p1-p6	GlnP2′ NE2	Asp30′ OD2	3.0	GlnP2′ OE1	Asp29 N	2.9
	ArgP4′ NE	Asp30′ OD1	3.2	GlnP2′ OE1	Asp30 N	2.7
	ArgP4′ NE	Asp30′ OD2	3.0			
	ArgP4′ NH2	Gln58′ OE1	3.4			
RT-RH	TyrP1′ OH	Arg8 NH1	3.2	GlnP3 NE2	Gly48 O	3.0
	AspP3′ OD1	Arg8 NH1	3.1			
RH-IN	AspP3′ OD1	Arg8 NH1	3.4			

[b] Seen in one of the two observed conformations of ArgP3′.

shape are conserved, a consensus volume was calculated based on the volume occupied by the six substrate peptides [46] when the terminal domains of the complexes were superimposed. This volume is asymmetric in shape, with a toroid on the unprimed side of the cleavage site and an extended shape on the primed side (Figure 1B). In the four peptides corresponding to the MA-CA, CA-p2, p2-NC, and RT-RH cleavage sites, the toroid is formed by the side chain at P1 packing against the side chain at P3, while the peptide backbone remains in an extended conformation. In the remaining two peptide structures corresponding to the p1-p6 and RH-IN cleavage sites, however, the toroid is formed not by the side chains but rather by a change in peptide backbone conformation. In the peptide p1-p6 (Table 1), the P3 and P4 side chains are glycine and proline residues, respectively, and the toroid shape is created by a tight turn at the glycine residue, placing the P4 proline in the space occupied by the P3 side chain of the four other structures described above. The RH-IN peptide has two very large side chains at P3 and P4, lysine and arginine, respectively, and forms the toroid by extending the lysine side chain at P3 along the direction of the backbone into solvent as the backbone of P4 and the arginine wrap back against the leucine at P1. The average sum of the buried surface area for the residues at P1 and P3/P4 is 300 Å^2 compared with 230 Å^2 for P1′ and P3′. Although this conservation of substrate shape could not be easily predicted from sequence homology, a specific shape is the likely determinant of whether or not a particular sequence is recognized as a substrate by HIV-1 protease.

Water Mediation in Substrate Recognition

As with the substrate side chain hydrogen bonds, the conserved hydrogen bonds between the peptides with water primarily involve backbone atoms. Five waters are completely conserved between the five structures that were determined to 2.0 Å resolution (Figure 2). These conserved waters include three that directly contact the peptides [31, 47] and two that stabilize them. W1 tethers the flaps of the protease at Ile50 N and Ile50′ N with P2 O and P1′ O. W2/W2′ hydrogen bond to the peptide at P3 O/ P2′ O and to the protease at Arg8 NE, Gly27′ O, and Asp29' OD1 (Arg8′ NE, Gly27 O, and Asp29 OD1) as well as to two other waters, W3/W3′. Although W3/W3′ do not directly contact the substrate peptides, they

do stabilize their conformation by binding to the protease at Thr26 O, Asp29 OD1, and Arg87 NE. These five waters stabilize the extended conformation of the peptides by hydrogen bonding to the carbonyl oxygens of the peptide to which the protease in the absence of water could not directly hydrogen bond.

Several additional waters surround each of the peptides but do not make direct hydrogen bonds (Figure 1C). The major role of these waters appears to be to occupy space. While the residues on the primed side of the peptides generally do not form the peptidic toroid, when water molecules within 4.5 Å of the peptide are analyzed, a cluster of waters is seen on the primed side but not on the unprimed side (Figure 1C). This cluster, which is located at the position between the P1′ and P3′ residues where a toroid could have formed if the peptides were more symmetric, further accentuates the asymmetry of the bound peptide. This cluster of water molecules might aid in the recognition of the peptide by the protease or be crucial for product release once the substrate is cleaved.

Analysis of the Complexes: Adaptation of the Protease

Overall Protease Conformation

When the protease binds a substrate, the structural symmetry of the protease homodimer is broken as the monomers adjust to accommodate the asymmetric peptides. Double difference plots of each dimer showed that their monomers varied in several regions (data not shown). The outer loop residues, 16–18, 35–41, and 65–70, which vary the most between monomers, however, have higher temperature factors and are near crystal contacts. To distinguish between structural changes in the protease monomers due to substrate binding from those due to crystal packing, the protease complexes were aligned and an average α-carbon structure was calculated based on the direction of the bound peptide (Figure 3A). Since the protease dimers were not all oriented the same way within the unit cell, the influence of crystal packing on the average structure was reduced. Nevertheless, variations in temperature factors tend to correspond to the same regions of the protein between the various complexes. When a double difference plot (data not shown) was calculated between the two monomers of the average protease structure, two additional regions were seen to vary: residues 48–54 (the tips of the flaps) and 78–83 (the P1-loop). Unlike the other loops, these two regions have below average temperature factors and line the walls of the active site (Figure 3B) directly contacting the peptide substrates.

Conformation of Substrate Binding Pockets in the Protease

The protease conformation around the different substrate sites is best divided into four primary pockets: P1/P3, P2, P1′/P3′ and P2′, with two more variable pockets P4 and P4′ (Figure 3C). The P1/P3 and P1′/P3′ pockets are surrounded by Arg8, Leu23, Asp29, Lys45, Met46, Ile47, Gly48, Gly49, Ile50, Phe53, Pro81′, Val82′, and Ile84′. These include residues in the two most adaptable low-temperature factor regions of the protein: the tips of the flaps and the P1-loop. In the P1/P3 pocket, the side chains of Arg8, Ile47, Phe53, Val82′, and Ile84′ adopt a variety of conformations, but, around the P1′/P3′ site, all the side chains have one conformation in all six complexes except for Val82. In addition, the backbone of the protein rearranges between residues 45–50 and 78′–82′, much more around the P1/P3 pocket than the P1′/P3′ pocket. Both the position and conformation of Phe53, which makes van der Waals contacts with the substrates in these pockets, are highly variable and depend on the substrate, especially in the P1/P3 pocket. This increased variability in the P1/P3 pocket reflects the more variable conformations adopted by that region of the peptide.

In contrast, the conformations of the residues that make up the P2 and P2′ pockets are much more rigid. These residues include Asn(Asp)25, Gly27, Ala28, Asp29, and Asp30, whose positions and conformations remain unchanged among the different protease-peptide complexes and on either side of the peptide. The side chain of Asp30 makes only minor adjustments between the different complexes. The only variability in the P2/P2′ pockets are residues Ile47, Ile50, and Ile84′. They are located on the edge of the P1/P3 pockets. Thus, although the P2 and P2′ pockets accommodate a variety of different types of side chains, their structures remain rigid.

Conservation of Hydrogen Bonds within the Protease

A total of 158 hydrogen bonds are conserved among at least five of the six substrate complexes (140 in all complexes) within the HIV-1 protease dimer. In the 99-residue monomer, only 40 side chains are polar, and only 16 of those 40 form hydrogen bonds. These 16 residues form a total of 16 hydrogen bonds between side chains and 28 hydrogen bonds to backbone atoms within the dimer. Twelve of the 16 residues are also extremely well conserved, mutating less than 1% of the time in protease sequences from treated patients (Table 4A) [48]. In fact, only Gln58, Asp60, Thr74, and Asn88 mutate more than 1% of the time, and Asp60, Thr74, and Asn88 are hydrogen bonded to each other and usually mutate to maintain the hydrogen bonds. The rest of the hydrogen bonding involves fairly nonspecific backbone β interactions, which account for the vast majority of the secondary structure in the protease.

Conservation of Water Structure within the Protease

Not only are the protein-protein hydrogen bonds highly conserved among the various complexes, but the internal water molecules are also highly conserved among the five complexes determined to 2.0 Å resolution. Each of these complexes contains between 73 and 101 water molecules, as resolved in the electron density map (Table 2). Of these, at least 52 are conserved among four of the five complexes (Figure 4), as determined by criteria described in the Experimental Procedures section. As with substrate binding, all five complex structures were crystallized in the same space group, but two complexes with peptides RT-RH and RH-IN crystallized such that their orientation was opposite to the other three peptide substrates. This set of independent structures thereby decreases the probability that the water molecules we are observing are artifacts of crystallization.

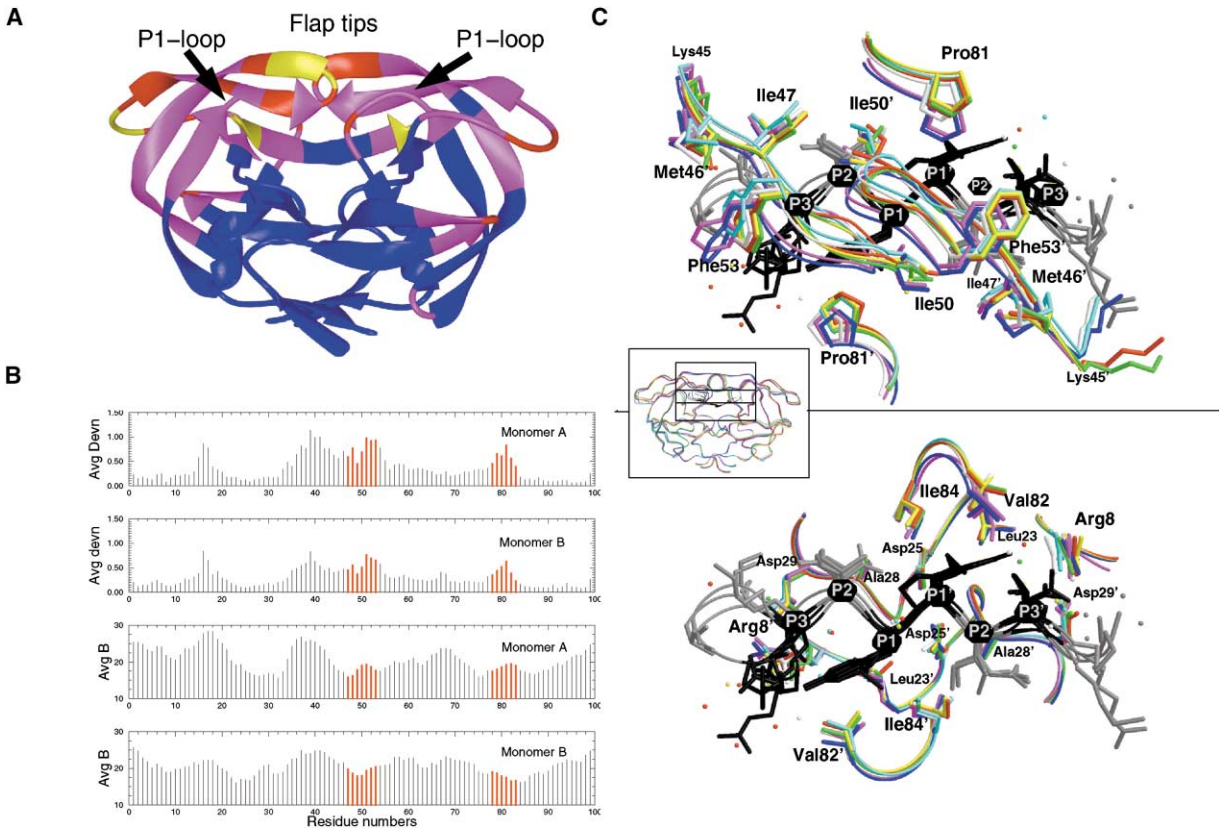

Figure 3. Deviations and Structural Adaptability between the Substrate Complexes

(A) Ribbon diagram of the inactive protease in complex with CA-p2. The figure is colored by the average deviation from the mean structure, after the structures were superimposed using the α-carbon atoms of residues 1–9 and 86–99 in both monomers. Blue, deviations between 0.0 and 0.3 Å; purple, deviations between 0.3 and 0.6 Å; red, deviations between 0.5 and 1.0 Å; yellow, deviations between 1.0 and 1.2 Å.

(B) The average deviations and B factors are plotted for both monomers. The tips of the flaps and P1-loop are in red, as these regions have relatively low B factors but deviate between the six complexes.

(C) Superposition of seven substrate complexes (two complexes with MA-CA from the two noncrystallographic dimers) for illustrating the relative mobility of the tips of the flaps and the P1-loop. MA-CA(1), magenta; MA-CA(2), blue; CA-p2, red; p2-NC, yellow; p1-p6, green; RT-RH, white; RH-IN, cyan; peptides, gray and black. The residues in the P1/P3 and P1′/P3′ are shown in black. The top panel shows the protease residues above the peptide plane, and the bottom panel shows the residues below the peptide plane. An α-carbon trace of these complexes is shown in the inset, with the boxes indicating the approximate regions of the two views.

Of the 52 conserved water molecules, 40 of them mediate hydrogen bonding between internal parts of the protease, 6 are on the surface of the flaps, and another 6 form only a single hydrogen bond with a surface residue. Only five water molecules are asymmetric, and all are found in the primed monomer. Among the five complexes, the root mean square deviation for the 40 water molecules involved in hydrogen bonding ranged between 0.2 and 1.0 Å from the mean position. Since many of these water molecules interact with the mobile and flexible regions of the enzyme, the rmsd indicates that they are structurally well conserved. Among the 40 water sites, 37 are highly symmetric: 18 in each monomer and 1 at the dimer 2-fold above the substrate. Five of the 37 symmetric water sites are found in the substrate recognition site (Figure 2) [31, 47]. The remaining water molecules are structural, maintaining the protease conformation and often bridging domains.

Water molecules that bridge regions of the active site are likely to be crucial to substrate recognition. As described above, the protease accommodates the substrate by adjusting the conformation of the P1-loop (78–83) and the tips of the flaps (48–54). These two regions are linked in both monomers through W4′ (Figure 4B), which bridges the carbonyl oxygen of Gly51 with the carbonyl oxygen of Pro79′. The relative position of this water is conserved even as these regions of the protease adjust in conformation, depending on the substrate peptide bound. In one monomer, the P1-loop is further stabilized by two additional water molecules: W5′ and W6′. Both of these waters hydrogen bond the internal surface of the P1′-loop to other parts of the protein. W5′ hydrogen bonds the backbone of Thr80′ to Thr89′, and W6′ hydrogen bonds the carbonyl oxygen of Val77′ to the side chains Glu35′ and Arg57′. These two waters are hydrogen bonded to the more variable P1′-loop.

The active site loop 21–34 is held in optimal position by a network of hydrogen bonds that mainly involves the backbone atoms of Thr74 and the side chain atoms of Asn83 and Asn88 at the base of the P1-loop. In addition to this network of hydrogen bonds, a total of 12 symmetric water molecules, 6 from each monomer, fur-

Table 4A. Hydrophilic Amino Acid Residues that Participate in Conserved Hydrogen Bonds, Their Rate of Mutations among Protease Genes from Patients Treated with Protease Inhibitors, and the Residues or Water Molecules to which They Bridge

		Hydrogen Bonds				Water Bridges					
		Monomer A		Monomer B		Monomer A			Monomer B		
Residue	Rate of Mutation (%)	Residues	Number[a]	Residues	Number[a]	Water	Residues	Number[a]	Water	Residues	Number[a]
Gln2	<1	**T96′, N98′**	5	**T96, N98**	5			0			0
Thr4	0		0		0	W20	T4, W6, K7	5	W20′	T4′, W6′, K7′	5
Arg8	<1	**D29′**	6	**D29**	6	W2	**D29′**, G27′, D29′	5	W2′	**D29**, G27, D29	5
D25N	0	**D25N′**, G27′, A28	6	**D25N**, G27, A28	6			0			0
Thr26	<1	T26′	6	T26	6			0			0
Asp29	0	**R87, R8′**	6	**R87′, R8**	6	W2, W3	**R8′, R87**, T26	5	W2′, W3′	**R8, R87′**, T26′	5
Asp30	5% N		0		0			3			3
Thr31	0	G86	6	G86′	6			0			0
Glu34	<1		0		0	W9	**N83**, K21, N83	4	W9′	**N83′**, K21′, N83′	2
Glu35	28% D	**R57**	1	**R57′**	5			2		**R57′**, V77′	4
Arg57	8% K	**E35**	1	**E35′**	5			0	W6′	**E35′**, V77′	4
Gln58	2% E		0		0			0			0
Asp60	6% E	T74	6	T74′	6			0			0
Gln61	2% E, 1% N		0		0	W13	**T74**, T74	4	W13′	**T74′**, T74′	4
Thr74	2% S, 1% A, 1% P	**D60, N88**	6	**D60′, N88′**	6	W13	**Q61**, Q61	4	W13′	**Q61′**, Q61′	5
Asn83	0	**N83**, K21	6	**N83′**, K21′	5	W8, W9	**E34**, K21	4	W8′, W9′	**E34′**, K21′	3
Arg87	<1	**D29**, L5′, W6′	6	**D29′**, L5, W6	6	W3	**D29**, T26	5	W3′	**D29′**, T26′	5
Asn88	2% D, 2% S	**T74**, T31	6	**T74′**, T31′	6	W10, W12	T31, T74	4	W10′, W12′	T31′, T74′	4
Thr91	0	R87, N88	6	R87′, N88′	5			0			0
Gln92	2% K, 1% N	I72	6	I72′	6			0			0
Thr96	0	**Q2′, N98**	6	**Q2, N98′**	6	W22	G94, N98′	4	W22′	G94′, N98	4
Asn98	0	**Q2′, T96′**	6	**Q2, T96**	6			0			0

[a]The number of structures (out of six) in which the hydrogen bond is found.
Bold indicates that the hydrogen bond is formed with another conserved side chain.

Table 4B. Patterns among Amino Acid Residue Types and Mutational Rates within the HIV-1 Protease Gene as Seen in 1260 Isolates from 742 Patients Treated with Protease Inhibitors [48]

Mutation Percentage	All Residues		Hydrophobic		Polar		Gly		Pro	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
<1	44	44.4	10	25.0	19	47.5	10	76.5	5	83.3
1–10	34	34.3	15	37.5	16	40.0	2	15.0	1	16.6
>10	21	21.2	15	37.5	5	12.5	1	7.5	0	0.0
Totals	99		40		40		13		6	

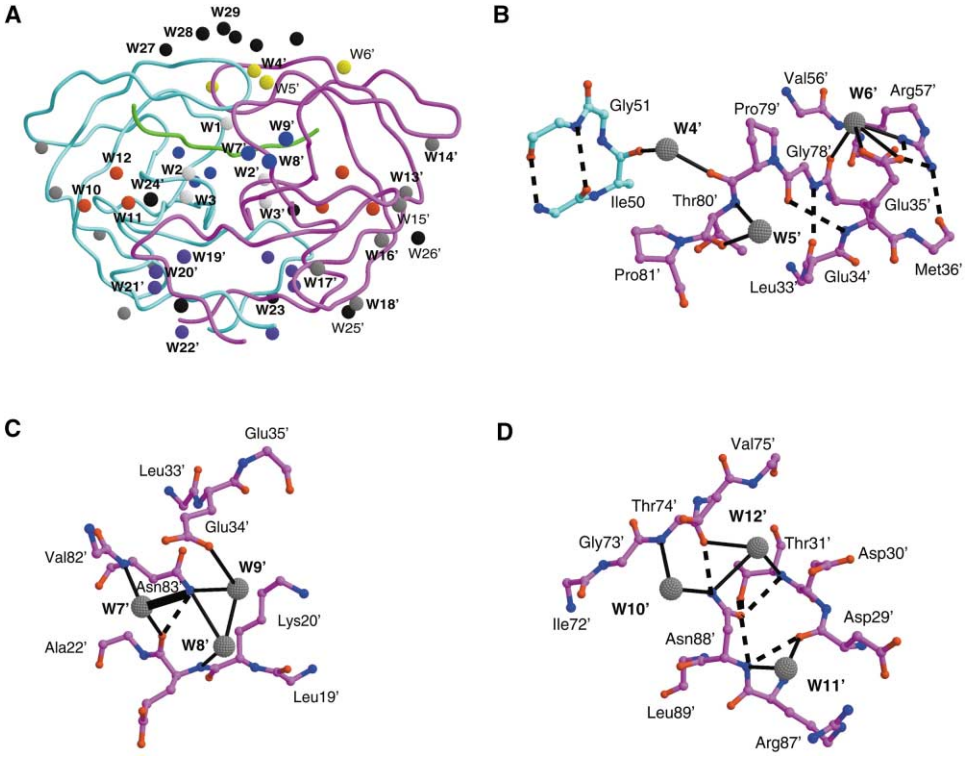

Figure 4. Conserved Water Structure

(A) Fifty-two water molecules that are conserved in at least four of the five structures determined to 2.0 Å resolution and form at least two protein-hydrogen bonds. Those waters found in both monomers are labeled in bold, and only one of the two is labeled to improve clarity. Equivalent clusters of waters are colored the same. Those waters around the substrate peptide are shown in white (Figure 2). Waters around the P1-loop and the tip of the flap are shown in yellow.

(B) On either side of the P1-loop, the conserved waters are shown in blue (C) and red (D). Hydrogen bonds between the protein and water are shown in bold, while hydrogen bonds between nearby protein atoms are shown in dashed lines.

ther stabilize the C-terminal and the N-terminal ends of the active site. Three water molecules (W7, W8, and W9) bridge Glu21 O and Glu34 OE2 with Asn83 ND2 (Figure 4C). Three more water molecules (W10, W11, and W12), located in the opposite side of the monomer, bridge Asp29 O, Thr31 N, and Thr31 O with Thr74 O and Asn88 ND2 (Figure 4D). These 12 water molecules occupy flat surfaces of the protease, and most of the residues with which these waters interact are already involved in a strong hydrogen bonding network. Perhaps these flat surfaces come within hydrogen bonding proximity to other protein domains, as HIV protease cleaves between larger protein domains that are either partially or fully folded, but not between isolated peptides.

The three most variable loops in HIV protease, 16–18, 35–41, and 65–70, are also connected to each other through a network of five symmetric (W13, W14, W16, W17, and W18) and one asymmetric (W15′) water molecule (Figure 4A). These water molecules thus mediate distant tertiary interactions between three separate loops, primarily through backbone atoms. The terminal domain, which is also the dimerization interface, contains an eight-water set (four of the waters are symmetric) that, like the domain itself, is very well ordered and highly conserved. The rmsds of these waters are only 0.2 Å. These waters (W19–W22) cap the ends of the small β sheet, which is made of the interdigitating termini of both monomers, in a manner consistent with that generally seen in protein structures [49–54]. The 12 remaining water molecules are conserved on the surface of the protease structure. Six of the 12 are associated with the flaps, which form an antiparallel β sheet; 3 molecules in each monomer form hydrogen bonds with the backbone atoms of Gly52, Phe53, and Lys55, mimicking an additional β strand and forming crystallographic symmetry-related contacts with other protease dimers. The other six waters are associated with only a single residue on the surface of the protein: two are found only in the primed monomer, and the other two are found in both monomers. The conserved water molecules thus play a crucial role in mediating inter- and intradomain interactions and largely affect the protein stability.

Discussion

What is necessary and sufficient for a particular peptide sequence to be specifically recognized as a substrate for HIV-1 protease? The structures of the six enzyme-substrate complexes we have determined lead us to conclude that the peptide must be able to form a relatively extended conformation to partake in the numerous backbone-backbone hydrogen bonds necessary for substrate binding. To achieve specificity, P1 and P3

(or the P3/P4 region) must contact each other on the unprimed side of the protease cleavage site to form a toroid. Rather than recognizing a particular amino acid sequence, the protease recognizes a shape. On the other side of the cleavage site, P1′ and P3′, there must be enough space for water to access the peptide. Whether this water is there for substrate recognition or for product release is still unclear. The conformational degrees of freedom must therefore allow the peptide to be sufficiently extended for the necessary hydrogen bonding to occur but not so extended that it precludes the formation of either a toroid or a cluster of water. That specificity of protein interaction is determined by shape rather than a particular set of hydrogen bonds has been noted before [55] but not in describing the specificity of HIV protease.

As only six of the ten substrate sequences in the Gag and Pol polyproteins were successfully crystallized, it is reasonable to ask whether the other four substrate sites would likely conform to this conserved shape. (Studies are currently underway to solve the rest of the complexes.) In fact, P1 and P3 in the remaining sequences tend to be larger than P1′ and P3′, suggesting that the trends described above will be borne out in protease complexes with these four remaining peptide substrates. The sequences of the four remaining peptides (Table 1) are highly hydrophobic and insoluble, probably contributing to their reluctance to crystallize under conditions that have been successful for most other protease complexes. Another possible contributing factor is that three of the four remaining peptides, trans frame peptide-protease (TF-PR), auto-proteolysis (AutoP), and protease-reverse transcriptase (PR-RT), correspond to substrate sites involving regions of HIV-1 protease [56] itself. Therefore, at least half of the sequences of these substrate peptides are identical with structured regions of the protease and may thereby disrupt crystal formation.

The hydrophobic toroid at P1–P3 may indicate a new avenue for future drug design. All the currently marketed drugs developed against HIV-1 protease bind predominantly to the protease between the P2–P2′ positions. Since these sites directly bridge the active site, even though the compounds are not themselves symmetric, they cover the same region in both monomers of the protease. When a drug-resistant mutation does occur, it can therefore affect the inhibitor's binding at two independent sites. The data presented in this paper suggest that this shortcoming of the current anti-HIV-1 drugs could be decreased or eliminated by designing an inhibitor that, rather than covering the region from P2–P2′, covers the region from P3–P1′. This design might produce a drug that interacts with the relatively invariant Asp29 and Arg8 and would likely produce inhibitors that are less susceptible to drug resistance. If such inhibitors mimic the toroid shape and extend across the catalytic aspartic acid regions, they could effectively compete with the natural substrates for HIV-1 protease.

The substrate recognition process has a further level of complexity. As with all HIV proteins, HIV protease undergoes frequent mutation due to the infidelity of its reverse transcriptase. HIV protease mutations in infected patients undergoing treatment are often selected if they confer some drug-resistant characteristics. Throughout all these changes, however, the enzyme must continue to function and recognize the various substrate sites within the Gag and Pol polyproteins. As the secondary structure of HIV-1 protease is primarily β sheet, the vast majority of the tertiary hydrogen bonds are between backbone atoms. A total of only 4 side chains are involved in protein-substrate hydrogen bonds, 17 side chains are involved in protein-protein hydrogen bonds, and 12 are involved in water-mediated protein-protein hydrogen bonds. A vast majority of these hydrogen bonds exist in both monomers of the protease and are highly conserved across the six substrate complexes that we have described. In addition, these hydrogen bond-forming side chains rarely mutate within the protease genes found in the treated patient population, thus confirming their importance to the enzyme (Tables 4A and 4B). Overall, 47.5 % of the hydrophilic residues within HIV protease mutate less than 1% of the time, and the rate of mutation is much less for those side chains forming hydrogen bonds, whereas only 25% of the hydrophobic residues are so highly conserved [48]. The most highly conserved residues are the structurally unique glycine and proline residues (Table 4B). By utilizing primarily a network of backbone and water-mediated hydrogen bonds, the protease can tolerate a high rate of mutation while still preserving its structural scaffold.

Biological Implications

Many of the most effective antiviral drugs used to treat AIDS to date target the HIV protease. All of these drugs were the successful results of structure-based drug design. The enzyme was a prime target, since it processes the Gag and Pol polyproteins at ten sites and allows for the maturation of the immature virion, thus enabling the spread of the virus. HIV-1 protease is a symmetric homodimer with the active site at the dimer interface, but its substrate sites are asymmetric and nonhomologous. Nevertheless, this enzyme is not a general protease, but its specificity is not easily determined from the sequences of the cleavage sites alone. In this study, we determined the crystal structures of six complexes of an inactive (D25N) HIV-1 protease bound to six decameric peptides corresponding to the cleavage sites within the Gag and Pol polyproteins. Analysis of these structures shows that the protease recognizes an asymmetric shape adopted by the substrate peptides rather than a particular amino acid sequence. The binding interface is approximately 1000 Å^2, on par with many other protein-protein molecular recognition surfaces. In binding the substrate, however, the protease adapts, and its symmetry is broken. Although substrate is recognized by a slightly altered enzyme conformation, the overall pattern of hydrogen bonds and water structure is conserved. The structural information gained from the asymmetric shape of the substrates could serve as a template for the design of a second generation of HAART drugs.

Experimental Procedures

Selection of the Peptide

In the crystallization of HIV-1 protease complexes, the bound ligand is often not uniquely oriented relative to the protease dimer in the

asymmetric unit. This lack of unique orientation results in structures of complexes actually being averaged with the ligand oriented 50% in one direction and 50% in the other. To prevent this occurrence, ten decameric peptides with sequences corresponding to the substrate sites within the Gag and Pol polyproteins were purchased from Quality Controlled Biochemicals Inc. (QCB), Hopkinton, MA (Table 1). Since the protease recognizes an octomeric site, binding these decameric peptides to the inactive protease should cause the peptide to protrude from either side of the active site, thereby differentiating one side of the homodimer from the other and hopefully influencing the orientation of the dimer within the crystal.

Protein Purification and Crystallization

Protein expression and purification were carried out as presented elsewhere [30, 33, 57]. The purified protein was concentrated to 2.5 mg/ml and equilibrated on ice for 30 min with a 5 M excess of the peptide. Crystals were obtained in hanging drops under more than one condition. The crystals used were grown with a reservoir solution of 126 mM phosphate buffer at pH 6.2, 63 mM sodium citrate, and 30%–35% ammonium sulfate [30, 58]. Each droplet contained 2.0–2.5 μl of the protease-peptide mixture and 2.0–2.5 μl of the reservoir solution. In all cases, the crystals were colorless, with maximum dimensions ranging from 0.1–0.2 mm.

Data Collection and Crystallographic Refinement

Intensity data were collected at room temperature using a R-AXIS IV image plate mounted on a Rigaku rotating anode X-ray generator. The raw data were reduced using DENZO and ScalePack [59] (Table 2).

Molecular replacement and crystallographic refinement were performed using Crystallography and NMR System (CNS) [60] in all cases. Our previously reported structure of the CA-p2 HIV-1 protease complex [30] was used as the model for solving the rest of the peptide complexes except for p1-p6, where 1MTR [61] was used. In all cases except MA-CA, isomorphous molecular replacement using only the protease atoms was employed for structure solution, as they all crystallized in the same space group, $P2_12_12_1$, with similar cell dimensions. In all cases, initial refinement was performed without the ligand, and the peptides were built into the difference density as the refinement progressed.

MA-CA crystallized in I222 space group with two dimers in the asymmetric unit. Molecular replacement was used to determine the three-dimensional structure of the MA-CA complex. The CA-p2 protease dimer was used as the search model for crossrotation. Only one solution was identified for which unambiguous translation solution was observed ($\theta_1 = 111.76$, $\theta_2 = 49.95$, $\theta_3 = 323.76$, $t_x = 40.07$, $t_y = 26.06$, $t_z = 29.44$). Once this solution was placed, another translation search, which yielded the solution for the second dimer as well, was performed ($\theta_1 = 45.62$, $\theta_2 = 79.91$, $\theta_3 = 259.90$, $t_x = 54.48$, $t_y = 18.51$, $t_z = 78.51$). Patterson correlation refinement [60] was used in the translation search for both dimers. Initial molecular replacement maps were unambiguous, and the difference density map ($F_o - F_c$) indicated the positions of the substrate in both dimers.

For all the complexes, refinement was carried out using a combination of simulated annealing, energy minimization, and individual B factor refinement with alternating cycles of model building using CHAIN [62]. Difference density maps using both $2F_o - F_c$ and $F_o - F_c$ as coefficients clearly indicated the positions of the residues that were changed to alanines during the initial refinements. Water molecules and acetate ions, which were present in the buffer during protein refolding, were added to the model as the refinement progressed. All the reflections were used for refinement, and no sigma cut off was applied. For the final models, the stereochemical parameters were checked using PROCHECK [63]. The final statistics are tabulated in Table 2. Figures were made using MIDASPlus [64] and GRASP [46].

Analysis of Water Structure

The structures of the five complexes, CA-p2, p2-NC, p1-p6, RT-RH, and RH-IN, were determined to 2.0 Å resolution and superimposed using the terminal domain and aligning the direction of the substrates. In order to include all relevant experimental water molecules, those waters related by crystal symmetry within 3.6 Å of any protein atom were generated before the superposition. These water molecules, enumerated for each structure, are tabulated in Table 2. The table also lists the number of side chain-water interactions and backbone-water interactions. Subsequent to the superposition, waters within 1.8 Å of each other were grouped. The distance criterion was relaxed up to 3.0 Å if the waters from different complexes were involved in a similar hydrogen bonding pattern. Thus, a water molecule was considered to be conserved between structures if it was within 1.8 Å of the cluster waters from other structures and formed at least one common hydrogen bond or if it was within 3.0 Å and most of the hydrogen bonds it formed were the same as others in the cluster. In addition, the water molecules were visually inspected using MidasPlus [64] to avoid discrepancies.

Acknowledgments

This research was supported by a grant from the American Cancer Society, RPG-99-213-01-MBC, and NIH GM64849. We would like to thank Charles Craik for the original D25N plasmid of HIV protease. We would also like to thank Jennifer Foulkes for help with the figures, Claire Baldwin for help with the editing, and Nancy King, Walter Scott, Kai Lin, Kendall Knight, and William Royer for many stimulating discussions.

Received: October 3, 2001
Revised: December 18, 2001
Accepted: December 18, 2001

References

1. Piot, P., and Coll Seck, A.M. (1997). International response to the HIV/AIDS epidemic: planning for success. Bull. World Health Organ. *79*, 106–112.
2. Carpenter, C.C., Fischl, M.A., Hammer, S.M., Hirsch, M.S., Jacobsen, D.J., Katzenstein, D.A., Montaner, J.S.G., Richman, D.D., Saag, M.S., Schooley, R.T., et al. (1998). Antiretroviral therapy for HIV infection in 1988: updated recommendations of the International AIDS Society—USA Panel. JAMA *280*, 78–86.
3. CDC. (1998). Report of the NIH panel to define principles of therapy of HIV infection. MMWR Morb. Mortal. Wkly. Rep. *47*, 1–41.
4. US Department of Health and Human Services and Henry J. Kaiser Family Foundation. (1998). Guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescents. MMWR Morb. Mortal. WKly. LA Rep. *47*, 43–82.
5. Luzuriaga, K.L., and Sullivan, J.L. (1998). Prevention and treatment of pediatric HIV infection. JAMA *280*, 17–18.
6. Hogg, R.S., O'Shaughnessy, M.V., Gataric, N., Yip, B., Craib, K., Schechter, M.T., and Montaner, J.S. (1997). Decline in deaths from AIDS due to new antiretrovirals. Lancet *349*, 1294.
7. Hogg, R.S., Heath, K.V., Yip, B., Craib, K.J.P., O'Shaughnessy, M.V., Schechter, M.T., and Montaner, J.S.G. (1998). Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. JAMA *279*, 450–454.
8. Palmer, S., Shafer, R., and Merigan, T.C. (1998). New drug combinations against highly drug-resistant HIV isolates in vitro. Antivir. Ther. *3*, 9–12.
9. Schouten, J.T. (1997). HIV drug resistance and the other causes of treatment failure. STEP Perspect *9*, 5–8.
10. Moyle, G.J. (1997). Viral resistance patterns selected by antiretroviral drugs and their potential to guide treatment choice. Expert Opin. Investig. Drugs *6*, 943–964.
11. Larder, B.A. (1995). Viral resistance and the selection of antiretroviral combinations. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. *10*, S28–S33.
12. Richman, D.D., and Staszewski, S. (1997). A Practical Guide to HIV Drug Resistance and Its Implications for Antiretroviral Treatment Strategies (London: International Medical Press).
13. Richman, D.D. (1997). Drug resistance and its implications in the management of HIV infection. Antivir. Ther. *2*, 41–58.
14. Schooley, R.T. (1997). Changing treatment strategies and goals. Antivir. Ther. *2*, 59–70.

15. Vella, S. (1995). HIV pathogenesis and treatment strategies. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. *10*, S20–S23.
16. McDonald, C.K., and Kuritkzkes, D.R. (1997). Human immunodeficiency virus type 1 protease inhibitors. Arch. Intern. Med. *157*, 951–959.
17. Flexner, C. (1998). HIV-protease inhibitors. New Engl. J. Med. *338*, 1281–1292.
18. Wlodawer, A., and Erickson, J.W. (1993). Structure-based inhibitors of HIV-1 protease. Annu. Rev. Biochem. *62*, 543–585.
19. Ji, J.P., and Loeb, L.A. (1992). Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochemistry *31*, 954–958.
20. Roberts, J.D., Bebenek, K., and Kunkel, T.A. (1988). The accuracy of reverse transcriptase from HIV-1. Science *242*, 1171–1173.
21. Roberts, J.D., Preston, B.D., Johnston, L.A., Soni, A., Loeb, L.A., and Kunkel, T.A. (1989). Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. Mol. Cell. Biol. *9*, 469–476.
22. Coffin, J.M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis and therapy. Science *257*, 483–489.
23. Coffin, J.M. (1996). HIV viral dynamics. AIDS *10*, S75–S84.
24. Hahn, B.H., Shaw, G.M., Taylor, M.E., Redfield, R.R., Markham, P.D., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C., Parks, E.S., and Parks, W.P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science *232*, 1548–1553.
25. Ho, D.D. (1997). Dynamics of HIV-1 replication in vivo. J. Clin. Invest. *99*, 2565–2567.
26. Henderson, L.E., Copeland, T.D., Sowder, R.C., Schultz, A.M., and Oroszlan, S. (1988). Human Retrovirus, Cancer and AIDS: Approaches to Prevention and Therapy (New York: Liss).
27. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M., and Sigal, I.S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. USA *85*, 4686–4690.
28. Debouck, C. (1992). The HIV-1 protease as a therapeutic target for AIDS. AIDS Res. Hum. Retroviruses *8*, 153–164.
29. Laco, G.S., Schalk-Hihi, C., Lubkowski, J., Morris, G., Zdanov, A., Olson, A., Elder, J.H., Wlodawer, A., and Gustchina, A. (1997). Crystal structures of the inactive D30N mutant of feline immunodeficiency virus protease complexed with a substrate and an inhibitor. Biochemistry *36*, 10696–10708.
30. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C.A. (2000). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. J. Mol. Biol. *301*, 1207–1220.
31. Mahalingam, B., Louis, J., Hung, J., Harrison, R., and Weber, I. (2001). Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. Proteins *43*, 455–464.
32. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., and Kent, S.B. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science *245*, 616–621.
33. Rosé, J.R., Babe, L.M., and Craik, C.S. (1995). Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. J. Virol. *69*, 2751–2758.
34. Babe, L.M., Rose, J., and Craik, C.S. (1995). Trans-dominant inhibitory human immunodeficiency virus type 1 protease monomers prevent protease activation and virion maturation. Proc. Natl. Acad. Sci. USA *92*, 10069–10073.
35. Mahalingam, B., Louis, J.M., Reed, C.C., Adomat, J.M., Krouse, J., Wang, Y.F., Harrison, R.W., and Weber, I.T. (1999). Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur. J. Biochem. *263*, 238–245.
36. Meek, T.D., Rodriguez, E.J., and Angeles, T.S. (1994). Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral proteases. Methods Enzymol. *241*, 127–156.
37. Polgár, L., Szeltner, Z., and Boros, I. (1994). Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease. Biochemistry *33*, 9351–9357.
38. Rasnick, D. (1997). Kinetics analysis of consecutive HIV proteolytic cleavages of the Gag-Pol polyprotein. J. Biol. Chem. *272*, 6348–6353.
39. Ridky, T.W., Kikonyogo, A., Leis, J., Gulnik, S., Copeland, T., Erickson, J., Wlodawer, A., Kurinov, I., Harrison, R.W., and Weber, I.T. (1998). Drug-resistant HIV-1 proteases identify enzyme residues important for substrate selection and catalytic rate. Biochemistry *37*, 13835–13845.
40. Silva, A.M., Cachau, R.E., Sham, H.L., and Erickson, J.W. (1996). Inhibition and catalytic mechanism of HIV-1 aspartic protease. J. Mol. Biol. *255*, 321–346.
41. Tozser, J., Zahuczky, G., Bagossi, P., Louis, J.M., Copeland, T.D., Oroszlan, S., Harrison, R.W., and Weber, I.T. (2000). Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur. J. Biochem. *267*, 6287–6295.
42. Weber, I.T., Wu, J., Adomat, J., Harrison, R.W., Kimmel, A.R., Wondrak, E.M., and Louis, J.M. (1997). Crystallographic analysis of human immunodeficiency virus 1 protease with an analog of the conserved CA-p2 substrate—interactions with frequently occurring glutamic acid residue at P2′ position of substrates. Eur. J. Biochem. *249*, 523–530.
43. Gustchina, A., Sansom, C., Prevost, M., Richelle, J., Wodak, S.Y., Wlodawer, A., and Weber, I.T. (1994). Energy calculations and analysis of HIV-1 protease-inhibitor crystal structures. Protein Eng. *7*, 309–317.
44. Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400.
45. Lawrence, M.G., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. J. Mol. Biol. *234*, 946–950.
46. Nicholls, A., Sharp, K., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.
47. Baldwin, E.T., Bhat, T.N., Gulnik, S., Liu, B., Topol, I.A., Kiso, Y., Mimoto, T., Mitsuya, H., and Erickson, J.W. (1995). Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophyenylnorstatine. Structure *3*, 581–590.
48. Shafer, R.W., Stevenson, D., and Chan, B. (1999). Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. *27*, 348–352.
49. Thanki, N., Thornton, J.M., and Goodfellow, J.M. (1988). Distribution of water around amino acid residues in proteins. J. Mol. Biol. *202*, 637–657.
50. Thanki, N., Umrania, Y., Thornton, J.M., and Goodfellow, J.M. (1991). Analysis of protein main-chain solvation as a function of secondary structure. J. Mol. Biol. *221*, 669–691.
51. Williams, M.A., Goodfellow, J.M., and Thornton, J.M. (1994). Buried waters and internal cavities in monomeric proteins. Protein Sci. *3*, 1224–1235.
52. Zhang, X.-J., and Matthews, B.W. (1994). Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. Protein Sci. *3*, 1031–1039.
53. Mattos, C., and Ringe, D. (1996). Locating and characterizing binding sites on proteins. Nat. Biotechnol. *14*, 595–599.
54. Nagendra, H.G., Sukumar, N., and Vijayan, M. (1998). Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration. Proteins *32*, 229–240.
55. Davis, A.M., and Teague, S.J. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. Angew Chem. Int. Ed. Engl. *38*, 736–749.
56. Wan, M., Takagi, M., Loh, B.N., Xu, X., and Imanaka, T. (1996). Autoprocessing: an essential step for the activation of HIV-1 protease. Biochem. J. *316*, 569–573.
57. Hui, J.O., Tomasselli, A.G., Reardon, I.M., Lull, J.M., Brunner, D.P., Tomich, C.S., and Heinrikson, R.L. (1993). Large scale purification and refolding of HIV-1 protease from *Escherichia coli* inclusion bodies. J. Protein Chem. *12*, 323–327.
58. Thanki, N., Rao, J.K., Foundling, S.I., Howe, W.J., Moon, J.B., Hui, J.O., Tomasselli, A.G., Heinrikson, R.L., Thaisrivongs, S., and Wlodawer, A. (1992). Crystal structure of a complex of HIV-1

protease with a dihydroxyethylene-containing inhibitor: comparisons with molecular modeling. Protein Sci. *1*, 1061–1072.

59. Otwinowski, Z. (1993). Oscillation data reduction program. In Data Collection and Processing, L. Sawyer, N. Isaacs, and S. Bailey, ed. (Warrington, United Kingdom: Daresbury Laboratory), pp. 56–62.
60. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.
61. Martin, J.L., Begun, J., Schindeler, A., Wickramasinghe, W.A., Alewood, D., Alewood, P.F., Bergman, D.A., Brinkworth, R.I., Abbenante, G., March, D.R., et al. (1999). Molecular recognition of macrocyclic peptidomimetic inhibitors by HIV-1 protease. Biochemistry *38*, 7978–7988.
62. Sack, J.S. (1988). CHAIN—a crystallographic modeling program. J. Mol. Graph. *6*, 224–225.
63. Laskowski, R.A., Mac Arthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
64. Ferrin, T.E., Huang, C.C., Jarvis, L.E., and Langridge, R. (1988). The MIDAS display system. J. Mol. Graph. *6*, 13–27.