Crystal Structures of the Inactive D30N Mutant of Feline Immunodeficiency Virus Protease Complexed with a Substrate and an Inhibitor[†,‡]

Gary S. Laco,[§,∥] Céline Schalk-Hihi,[§,⊥] Jacek Lubkowski,[⊥] Garrett Morris,[∥] Alexander Zdanov,[⊥] Arthur Olson,[∥] John H. Elder,[∥] Alexander Wlodawer,[⊥] and Alla Gustchina*[,⊥]

Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21702, and Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received March 31, 1997; Revised Manuscript Received June 25, 1997[⊗]

ABSTRACT: Crystal structures of complexes of a D30N mutant of feline immunodeficiency virus protease (FIV PR) complexed with a statine-based inhibitor (LP-149), as well as with a substrate based on a modification of this inhibitor (LP-149S), have been solved and refined at resolutions of 2.0 and 1.85 Å, respectively. Both the inhibitor and the substrate are bound in the active site of the mutant protease in a similar mode, which also resembles the mode of binding of LP-149 to the native protease. The carbonyl oxygen of the scissile bond in the substrate is not hydrated and is located within the distance of a hydrogen bond to an amido nitrogen atom from one of the two asparagines in the active site of the enzyme. The nitrogen atom of the scissile bond is 3.25 Å from the conserved water molecule (Wat301). A model of a tetrahedral intermediate bound to the active site of the native enzyme was built by considering the interactions observed in all three crystal structures of FIV PR. Molecular dynamics simulations of this model bound to native wild-type FIV PR were carried out, to investigate the final stages of the catalytic mechanism of aspartic proteases.

Retroviruses encode an essential protease (PR)[1] responsible for the cleavage of polyprotein precursors into individual proteins, leading to proper virion assembly and maturation. On the basis of the presence of a sequence Asp-Ser/Thr-Gly in the active sites of retroviral proteases (*1*) and their inhibition *in vitro* by pepstatin (*2*), these enzymes have been classified as members of the aspartic protease family (*3*−*7*). Crystal structures have been determined for the proteases from Rous sarcoma virus (RSV) (*8*), human immunodeficiency virus type 1 (HIV-1) and its mutants (*9, 10*), HIV-2 (*11*), simian immunodeficiency virus (*12*), equine infectious anemia virus (*13*), and feline immunodeficiency virus (FIV) (*14*). Retroviral proteases are active as homodimers with a single active site formed by two catalytic aspartates, derived from the two subunits, which are within hydrogen-bonding distance. Mutations of the HIV-1 PR active site Asp25 into Asn (*15, 16*), Thr (*3*), or Ala (*4, 17, 18*) led to an inactive enzyme. Similarly, RSV PR was inactivated by mutation of its active site Asp to Ile (*19*). In apoenzymes of both the retroviral and nonviral subclasses of the family of aspartic proteases, a water molecule is bound to both catalytic aspartates (*10, 20*). It has been proposed that this polarized water molecule is involved in a nucleophilic attack on the carbonyl carbon of the scissile peptide bond (*21*−*24*).

Retroviral proteases have become a major target for the rational design of drugs against AIDS (*25*), and hence a large number of crystal structures of their inhibitor complexes have been determined by X-ray crystallography (*26, 27*). These analyses show that peptidomimetic inhibitors usually bind to the enzyme in an extended conformation, making hydrogen bonds and hydrophobic contacts within the binding pockets. It is expected that a substrate bound to the protease would be involved in similar interactions. However, since a substrate would be rapidly processed, no structural data on the complexes of the native enzyme with the substrate are currently available. Such data, however, would be crucial for studies of the specificity and of the catalytic mechanism of aspartic proteases.

We expressed a mutant of FIV PR in which the catalytic Asp30 was mutated to Asn, leading to inactive protease [designated FIV PR(D30N)]. To investigate the extent of perturbation of the active site of FIV PR caused by this mutation, we determined at 2.0 Å resolution the crystal structure of FIV PR(D30N) in a complex with LP-149, a statine-based inhibitor patterned after the junction between the capsid and nucleocapsid in the HIV-1 Gag polyprotein. We then compared this structure with the structure of the wild-type enzyme, FIV PR(wt), complexed with the same inhibitor. This comparison revealed that the mode of binding of LP-149 to FIV PR(D30N) is similar to that of LP-149 to FIV PR(wt), making the mutant a valuable model in which to study the interactions of substrates with FIV PR, and retroviral proteases in general.

[†] Dedicated to Dr. David Davies on the occasion of his 70th birthday. This work was supported by the National Cancer Institute, DHHS, under contract with ABL, and by NIH Grant P01GM48870. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank (filenames 2fiv and 3fiv).

* Author to whom correspondence should be addressed. Tel: 301-846-5031. Fax: 301-846-5991. E-mail: alla@orion.ncifcrf.gov.

[§] The first two authors contributed equally to this work.

[∥] The Scripps Research Institute.

[⊥] NCI-Frederick Cancer Research and Development Center.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1997.

[1] Abbreviations: PR, protease; RSV, Rous sarcoma virus; HIV, human immunodeficiency virus; FIV, feline immunodeficiency virus.

S0006-2960(97)00743-5 CCC: $14.00

We carried out several attempts to grow crystals of the complexes of FIV PR(D30N) with a number of different substrates, without success. To provide a crystallizable substrate, we synthesized a hexapeptide based on the primary structure of the inhibitor LP-149, incorporating in place of the statine a Leu-Ala junction with an unmodified peptide bond (thus restoring the sequence of the capsid/nucleocapsid junction in HIV-1 Gag). A crystal structure of the complex of FIV PR(D30N) with this substrate was solved at 1.85 Å resolution. We built a model of a tetrahedral intermediate bound to the active site of the native enzyme by considering the interactions observed in all three crystal structures of FIV PR. We subjected this model to molecular dynamics simulations to investigate their effect on the ligand/enzyme interactions.

MATERIALS AND METHODS

Constructs, Expression, and Purification. For pT7-D30N/PR, the feline immunodeficiency virus 34TF10 infectious molecular clone (FIV34TF10) was used as the template in a polymerase chain reaction (PCR) with the primer 5′PR/D30N (5′CCTATAAAATTTTTATTAAACACAGGA-GCA-GATATAACAATT3′, nucleotides 2051−2093), which mutated the FIV PR aspartic acid codon no. 25 to asparagine, and primer 3′PR/*Hin*dIII (5′ATCAGAAAGCTTTTACAT-TACTAACCTG-ATATTAAATTT3′, complementary to nucleotides 2306−2344), which added a stop codon and *Hin*dIII restriction site 3′ to the determined C-terminal methionine codon no. 116 of the PR open reading frame (ORF) (*28*), after incorporation into the PCR product. The resulting ∼300 bp PCR product was used as a negative strand primer ("megaprimer") in a second PCR reaction with primer 5′PR/*Nde*I (5′ACTATTGGACATATGGCATATAATAAAGTA-GGTACTACTAC3′, nucleotides 1964−2004), which adds an *Nde*I restriction site, initiation methionine, and alanine codon 5′ to the determined 5′ tyrosine codon of FIV PR, when incorporated into the PCR product (*28*). The resulting PCR product was digested with *Nde*I and *Hin*dIII and ligated into pT7-7 (*29*), which had been digested with *Nde*I and *Hin*dIII, resulting in pT7-D30N/PR. The mutated PR ORF was sequenced to verify authenticity of the construct.

pT7-D30N/PR was transformed into *Escherichia coli* strain BL21(DE3) pLysS (*30*) which contains the T7 RNA polymerase gene under the control of the *lac* promoter, as well as a plasmid-encoded lysozyme gene (pLysS) under the control of a constitutive promoter. Cultures were induced at $OD_{600} = 0.5$ with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 5 h; cells were pelleted and resuspended in 10 mM Tris, pH 8.0, and 5 mM ethylenediaminetetraacetic acid (EDTA) and frozen. The cell solution was then thawed in a 20 °C water bath, resulting in cell lysis. The released cellular DNA was sheared in a Waring blender for 10 s, and Nonidet P-40 (NP-40) was added to 0.1% followed by the addition of urea to 2 M. The solution was then centrifuged at 5000*g* to pellet the PR inclusion bodies. The inclusion bodies were washed extensively with deionized H_2O, pelleted again, and then solubilized in buffer A (8 M urea, 10 mM Tris, pH 8.0, 5 mM EDTA).

The soluble protease solution was loaded onto a Q-Sepharose Fast Flow column (Pharmacia) equilibrated in buffer A. The protein fraction that did not bind to the column was collected and brought to 20 mM sodium acetate, pH 5.0. The protein was then loaded onto a Resources S-column (Pharmacia) equilibrated in buffer B (20 mM sodium acetate, pH 5.0, 8 M urea, 5 mM EDTA). The protease that was bound to the column was eluted with a 0−1 M NaCl gradient in 8 M urea. The Resource S-bound protease fraction was diluted in 8 M urea, brought to 20 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS), pH 10.5, and loaded onto a Resource Q column (Pharmacia) equilibrated in buffer C (20 mM CAPS, pH 10.5, 8 M urea, 5 mM EDTA). The protease that did not bind to the column was collected.

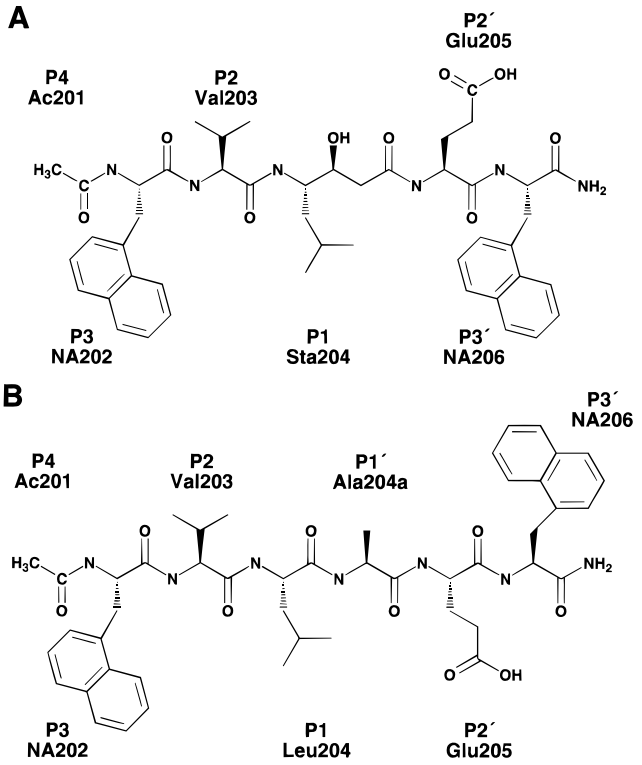

FIGURE 1: Ligands used in these studies. (A) Chemical formula of the inhibitor LP-149 (Ac-NA-Val-Sta-Glu-NA-NH_2, where NA is naphthylalanine and Sta is statine). (B) The substrate Ac-NA-Val-Leu-Ala-Glu-NA-NH_2 used for cocrystallization with the D30N mutant of FIV PR.

The Resources Q pH 10.5 protease fraction was brought to 0.1 M dithiothreitol (DTT) and 0.15 M NaCl and dialyzed against 25 mM $NaPO_4$, pH 7.4, 150 mM NaCl, 5 mM EDTA, and 2 mM DTT at 5 °C, using 6−8 kDa MWKO membrane (Spectropore) (*14*). The renatured protease solution was brought to 5% 2-propanol and concentrated at 5 °C to 1 mg/mL using an Amicon stirred cell concentrator with a YM 10 kDa cutoff membrane. The protease sample was either stored at −70 °C or brought to 50% glycerol and stored at −20 °C. The FIV PR and FIV PR(D30N) were tested for activity with a fluorogenic peptide substrate (M. C. Fitzgerald, personal communication). The FIV PR(D30N) had no detectable PR activity on the substrate even after incubation for 4 h at 37 °C (G. S. Laco, unpublished).

Preparation of the Ligands. The statine-based inhibitor LP-149 (Figure 1A), prepared at Lilly Research Laboratories, was a gift of K. Hui (unpublished results). The sequence was based on the HIV-1 capsid/nucleocapsid junction Arg-Val-Leu/Ala-Glu-Ala (/ indicates the cleavage site). Synthesis of the substrate LP-149S, in which Leu-Ala replaced the statine residue of the inhibitor (Figure 1B), was conducted as follows. Manual solid-phase synthesis was carried out using Boc chemistry *in situ* neutralization protocols as

Table 1: Summary of Data Collection and Refinement Statistics for the Complexes of FIV PR(D30N)

		LP-149	LP-149S
unit cell parameters	*a*	50.49 Å	50.89 Å
	c	74.29 Å	74.53 Å
no. of reflections	total measured	63643	40465
	unique $P3_121$	8646	
	unique $P3_1$	14984	17411
completeness	$P3_121$	95.2% (40−1.9 Å)	
		87.2% (1.93−1.9 Å)	
	$P3_1$	89.5% (40−1.9 Å)	89.5% (20−1.8 Å)
		74.2% (1.93−1.9 Å)	68.5% (1.83 −1.8 Å)
R_{sym}	$P3_121$	0.089	
	$P3_1$	0.080	0.091
R_{merge}[a]	$P3_1$ vs $P3_1$	0.015	
resolution limits in refinement		10.0−2.0 Å	8.0−1.85 Å
no. of reflections used in refinement		12589	15187
R-factor (final model)	$P3_1$	0.162 (0.159)[b]	0.172
	$P3_121$	0.159	
rmsd (monomer vs monomer, all atoms)		0.002 Å (max 0.009 Å)	0.007 Å (max 0.018 Å)
rmsd from ideality[c]	bonds	0.013 Å	0.014 Å
	bond angles	1.69°	1.82°
	torsion angles	27.9°	27.6°
	improper angles	1.57°	1.52°
average *B*-factor	all non-H atoms	25.0 Å^2	28.0 Å^2
PDB designation		2fiv	3fiv

[a] One X-ray data set is the result of data reduction in space group $P3_1$ whereas the second data set was generated by symmetry operations from X-ray data in space group $P3_121$. [b] The values shown in parentheses correspond to the data set generated by symmetry operations from X-ray data in space group $P3_121$. [c] Ideal targets for refinement were those published by Engh and Huber (*65*).

described (*31*). Naphthylalanine (NA) was substituted for the P3 and P3′ position amino acids, and the peptide was acetylated (Ac) at the N-terminus with 20% acetic anhydride/5% diisopropylamine (DIEA) in dimethylformamide (DMF) for 10 min. The peptide/resin was deprotected, with the peptide cleaved from the resin with hydrogen fluoride (HF) followed by washing and lyophilization as described (*31*). This procedure resulted in the peptide Ac-NA-Val-Leu/Ala-Glu-NA-NH_2 (LP-149S peptide).

Substrate Assay Conditions. The LP-149S peptide (115 μM) was incubated with FIV PR (15 μM) in 50 mM $NaPO_4$, pH 7.2, 100 mM NaCl, and 2% glycerol at 24 °C for 72 h. The mixture was then placed into a glass vial to which FIV PR bound irreversibly, thus removing it from the reaction. The solution was injected into a reverse-phase HPLC column (C_4, Vydac) equilibrated in H_2O and 0.1% trifluoroacetic acid (TFA). A 0−67% acetonitrile gradient in 0.1% TFA was run, with peak fractions collected and analyzed by mass spectrometry.

Mass Spectroscopy. Atmospheric pressure ionization mass spectrometry [API/MS Perkin/Elmer/Sciex; (*32*)] was used to determine the precise molecular mass of the PRs and peptides. The purified protein was dialyzed against H_2O to remove salts and then diluted in 40% acetonitrile and 0.1% TFA. The sample was then sprayed from a metal syringe (+5000 V) into the orifice of the mass analyzer with the assistance of concentrically applied nebulizer gas, where the ion spray was directed into the quadrupole filter to determine the *m/z* ratio of the proteins, which are in various charge states. MacBioSpec algorithms (Sciex) were used to convert the family of ion peaks, which resulted from the protein being in various charge states, to an accurate molecular mass (± 1 Da) and into a single peak representing the uncharged protein. The calculated molecular mass of the D30N/PR (13304 Da) was found to be within the expected error (±1 Da) of the determined molecular mass (13305 Da).

Crystallization and Data Collection. Before crystallization trials, the protein was subjected to gel filtration on Superdex-75 (Pharmacia) in 50 mM sodium/potassium phosphate buffer, pH 7.4, containing 1 mM EDTA, 50 mM 2-mercaptoethanol, 150 mM NaCl, 5% glycerol, and 5% 2-propanol, as described previously (*14*). Crystallization of the complexes of the D30N mutant of FIV PR with both the inhibitor and the substrate was carried out under similar conditions at 4 °C, using the hanging-drop vapor diffusion method as follows: 2.5 μL of FIV PR(D30N) complexed with LP-149 at 7 mg/mL (1:4 molar ratio), with substrate at 5 mg/mL (1:4 molar ratio) in 50 mM imidazole-hydrochloride, pH 7.0, containing 1mM EDTA and 1 mM DTT were mixed with an equal volume of 2 M ammonium sulfate and 0.1 M sodium acetate, pH 4.6 (Hampton Crystal Screen, solution no. 47). In both cases, crystals appeared within a few days and reached the size of 0.2 mm × 0.2 mm × 0.4 mm in 1 week.

X-ray diffraction data were collected at room temperature using a MAR Research 300 mm image plate detector mounted on a Rigaku RU-200 rotating anode generator operated at 50 kV and 100 mA. Each data set was collected from a single crystal and processed using the program DENZO (*33*). Crystals belong to the trigonal system, space group $P3_121$, $a = b = 50.49$ Å, $c = 74.29$ Å, with half the dimeric protease molecule in the asymmetric unit, and are isomorphous with the crystals of the wild-type enzyme. Since for technical reasons the last steps of the refinement were carried out in the lower symmetry space group $P3_1$ (see below), the X-ray data were also reduced in this lower symmetry space group. The statistics of the data are summarized in Table 1.

Molecular Dynamics Calculations. Molecular dynamics simulations of the model of the tetrahedral intermediate bound to the active site of FIV PR(wt) were performed using the Discover module of InsightII (Biosym Technologies) running on a Silicon Graphics Power Challenge Supercomputer Server. The force field used was CFF91 (*34, 35*), a so-called "class II" force field, which is based on the shape of *ab initio* potential energy surfaces. Molecular dynamics

were performed on the transition-state complex using a step size of 1 fs and a trajectory output frequency every 100 fs; a total of 5 ps of equilibration was followed by 50 ps of simulation. A simulation temperature of 500 K was used. The entire transition state, water-301, flap residues A58−A61 and B58−B61, and residues adjacent to and including the catalytic aspartates A29−A31 and B29−B31 were allowed to move.

The initial coordinates were modeled as follows. All hydrogens were added to FIV PR; ionizable side chains were protonated with the pH set to 4.0. Only one of the catalytic aspartic acid side chains (AspB30) was protonated: in the first model at OD1 (the oxygen atom which is closer to AspA30) and in the second model at OD2. The second catalytic aspartate (AspA30) was negatively charged in all cases.

The transition state was modeled as the *gem*-diol product of nucleophilic attack by water upon the carbonyl carbon of the scissile peptide bond. The initial conformation of the transition state was modeled by carefully superimposing the hydrogen-bonding heteroatoms along the backbone, as well as the *gem*-diol hydroxyls of the hydrated peptide on the statyl hydroxyl of LP-149 and the carbonyl oxygen of the substrate. The location and interactions of the *gem*-diol moiety with the aspartates in the structure of a pyrrolidine-containing α-keto amide inhibitor, bound to HIV-1 PR, were also taken into account (*36*). The side chains at P2, P3, P2′, and P3′ were modeled according to the positions of the corresponding atoms of the structures of the inhibitor LP-149 and the substrate complexed with FIV PR.

RESULTS

Cleavage of LP-149S Peptide by FIV PR. The LP-149S peptide (Figure 2A) was specifically cleaved by FIV PR-(wt). The two cleavage products were determined to be within ±1 Da of the calculated masses for the predicted cleavage products (Figure 2B). This result confirms that LP-149S is a substrate for FIV PR. However, it should be noted that the LP-149S cleavage rate was slow compared with longer substrates (*31*) and was detected only when high concentrations of FIV PR, and extended incubation times, were used (0.4 mg/mL for 72 h vs standard assay at 0.004 mg/mL for 5 min).

Structure Solution and Refinement. The crystallographic refinement of the complexes of FIV PR(D30N) with LP-149 and with the synthetic substrate LP-149S was carried out with the same protocol. A major complication to the refinement was the 2-fold disorder of the ligand resulting from the presence of only half of the protease dimer in the asymmetric unit. The electron density around Asn30 in the inhibitor complex clearly indicated that this residue also adopted more than one conformation. Furthermore, it was evident that when only a single conformer was modeled, the distance between these asparagines from both monomers was unusually short (2.3 Å). This finding indicated that the crystallographic symmetry of space group $P3_121$ may be violated by at least a few active site residues of FIV PR-(D30N). To overcome these problems, we refined the structure using X-PLOR in the space group $P3_1$, with residues 29−32 modeled in two alternative conformations. Protease molecules with each of these conformations were denoted A and A*, respectively, while their dimer mates were B and B*. On the basis of careful analysis of the interactions between the enzyme and the ligand molecule, and of the electron density, the unique active site could then be defined as consisting of AsnA30*, AsnB30, and the inhibitor molecule A201−A206. The asymmetric part of the $P3_1$ unit cell thus contained the complete dimer of the enzyme with two structurally identical molecules of the inhibitor related by 2-fold symmetry. It needs to be stressed that the 2-fold symmetry of the unique dimer (A-B* or A*-B) is violated by the bound inhibitor molecule and by the residues adjacent and including Asn30. Since the crystal contains an equal number of both dimers, the diffraction data exhibit higher symmetry due to statistical averaging (Table 1). In such a case, it is a matter of choice whether to complete the refinement in the higher symmetry space group or in the lower symmetry with very strong noncrystallographic symmetry restraints. Very low monomer−monomer deviations listed in Table 1 indicate that the results of both procedures would be the same.

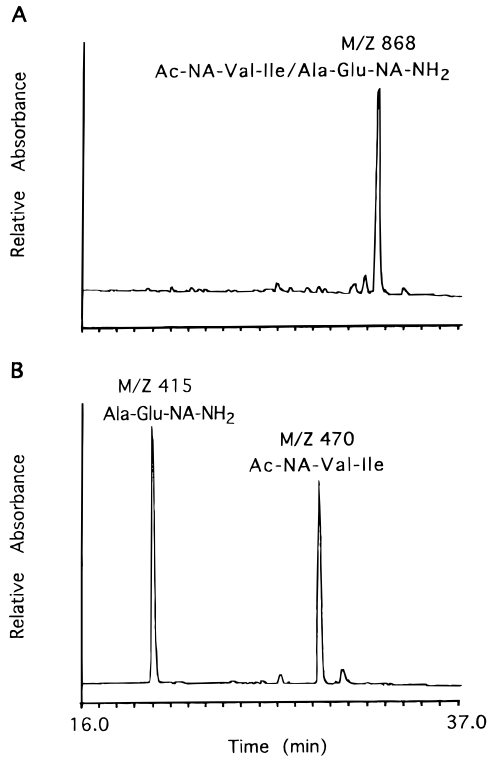

FIGURE 2: Chromatography of LP-149S and the results of enzymatic assays. (A) Analysis of the peptide by reverse-phase HPLC (shown), with mass determined by mass spectrometry (see Materials and Methods). (B) Specific cleavage of the LP-149S peptide by FIV PR. Cleavage products were purified by reverse-phase HPLC and analyzed by mass spectrometry. *m/z*, = mass-to-charge ratio.

Because the crystals of the FIV PR(D30N) complexes were isomorphous with those of the FIV PR(wt)/LP-149 complex (*14*), the crystallographic coordinates of FIV PR(wt) dimer (*14*) were used for initial refinement. The rigid-body refinement of the protein alone in the FIV PR(D30N)/LP-149 complex using X-PLOR (*37*), at the resolution range 10−2.0 Å, lowered the *R*-factor to 0.267. This was then followed by positional refinement and manual corrections made with FRODO (*38*), after which the inhibitor was built into the enzyme's active site. Initially, refinement of the complex was performed using PROLSQ (*39, 40*) at the resolution range 10−2.0 Å. The electron density for residues

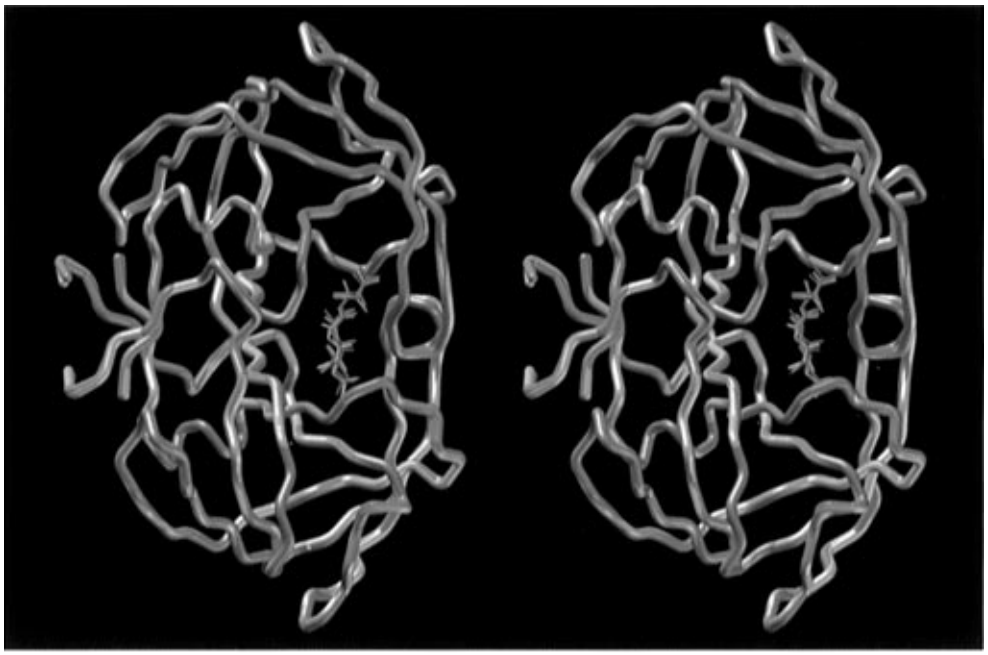

FIGURE 3: Stereoview showing the superposition of the main chains of the crystal structure of FIV PR(wt) complexed to LP-149 (blue) and the crystal structures of the D30N mutant of FIV PR complexed to LP-149 (orange) and the substrate (pink).

59−61 (the tip of the flap) and for the side chains of IleA-(B)37 indicated two conformations for each of these fragments, which were subsequently modeled. A very good agreement between the structure and the electron density, after a few rounds of refinement, validated the presence of alternative conformations of both fragments. The subsequent cycles of refinement improved both the overall geometrical and stereochemical quality of the model, as well as its agreement with the electron density. The refined model of the FIV PR(D30N)/LP-149 complex contained 1871 protein atoms, 122 inhibitor atoms, 144 water molecules, and 2 sulfate anions. The final value of the *R*-factor was 0.162 for 12 589 reflections at the resolution range 10.0−2.0 Å [with $|F| \geq 3.0\sigma(F)$].

The electron density maps for the asparagines in the substrate structure did not indicate any obvious disorder for their side chains. Therefore, we built two models for this structure, which differed only by the conformations of asparagine residues in the active site of the enzyme. The first model of the active site had two pairs of asparagine residues, related by crystallographic symmetry operation, with an asymmetric assignment of the atoms in each pair [similar to their arrangement in the active site of FIV PR-(D30N) with bound LP-149]. An alternative model of the active site had only one conformation for each asparagine with a symmetric assignment of the atoms in the asparagine's side chains. In this model, both amide groups had their oxygens adjacent to each other, and the NH_2 groups were located at the opposite sides of the asparagine pairs. The final structure of the FIV PR(D30N)/substrate complex was refined as above at the resolution range 10.0−1.85 Å [space group $P3_1$; 15 187 unique reflections with $|F| \geq 3.0\sigma(F)$; *R*-factor of 0.172 for both models]. This structure contained 1836 protein atoms, 126 substrate atoms, 148 water molecules, and 2 sulfate anions. The resulting geometric and stereochemical descriptions of the structures of both complexes, as well as the Protein Data Bank designation of the deposited coordinates, are shown in Table 1.

Structure Description. The crystal structures of the complexes of FIV PR(D30N) with the inhibitor LP-149 and the substrate LP-149S showed that, as expected, the single active site mutation of the catalytic Asp30 to Asn did not affect the overall structure of the protein (Figure 3). The root-mean-square (rms) deviation between the Cα atoms of FIV PR(wt) and FIV PR(D30N), both complexed to LP-149, was 0.175 Å for 234 pairs; between FIV PR(wt) complexed to LP-149 and FIV PR(D30N) complexed to LP-149S, 0.182 Å for 229 pairs; and between the complexes of FIV PR(D30N) with LP-149S and LP-149, 0.179 Å. For the backbone atoms and all atoms, the respective values were 0.166 Å (919 pairs) and 0.188 Å (1603 pairs), 0.182 Å (906 pairs) and 0.200 Å (1514 pairs), and 0.175 Å (887 pairs) and 0.190 Å (1576 pairs).

The monomer of FIV PR(D30N) contains 117 residues, including an extra N-terminal alanine (numbered 0). To simplify the numbering of the dimer, we used letters A and B for the residues from the first and second monomers, respectively; for the residues with multiple conformations of their side chains, we denoted their alternate conformations by asterisks. The ligand molecules are bound in the active site of the enzyme, interacting with both asparagines−AsnA30 and AsnB30. The flaps formed by residues A49−A69 and B49−B69 fold over the ligand in the active site of the D30N mutant. Their position and conformation are similar to those observed in the wild-type structure (*14*). Residues A59−A61 (and B59−B61), which form the tip of the flaps, adopt two conformations, similar to the FIV PR(wt)/LP-149 complex (*14*), with equal occupancy for each conformation. The flap structure is stabilized by a hydrogen bond between the carbonyl oxygen atom of ValA59 and the amide nitrogen atom of GlyB60*. All three complexes were crystallized in the same space group, and the symmetry of the crystals is the reason for the 2-fold disorder of the ligand, which is bound in the active site of the enzyme with two opposite directions of the main chain with equal occupancy (0.5). The conserved network of hydrogen bonds between the backbone

Table 2: Hydrogen Bond Distances between the Interacting Pairs of Atoms in the Active Site of FIV PR(wt) Complexed with Inhibitor LP-149 Compared with Those in the Active Sites of FIV PR(D30N) Complexed with LP-149 and with Substrate LP-149S

	FIV PR(wt)/LP-149			FIV PR(D30N)/LP-149			FIV PR(D30N)/LP-149S		
ligand residue	hydrogen bond	residue	distance Å	hydrogen bond	residue	distance Å	hydrogen bond	residue	distance Å
P4	CO−NH	Ile57	3.0	CO−NH	Ile57	2.8	CO−NH	Ile57	3.2
P3	NH−OD2	Asp34	2.6	NH−OD2	Asp34	2.9	NH−OD2	Asp34	2.8
	CO−NH	Asp34	2.5	CO−NH	Asp34	2.9	CO−NH	Asp34	2.7
P2	NH−CO	Ile57	3.2	NH−CO	Ile57	2.9	NH−CO	Ile57	3.0
	CO−OH	Wat301	2.7	CO−OH	Wat301	2.8	CO−OH	Wat301	2.8
P1	NH−CO	Gly32	3.1	NH−CO	Gly32	3.1	NH−CO	Gly32	3.2
	OH−OD2	AspA30	2.6	BOH−ND2	AsnA30	2.7	CO−ND2	AsnA30	2.6
	OH−OD1	AspA30	2.5	BOH−OD1	AsnA30	2.8			
	OH−OD2	AspB30	3.1	BOH−ND2	AsnB30′	2.8			
P1′							NH−OH	Wat301	3.2
	CO−OH	Wat301	2.6	CO−OH	Wat301	2.7	CO−OH	Wat301	2.8
P2′	NH−CO	Gly32	3.2	NH−CO	Gly32	3.2	NH−CO	Gly32	3.0
	OE2−NH	Ile35	3.0	OE2−NH	Ile35	3.3	OE2−NH	Ile35	3.3
	OE2−NH	Asp34	2.8	OE2−NH	Asp34	3.1	OE2−NH	Asp34	2.9
	CO−OH	Wat338	3.2	CO−OH	Wat302	3.1	CO−OH	Wat321	2.9
P3′	NH−CO	Ile57	2.5	NH−CO	Ile57	2.6	NH−CO	Ile57	2.7
	CO−NH	Ile57	2.6	CO−NH	Ile57	2.8	CO−NH	Ile57	2.8
P4′	NH2−OD2	Asp34	3.0	NH2−OD2	Asp34	2.9	NH2−OD2	Asp34	3.4

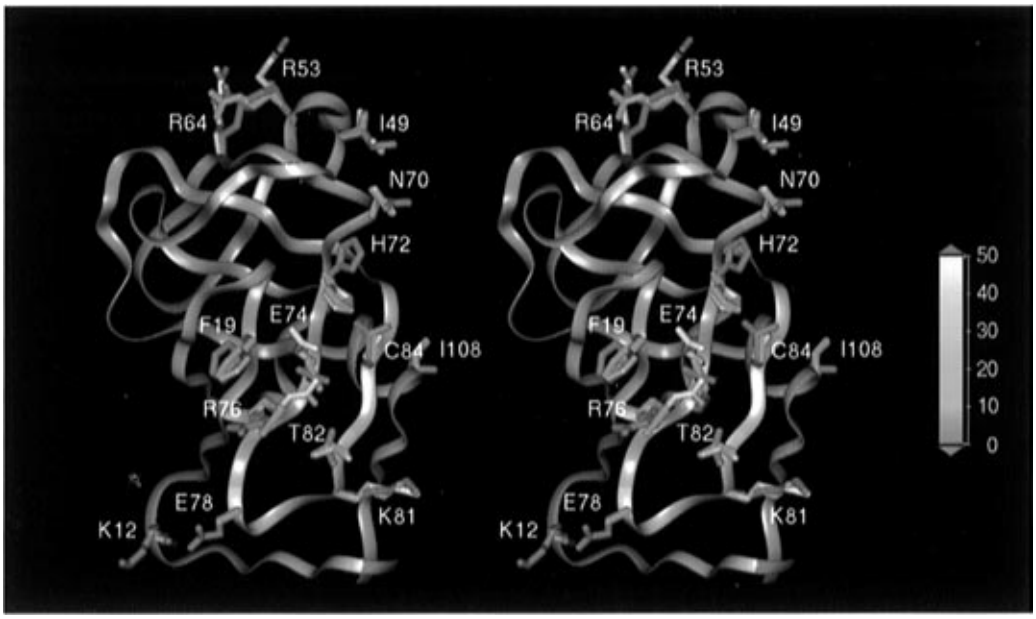

FIGURE 4: Stereo ribbon diagram of a monomer of FIV PR. The trace of the main chain is shown in a ribbon representation, and the color is applied according to the temperature factor value (shown on the right). The residues found to have a different conformation in compared structures are shown in different colors: green and magenta for FIV PR(D30N) complexed to LP-149 and the substrate, respectively, and yellow for the wild-type enzyme.

of the ligands and main chains/side chains of the enzyme (*41*) is preserved in all three complexes, and the information about the hydrogen bond distances, given in comparison with structurally equivalent interacting pairs, is summarized in Table 2.

A conserved water molecule (known as Wat301 or "flap" water), mediating the interactions between the tips of the flaps in the protease and the inhibitor, was identified in the FIV PR(wt)/LP-149 complex (*14*). The flap water is also present in the crystal structures of the complexes of FIV PR(D30N) with both the inhibitor and the substrate. This water is located on the pseudodyad axis, between the inhibitor and the flaps, and has nearly perfect tetrahedral coordination, accepting two hydrogen bonds from the amide nitrogen atoms of the valine residues (A59 and B59*) at the tip of the flap and donating two hydrogen bonds to the carbonyl oxygen atoms of inhibitor residues 203 and 204 (residues 203 and 204A in the substrate molecule).

Although the overall structure of FIV PR(D30N) is very similar to that of FIV PR(wt) (*14*), differences are evident in the conformations of a few protein side chains (Figure 4). As in FIV PR(wt), the four N-terminal residues (including the N-terminal alanine) are disordered and could not be located in the electron density map. Thr7, Phe19, Ile49, Asn70, His72, Glu74, Arg76, Lys81, Thr82, Cys84, and Ile108 were found to have variable conformations in the structures that we are comparing. The side chain of Cys84, which adopted two conformations in the FIV PR(wt)/LP-149 complex, had only one in the structures of FIV PR(D30N). In the structure of the complex of FIV PR (D30N) with the substrate, the electron density maps at the very early stages of the refinement showed positive densities adjacent to the sulfur atom of Cys84 at distances of approximately 1.6−1.7 Å. We interpreted this as the result of oxidation of

the cysteine residue, similar to what was found in HIV PR for Cys67 (*42*). In the structures of both the wild-type and D30N mutant of FIV PR complexed to LP-149, the hydrogen bond between Cys84 and His72 was observed. This interaction was not present in the substrate complex, since the side chain of His72 changed its position due to the presence of oxidized Cys84 (Figure 4).

The side chain of Arg53, modeled in two orientations in the FIV PR(wt)/LP-149 complex, was disordered in the mutant/LP-149 complex and did not show well-defined density for all the atoms of the side chain. In the mutant/substrate complex, one orientation of the Arg53 side chain could still be traced, although poorly. The side chain of Ile37, which had only one orientation in the wild-type structure, was modeled in two orientations in both complexes of the mutant. Moreover, the side chain of Lys12, which was not visible in the wild-type structure, was found in two different conformations in both mutant structures. In the case of Glu78, however, the side chain could be located only in the wild type, while it was disordered in both structures of the mutant complexes and could not be located at all in the electron density map. Arg64 adopted two orientations in the wild-type enzyme, had one orientation in mutant/substrate structure, and was disordered in mutant/LP-149 complexes.

A sulfate anion, which was not seen in the FIV PR(wt)/LP-149 complex, was identified at the interface between two molecules in both mutant complexes, making similar types of the interactions. This anion formed hydrogen bonds with the hydroxyl group of TyrA23, the nitrogen atom of LysA46, and the amide nitrogen atom of ArgB13. None of these residues interacted with the inhibitor molecule.

Active Site Region. The major differences in the structures of the wild-type and D30N FIV PRs with their ligands were located in the vicinity of the catalytic residues at the active site region. Since the asymmetric unit of the crystals contained only one monomer, both halves of the active site, formed by the residues from monomers A and B, were related by the symmetry operation. In the structure of FIV PR(wt), the two catalytic aspartates were related by crystallographic symmetry and had only one conformation each, while in the structure of FIV PR(D30N)/LP-149, the side chains of both asparagines, AsnA30 and AsnB30, had dual conformations. For active site residues A29−A32 (B29−B32), which include the mutated Asn30, the second conformation was built and numbered A29*−A32* and B29*−B32*, respectively. Both conformations had an occupancy of 0.5. As in the FIV PR(wt)/LP-149 complex (*14*), the inhibitor or the substrate molecules, when bound to the active site of FIV PR(D30N), adopted two orientations, numbered A201−A206 and B201−B206. Analysis of the distances between the catalytic aspartates (or asparagines in the D30N mutant) and the atoms of the ligands interacting with them led us to define two distinct triplets composed of two asparagine residues and one inhibitor (or substrate) molecule. These two triplets reflected the symmetry of the crystals. The first triplet consisted of AsnA30, AsnB30*, and the ligand molecule B201−B206 (B201−B207 in the substrate). The second triplet consists of AsnA30*, AsnB30, and the ligand molecule A201−A206. Other combinations of asparagine residues and the ligand molecule were not possible. The side chain amino nitrogen, ND2, of Asn30 was within hydrogen-bonding distance of the oxygen atom, OD1, of the side chain of Asn30*.

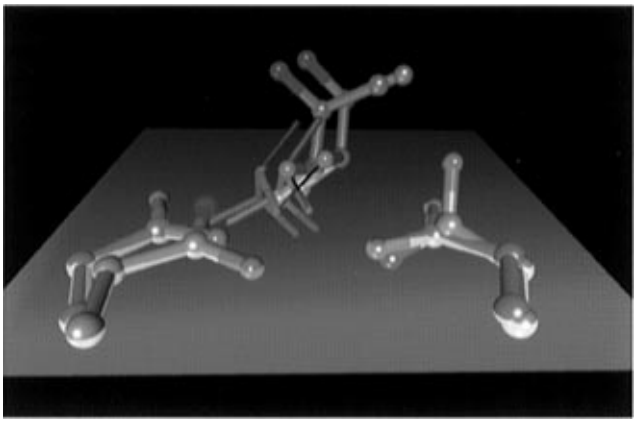

FIGURE 5: Superposition of the active sites of the wild-type FIV PR and the inactive D30N mutant complexed to LP-149. The aspartates in the active enzyme, as well as the statyl hydroxyls of the bound inhibitors, are shown in blue. The asparagines of the inactive mutant, and corresponding statyl groups, are shown in yellow. The aspartates are coplanar, with their plane marked in gray. The deviation from planarity of the asparagine side chains is clearly visible.

The binding sites of the inhibitor/substrate side chains were very similar in the structures of their complexes with FIV PR(wt) and FIV PR(D30N). Because of the crystallographic 2-fold symmetry, prime and nonprime binding pockets were identical, i.e., S1 was identical to S1′, S2 was identical to S2′, etc. The residues forming these subsites were listed previously (*14*). These residues were very well defined in both FIV PR(D30N) structures, and their overall position and conformation in the active site were similar to those observed in the FIV PR(wt)/LP-149 complex. In addition to the side chain of Val59, which was disordered in all three structures, Ile37 was found to have multiple conformations in both complexes of FIV PR(D30N).

Differences in the Active Sites of FIV PR(wt) and FIV PR(D30N) Complexed to LP-149. The conformations of aspartate/asparagine side chains, as well as the relative positions of the statyl hydroxyl groups of the inhibitor molecules, were found to be significantly different between the complexes of FIV PR(wt) and FIV PR(D30N) with LP-149. The line passing through the hydroxyl oxygens of statine residues A204 and B204 in the FIV PR(D30N) structure was roughly perpendicular to the line through the equivalent atoms in the wild-type structure (Figure 5). In addition, the two statyl hydroxyl oxygens were further apart (1.7 Å) in the mutant structure than in the wild-type structure (1.1 Å). Also, the deviations from planarity for both the asparagines and the statyl hydroxyl were much more significant in the case of the FIV PR(D30N) structure than in the wild-type enzyme, where both aspartates and the statyl hydroxyl were approximately in the same plane (Figure 5).

In both cases, the hydroxyl of the statine residue formed closer interactions with one aspartate/asparagine than with the other, in a manner similar to that found in the structure of acetylpepstatin bound to HIV-1 PR (*43*), although the network of hydrogen bonds was not exactly the same. In the active site of FIV PR(D30N), the statyl hydroxyl formed two short hydrogen bonds with both the OD1 and ND2 atoms of Asn B30 and only one short hydrogen bond with the ND2 atom of Asn A30*, located at the opposite side of the amide group in respect to the second asparagine (AsnB30) (Figure 6A). A pattern of three hydrogen bonds between the statyl

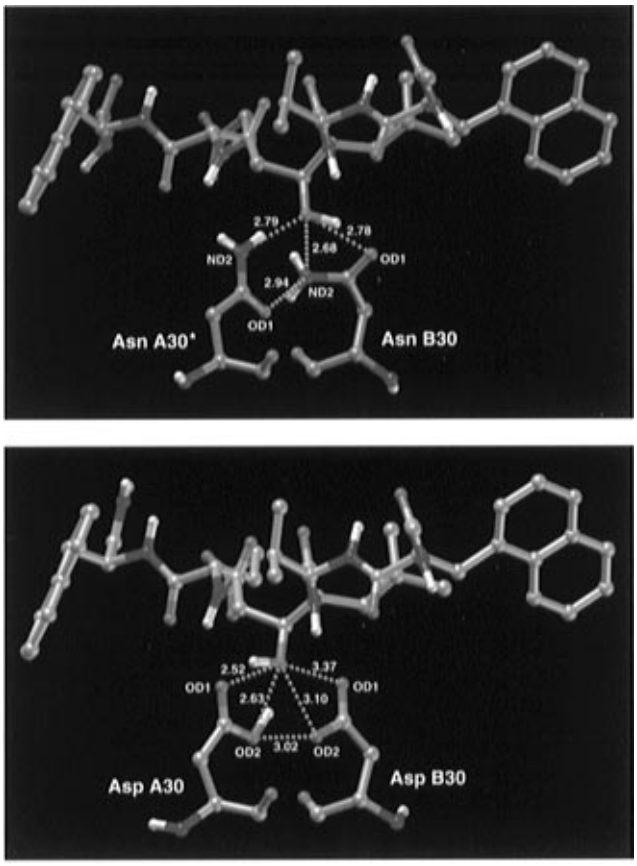

FIGURE 6: Active sites of native and mutated (D30N) FIV PRs complexed with the inhibitor LP-149. (A, top) Hydrogen-bonding network in the mutated active site of FIV PR(D30N) complexed with LP-149, involving AsnA30*, AsnB30, and LP-149. Dashed lines indicate possible hydrogen bond contacts. (B, bottom) Hydrogen-bonding network in the active site of FIV PR(wt) complexed with LP-149, involving AspA30, AspB30, and LP-149. Dashed lines indicate possible hydrogen bond contacts.

hydroxyl and aspartates was found in the majority of crystal structures of the complexes of various aspartic proteases with pepstatin or statine-based inhibitors. However, the sole hydrogen bond was usually formed with the carboxylate oxygen adjacent to another aspartate (*44*−*49*).

The distance between the two aspartates in the wild-type enzyme was close to 3 Å, nearly the same as the distance separating atoms AsnA30*(OD1) and AsnB30(ND2) in FIV PR(D30N) (Figure 6). Since protons are not observed in the electron density maps at this resolution, we cannot be sure of the assignment of the protonation states of the aspartates but can only postulate that the hydrogen bond network in the wild-type enzyme is as shown in Figure 6B. Recent studies of the ionization states of the catalytic aspartates in HIV-1 PR using NMR (*50*) have shown that, in the complexes of HIV-1 PR with pepstatin A, the catalytic aspartate that interacts with the statyl hydroxyl is protonated. These data are in good agreement with the interaction scheme deduced from the crystal structure of FIV PR(wt)/LP-149. The two conformations of the asparagine residues differed also in their temperature factors. AsnA30* and AsnB30*, with the oxygens adjacent to each other, had *B*-factors of ∼8 Å^2, while the other pair of asparagines, AsnA30 and AsnB30, with NH_2 groups at the same positions, had *B*-factors of ∼15 Å^2. The oxygen atom, if located in the center of an asparagine pair, formed a hydrogen bond of length ∼3 Å with an NH group of Gly32. This interaction was not possible when the amido group of the asparagine was flipped 180°, and an NH_2 group is located at the former position of an oxygen atom. The distance between the two nitrogen moieties (NH_2 of Asn30 and NH of Gly32) increased in this case to 3.4 Å. It should be noted that, in the structure of FIV PR(wt)/LP-149, the hydrogen bond between the inner oxygen of the catalytic aspartate and NH group of Gly32 was 3.15 Å, which may reflect the interactions of the aspartates with the statyl hydroxyl.

The conserved network of hydrogen bonds between the main chain of the ligand and the protease (*41*) was preserved in the structures of all three complexes, with the comparison of their distances shown in Table 2.

Interactions of the Substrate Bound to the Active Site of FIV PR(D30N). The structural information obtained from the crystal structure of the FIV PR(D30N)/LP-149 complex raised some questions concerning the catalytic mechanism of aspartyl proteases. It has been shown previously that the hydroxyl group of statine-based inhibitors replaces the nucleophilic water molecule present in the active site (*20*, *49*). Since we observed that the binding mode of LP-149 in the inactive mutant was similar to that in the wild-type enzyme, this similarity implies that the nucleophilic water molecule may also be present in the inactive, uncomplexed FIV PR(D30N), with the location of the statyl hydroxyl group indicating its possible position. Considering these results, one can address the following question: if the nucleophilic water molecule is present between the two asparagines, why is FIV PR(D30N) unable to cleave the peptide bond of a substrate? One possible answer is that the water molecule might not be activated as strongly by asparagines as by aspartates, thus not allowing successful nucleophilic attack on the carbonyl carbon of the scissile bond. Another possibility is that the nitrogen atom of the scissile bond, which was thought to be protonated by the catalytic aspartates, cannot be protonated in the D30N mutant and is therefore no longer a good leaving group. We solved the structure of the substrate bound to FIV PR(D30N) to clarify these points.

Crystals of the complexes of FIV PR(D30N) and the substrate LP-149S were successfully obtained under the same conditions as the complexes with the inhibitor and were isomorphous with the latter. As mentioned in the Materials and Methods section, two models with a different arrangement of asparagines in the active site had to be built and refined. In the first model, asparagine residues were built in two conformations each, with an asymmetric assignment of the atoms in each pair (similar to the structure of the active site with bound LP-149). Since the electron density for the asparagine side chains did not indicate unambiguously multiple conformations, we built an alternative model, in which both asparagines were related by a symmetry operation and had a single conformation each. We refined models to the same of *R*-factor value, so that they could not be distinguished on this basis.

The electron density and the refined structure of LP-149S bound to FIV PR(D30N) did not indicate any distortion from planarity for the P1−P1′ peptide bond. The interactions between the substrate and asparagines were limited to the formation of one hydrogen bond between the carbonyl oxygen at P1 and the nitrogen atom of the amide group of the asparagine (Figure 7). We assigned the atoms in the model with the symmetric arrangement of asparagines in the

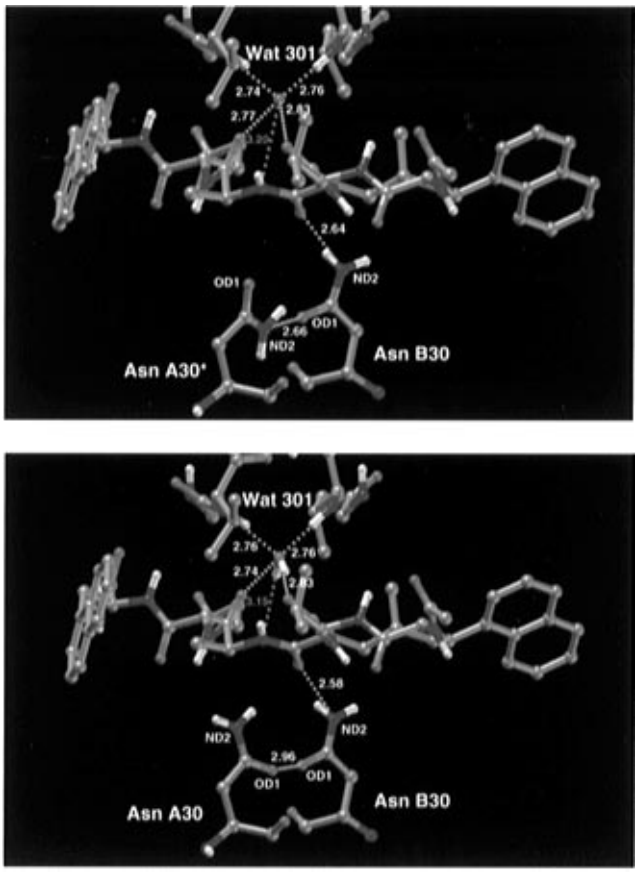

FIGURE 7: "Active site" of the inactive mutant FIV PR(D30N) complexed with the substrate LP-149S. (A, top) Interactions of a substrate bound to FIV PR(D30N) with an asymmetric assignment of the atoms in the asparagines. Dashed lines (in white) indicate possible hydrogen bonds. The distance between the amide nitrogen of the scissile bond and Wat301 is shown in green. (B, bottom) Analogous interactions with the symmetric assignment of asparagines.

active site to fulfill the requirement for the formation of this hydrogen bond, putting two oxygens "inside" and two NH_2 groups "outside" of the pair of asparagines. In the first model, the nitrogen atom of the amide group of the second asparagine was within the hydrogen-bonding distance of an oxygen atom of the first asparagine (~2.7 Å, Figure 7A). The "conserved" hydrogen bond between one inner oxygen atom of the AsnB30 and the NH group of GlyB32 was present (2.8 Å), while in the other monomer the side chain of AsnA30*, with the NH_2 group in the "middle" of the asparagine pair, was slightly rotated, increasing the distance between its nitrogen to the NH of GlyA32 to ~3.1 Å. Because of less favorable contacts of the NH_2 moiety when it was located in the center of the asparagine pair, AsnA30* and AsnB30* had higher temperature factors than those of residues AsnA30 and AsnB30 [~19 vs ~9 $Å^2$, similar to the structure of FIV PR(D30N)/LP-149]. In the second model, the distances between the inner oxygens of the asparagines increased to ~3 Å (Figure 7B), while both conserved hydrogen bonds between these and the NH groups of GlyA32 and GlyB32 were restored (~2.8 Å).

The flap water molecule (Wat301) was present in near-perfect tetrahedral coordination at the position that is conserved in all structures of retroviral proteases with peptidomimetic inhibitors. The nitrogen atom of the scissile bond was found to be fairly close to this water (~3.2 Å). With the exception of the atoms of the unmodified peptide bond Leu-Ala, both the N-terminal and C-terminal halves of the substrate LP-149S had conformations similar to that of LP-149 when bound to FIV PR(wt) and FIV PR(D30N). It should be noted, however, that the N-terminal halves of the substrate and inhibitor were more similar than the C-terminal halves. The differences in the conformation of C-terminal moieties increased toward the C-terminus of the substrate with a significantly different position of the naphthylalanine side chain at P3′. However, all of the conserved substrate/enzyme interactions were present (see Table 2).

After refinement, the active site residues from both models fit the electron density map equally well; thus, we were unable to distinguish between them. While the asymmetric assignment of the atoms for the pair of asparagines (Figure 6A) was experimentally found in the active site of LP-149 complexes with the enzyme, we cannot rule out the possibility of a symmetrical arrangement of the atoms (Figure 7B) in the structure of the substrate complexes, neither on experimental grounds nor by considering the substrate/enzyme interactions. These differences are, however, irrelevant to the discussion of the catalytic mechanism.

DISCUSSION

We solved a series of crystal structures of the complexes of FIV PR(wt) and FIV PR(D30N), with a statine-based inhibitor and with substrate, in order to study the effects of the mutation, the recognition principles of retroviral proteases, and their mechanism of action. The substitution of the catalytic aspartates by asparagines did not affect the overall enzyme structure, but the conformation of the asparagines and the position of the LP-149 statyl hydroxyl interacting with them were different from the equivalent aspartates in the wild-type enzyme. Despite this, the networks of hydrogen bonds found in both active sites, involving the catalytic aspartates and asparagines, were very similar. Our results suggest that the nucleophilic water bound to catalytic residues in the native enzyme is also present in the uncomplexed D30N mutant, although its position may be shifted in comparison with its actual location in the wild-type enzyme. We solved the crystal structure of FIV PR(D30N) complexed with the substrate LP-149S in order to answer the question whether the D30N mutant is inactive because the asparagines cannot activate this water to provide a successful nucleophilic attack on a carbonyl carbon atom or because the catalysis is interrupted at a later stage.

The crystal structure of FIV PR(D30N)/LP-149S revealed that the P1−P1′ peptide bond of the bound substrate did not deviate from planarity, indicating that the tetrahedral intermediate was not formed, and thus the asparagines most likely did not polarize a water molecule sufficiently to become a strong nucleophile. Another possible reason for the inactivation of the enzyme by the D30N mutation may be the different location of the nucleophilic water molecule with respect to the asparagines (vs aspartates in the active enzyme) and the bound substrate, which is indicated by the different positions of the statyl hydroxyls in their active sites (the distance between their positions in the native and mutated enzymes is ~1 Å).

The structural data presented here contain unique information about the interactions of the intact substrate molecule

bound in the active site of the retroviral enzyme. They experimentally confirm the assumption about the uniformity of the interactions between the enzyme and any bound ligands of a peptidomimetic type. The interactions of the atoms of an unmodified peptide bond, P1−P1′, were observed and described in this report. The carbonyl oxygen of the scissile bond interacts with one asparagine, forming a hydrogen bond with the NH_2 group of Asn30. This is the only interaction found in the crystal structure between the substrate and the asparagines. The nitrogen atom of the scissile bond is not involved in any interactions with the enzyme but is near (∼3.2 Å) the conserved water molecule (Wat301), which mediates the interactions between the substrate and the flaps of the enzyme.

Wat301 (or its substitute) is invariably present in all crystal structures of the retroviral proteases with inhibitors. NMR experiments have also confirmed that this water is present in the complexes of HIV-1 PR with peptidomimetic inhibitors KNI-272 (*51*) and P9941 (*52*) in solution. In the latter report, it was shown that the long-lived water, found in the expected position in the structure with the linear inhibitor P9941, is replaced by a urea oxygen in the complexes of HIV-1 PR with a cyclic urea-based inhibitor, DMP323. Several recent crystal structures of the complexes of HIV-1 PR with various products of hydrolysis, solved at high resolution, revealed that Wat301 is also retained (*53*). The fact that Wat 301 is present throughout all steps of enzymatic activity and maintains the hydrogen bonds with both the intact substrate and the reaction products confirms that Wat301 plays an important role, although the full extent of its involvement is not yet clear.

To investigate the interactions of the tetrahedral intermediate formed during catalysis with the enzyme in the vicinity of the aspartates, we built two models of the catalytic intermediate, bound in the active site of FIV PR(wt) with different protonation states of the aspartates. In one of the models, the proton was shared by both aspartates, while in the other model only one aspartate was protonated. We built a model of the tetrahedral intermediate (*gem*-diol) by using the structures described here and that of a pyrrolidine-containing α-keto amide inhibitor, bound to HIV-1 PR (*36*). This model showed that the inhibitor used in the latter study was not hydrated until bound to the protease. The *gem*-diol was formed by nucleophilic attack of the water molecule bound in the active site of HIV-1 PR; thus, the structure of the hydrated ketone is likely to be similar to the structure of the tetrahedral intermediate formed during the hydrolysis of a peptide substrate. Molecular dynamics simulations were carried out with both models. It was found that the protonation state of the aspartates had a strong effect on the results of the calculations. Nevertheless, in both models the distance between the sp^3 nitrogen of the scissile bond and Wat301 decreased to ∼3.0 Å, close to the 3.2 Å observed for the corresponding atoms in the crystal structure of FIV PR(D30N) complexed with the substrate, although in both initial models this distance was close to 4 Å. Clearly, the distance resulting from the molecular dynamics calculations is close to the one observed experimentally.

Several proteolytic mechanisms have been proposed for the family of aspartic proteases, including nonviral and retroviral enzymes (*24, 54*−*58*). All of these mechanisms assume nucleophilic attack by a central water molecule, which is bound to the catalytic aspartates in the active site of the native enzyme. This water molecule is replaced by an oxygen atom of the statyl hydroxyl in several crystal structures of various aspartic proteases with statine-based inhibitors (*20, 49*). The structural data presented here allowed us to investigate the disposition of the scissile bond in our substrate structure with respect to the predicted position of the nucleophilic water in the native enzyme (*59*). If the structure of LP-149S bound to FIV PR(D30N) is superimposed on the structure of LP-149 bound to FIV PR-(wt), the position of the statyl hydroxyl should indicate the position of the catalytic water. In this case, the angle O−C═O of nucleophilic attack on the carbon of the carbonyl will be ∼83°. This angle is close to the optimal value for such reactions, namely ∼109°, according to the Baldwin rules (*60*). Thus, the observed conformation of the scissile bond in a substrate bound to FIV PR(D30N) seems to satisfy the requirements for productive nucleophilic attack of catalytic water in the native enzyme. This conclusion is not surprising, but is noteworthy nonetheless, since the conformation of the scissile bond is strongly constrained by numerous enzyme/substrate interactions.

Similar interactions were found in the crystal structures of the complexes of N- and C-terminal products of peptide hydrolysis bound to HIV-1 PR (*53*). The carboxyl group at the C-terminus of the N-terminal product interacts with aspartates in such a way that the model of a substrate, reconstructed from the crystal structures of both N- and C-terminal products, has an arrangement near the aspartates very similar to that found in the substrate structure presented here. A similar binding mode has also been observed recently in a structure of HIV-1 PR with bound substrate (A. Silva, personal communication).

A revision of the proposed catalytic mechanisms, based on the new structural evidence presented here, can indicate what aspects are established and what questions still remain. The formation of a hydrogen bond between the substrate's carbonyl oxygen and an "outer" oxygen atom of one of the catalytic aspartates in order to polarize the carbonyl oxygen and thus make the carbon atom more susceptible to nucleophilic attack has been proposed for a number of catalytic mechanisms (*55*−*58, 61, 62*). This interaction agrees well with the experimental data observed in the crystal structures of the complexes of native and mutant FIV PRs presented here and of mutant and native HIV-1 PRs complexed with a substrate (A. Silva, personal communication) and with products (*53*), respectively. Thus, it can be accepted as one known step of the proteolysis. The other significant role of this hydrogen bond is in holding the scissile bond in a favorable position for nucleophilic attack by a catalytic water molecule. There is less consensus about the coordination of the water molecule during the nucleophilic attack (*55*), although the majority suggests that the water molecule forms three hydrogen bonds with the catalytic aspartates at the initial stage of the catalysis. Two of these bonds are with one aspartate; the third one is with the other aspartate, forming a network of hydrogen bonds very similar to that of the statyl hydroxyl. This scheme is also in good agreement with crystallographic data on the interactions with the aspartates of the *gem*-diol moieties in various inhibitors (*36, 56*−*58, 63*). A more symmetric disposition of the nucleophilic water and the *gem*-diol intermediate with respect to the catalytic aspartates was proposed by Rose et al. (*53*) on the basis of the experimentally observed interactions of

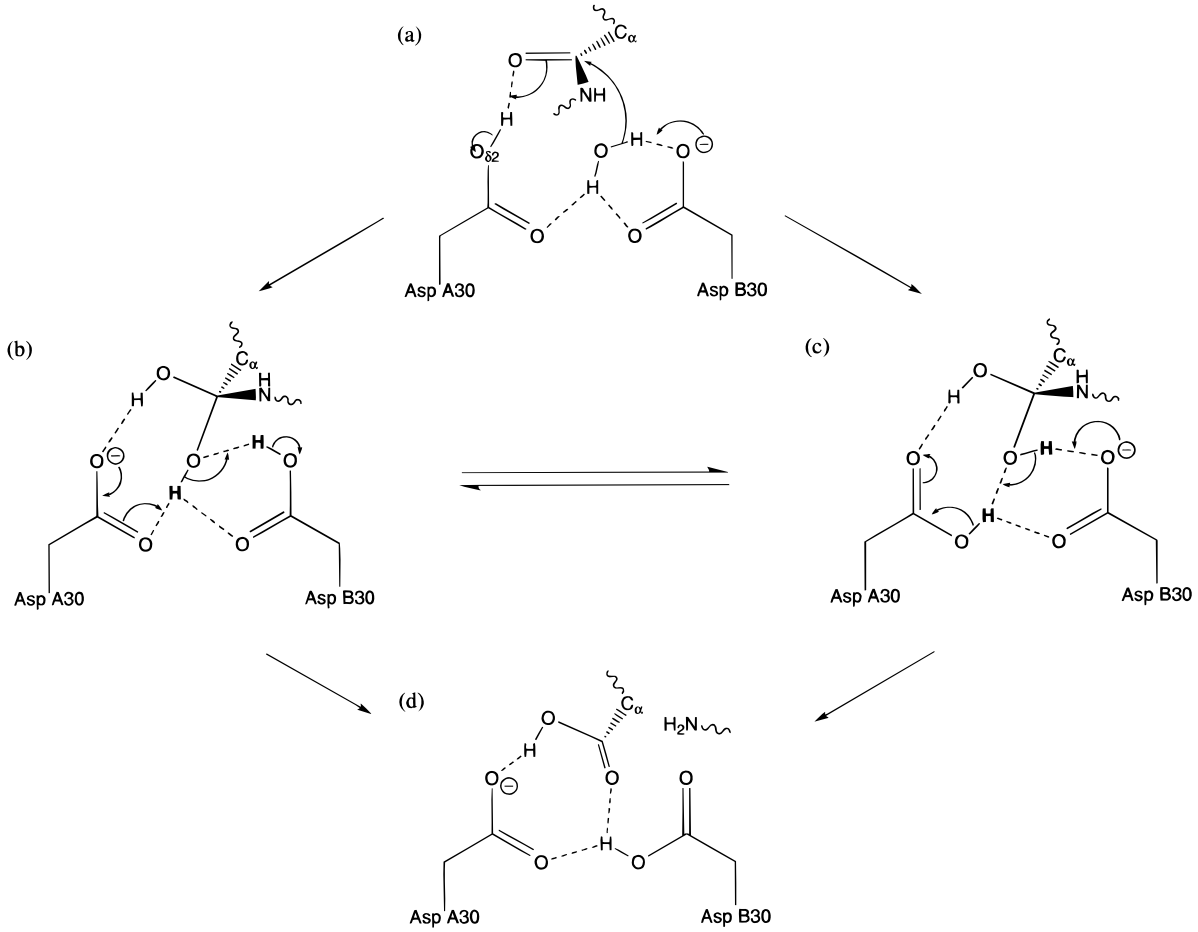

FIGURE 8: Essential steps of the catalytic mechanism of aspartic proteases that are pertinent to this study. (a) The first step in catalysis: the formation of a Michaelis complex, which involves the abstraction of a proton from the so-called catalytic water molecule by AspB30 in order to assist the nucleophilic attack of the water's oxygen atom upon the carbonyl carbon atom of the scissile bond. Note the hydrogen bond donated by OD2 of AspA30 to the carbonyl oxygen atom. (b and c) The enzyme-bound amide hydrate, in which the substrate has become a tetrahedral *gem*-diol intermediate. Possible pathways of proton transfer between the two catalytic aspartates and the *gem*-diol hydroxyls are shown. (d) An example of possible reaction product formation, in which the C-terminus of the N-terminal product is hydrogen bonded to AspA30.

the reaction products. However, the majority of the structural data indicates that the most probable hydrogen-bonding mode for the Michaelis complex is as shown in Figure 8a. The activation of the nucleophilic water molecule by the abstraction of a proton and the attack of the emerging hydroxyl upon the carbonyl carbon of the scissile bond occur simultaneously with the protonation of the carbonyl oxygen by AspA30, resulting in the formation of a *gem*-diol intermediate. Possible pathways for proton transfer are shown in Figure 8, parts b and c. Similar schemes of the interactions in these stages of catalysis have been proposed elsewhere and are as shown in Figure 8b (*24*, *57*, *58*, *61*, *63*), and in Figure 8c (*53*, *55*, *56*). There is a reversible equilibrium between these two states of the *gem*-diol intermediate, and a similar arrangement was described for the chemical mechanism proposed by Hyland et al. (*55*).

During proteolysis, the amide nitrogen, as well as the carbonyl carbon atom, undergoes a change in hybridization, from sp^2 to sp^3. A proton must be transferred to the lone pair of the sp^3 nitrogen atom in order to make the C-terminal product a good leaving group. For a long time, one of the catalytic aspartates was considered to be the source of this proton (*24*, *55*−*58*, *61*). To accomplish the proton transfer, rotation around the C(OH)N bond is required to bring the nitrogen's lone pair close to the aspartates (*57*, *58*). In the crystal structure of FIV PR(D30N) with the substrate LP-149S, the closest distance between the nitrogen of the scissile bond and the asparagines is approximately 4 Å. The same distance was observed in a modeled structure of a substrate bound to HIV-1 PR, leading to the suggestion that one of the *gem*-diol's hydroxyls might be the source of the proton that is donated to the nitrogen of the cleaved peptide (*53*).

Molecular dynamics simulations carried out here investigated where this proton might come from. By measuring the distance from the sp^3 nitrogen to each candidate proton, we were able to infer which protons were unavailable for donation and which one was most likely to be donated. In the minimum-energy structure from the simulation in which the two catalytic aspartates shared a proton (Figure 8c), the distances from the sp^3 nitrogen and to the protons of Wat301 were 2.91 and 2.78 Å, to the proton on the carboxylate oxygen of AspA30, 2.94 Å, and to the two *gem*-diol protons, 2.96 and 1.95 Å. Thus, a *gem*-diol proton is nearest to the soon-to-be-released sp^3 nitrogen.

Quantum chemical calculations have shown that the direct transfer of a proton from a *gem*-diol hydroxyl to the amide nitrogen of the peptide bond has a high energy barrier (∼30−40 kcal/mol) (*64*). However, in the presence of two aspartates, the *gem*-diol proton, which is nearest to the amide nitrogen, as well as a second proton (both shown as bold in

Figure 8b,c) may be transferred along the same pathway, but in two opposite directions, from the "inner" oxygen of AspA30 to the *gem*-diol oxygen atom, which simultaneously transferred its proton to the "outer" oxygen of AspB30 and back. This motion is accompanied by a switch of the negative charge from one aspartate to another, and as a result, two reversible states, shown in Figure 8b,c, are formed. The fact that this proton can be exchanged *between* the oxygen atoms of the *gem*-diol and AspB30 makes its transfer to the nitrogen atom more favorable than if this proton were located at either of these oxygen atoms. Thus, the transfer of a *gem*-diol proton to the amide nitrogen is assisted by the presence of both aspartates. The products are formed as soon as the proton is transferred to the amide nitrogen (Figure 8d), while the electronic state of the catalytic aspartates with a shared proton located between them is restored.

Although this interpretation of the results of the studies of active and inactive FIV PRs is in agreement with structural data, we are not yet in a position to provide a detailed, unambiguous mechanism of action of aspartic proteases. Further crystallographic and kinetic studies of various ligand complexes of this and other aspartic proteases are still necessary in order to elucidate those aspects of the catalytic mechanism that remain uncertain.

ACKNOWLEDGMENT

We thank Dr. Kwan Y. Hui for the gift of the inhibitor LP-149, Drs. Stephen B. Kent, Gottfried Palm, Pavel Majer, and Sergei Gulnik for thoughtful discussions, Michael L. Fitzgerald for technical support, and Anne Arthur for editorial assistance.

REFERENCES

1. Toh, H., Ono, M., Saigo, K., and Miyata, T. (1985) *Nature 315*, 691.
2. Hansen, J., Billich, S., Schulze, T., Sukrow, S., and Moelling, K. (1988) *EMBO J. 7*, 1785−1791.
3. Seelmeier, S., Schmidt, H., Turk, V., and von der Helm, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6612−6616.
4. Darke, P. L., Leu, C. T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A., and Sigal, I. S. (1989) *J. Biol. Chem. 264*, 2307−2312.
5. Richards, A. D., Roberts, R., Dunn, B. M., Graves, M. C., and Kay, J. (1989) *FEBS Lett. 247*, 113−117.
6. Schneider, J., and Kent, S. B. (1988) *Cell 54*, 363−368.
7. Umezawa, H. (1976) *Methods Enzymol. 45*, 678.
8. Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J., and Wlodawer, A. (1989) *Nature 337*, 576−579.
9. Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature 337*, 615−620.
10. Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) *Science 245*, 616−621.
11. Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., and Anderson, P. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8387−8391.
12. Zhao, B., Winborne, E., Minnich, M. D., Culp, J. S., Debouck, C., and Abdel-Meguid, S. S. (1993) *Biochemistry 32*, 13054−13060.
13. Gustchina, A., Kervinen, J., Powell, D. J., Zdanov, A., Kay, J., and Wlodawer, A. (1996) *Protein Sci. 5*, 1453−1465.
14. Wlodawer, A., Gustchina, A., Reshetnikova, L., Lubkowski, J., Zdanov, A., Hui, K. Y., Angleton, E. L., Farmerie, W. G., Goodenow, M. M., Bhatt, D., Zhang, L., and Dunn, B. M. (1995) *Nat. Struct. Biol. 2*, 480−488.
15. 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) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686−4690.
16. Loeb, D. D., Hutchison, C. A., III, Edgell, M. H., Farmerie, W. G., and Swanstrom, R. (1989) *J. Virol. 63*, 111−121.
17. Le Grice, S. F., Mills, J., and Mous, J. (1988) *EMBO J. 7*, 2547−2553.
18. Mous, J., Heimer, E. P., and Le Grice, S. F. (1988) *J. Virol. 62*, 1433−1436.
19. Kotler, M., Katz, R., and Skalka, A. M. (1988) *J. Virol. 62*, 2696−2700.
20. Gilliland, G. L., Winborne, E. L., Nachman, J., and Wlodawer, A. (1990) *Proteins 8*, 82−101.
21. Fruton, J. S. (1976) *Adv. Enzymol. Relat. Areas Mol. Biol. 44*, 1−36.
22. Marczinovits, I., Molnar, J., and Patthy, A. (1994) *J. Biotechnol. 37*, 79−83.
23. Pearl, L. H. (1987) *FEBS Lett. 214*, 8−12.
24. Jaskólski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., and Wlodawer, A. (1991) *Biochemistry 30*, 1600−1609.
25. Kaplan, A. H. (1996) *AIDS Res. Hum. Retroviruses 12*, 849−853.
26. Wlodawer, A., and Erickson, J. W. (1993) *Annu. Rev. Biochem. 62*, 543−585.
27. Vondrasek, J., van Buskirk, C. P., and Wlodawer, A. (1997) *Nat. Struct. Biol. 4*, 8.
28. Elder, J. H., Schnolzer, M., Hasselkus-light, C. S., Henson, M., Lerner, D. A., Philips, T. R., Wagaman, P. C., and Kent, S. B. H. (1993) *J. Virol. 67*, 1869−1876.
29. Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074−1078.
30. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendoff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
31. Schnolzer, M., Rackwitz, H. R., Gustchina, A., Laco, G. S., Wlodawer, A., Elder, J. H., and Kent, S. B. (1996) *Virology 224*, 268−275.
32. Chait, B. T., and Kent, S. B. (1992) *Science 257*, 1885−1894.
33. Otwinowski, Z. (1992) in *An Oscillation Data Processing Suite for Macromolecular Crystallography*, Yale University, New Haven, CT.
34. Maple, J. R., Dinur, U., and Hagler, A. T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5350−5354.
35. Maple, J. R., Thatcher, T. S., Dinur, U., and Hagler, A. T. (1990) *Chem. Des. Autom. News 5*, 5−10.
36. Slee, D. H., Laslo, K. L., Elder, J. H., Ollmann, I. R., Gustchina, A., Kervinen, J., Zdanov, A., Wlodawer, A., and Wong, C.-H. (1995) *J. Am. Chem. Soc. 117*, 11867−11878.
37. Brünger, A. (1992) in *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.
38. Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.
39. Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252−270.
40. Finzel, B. C. (1987) *J. Appl. Crystallogr. 20*, 53−55.
41. Gustchina, A., Sansom, C., Prevost, M., Richelle, J., Wodak, S. Y., Wlodawer, A., and Weber, I. T. (1994) *Protein Eng. 7*, 309−317.
42. Thanki, N., Rao, J. K. M., Foundling, S. I., Howe, W. J., Moon, J. B., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Thaisrivongs, S., and Wlodawer, A. (1992) *Protein Sci. 1*, 1061−1072.
43. Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C.-T., Herber, W. K., Dixon, R. A. F., and Darke, P. L. (1990) *J. Biol. Chem. 265*, 14209−14219.
44. James, M. N., Sielecki, A., Salituro, F., Rich, D. H., and Hofmann, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6137−6141.
45. Suguna, K., Padlan, E. A., Bott, R., Boger, J., Parris, K. D., and Davies, D. R. (1992) *Proteins 13*, 195−205.
46. Fraser, M. E., Strynadka, N. C., Bartlett, P. A., Hanson, J. E., and James, M. N. (1992) *Biochemistry 31*, 5201−5214.
47. Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C., and James, M. N. (1995) *Protein Sci. 4*, 960−972.
48. Gulnik, S., Baldwin, E. T., Tarasova, N., and Erickson, J. (1992) *J. Mol. Biol. 227*, 265−270.

49. Bailey, D., Cooper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M., and Szelke, M. (1993) *Biochem. J. 289*, 363−371.
50. Smith, R., Brereton, I. M., Chai, R. Y., and Kent, S. B. (1996) *Nat. Struct. Biol. 3*, 946−950.
51. Wang, Y., Freedberg, D. I., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., Bhat, N., Erickson, J. W., and Torchia, D. A. (1996) *J. Am. Chem. Soc 118*, 12287−12290.
52. Grzesiek, S., Bax, A., Nicholson, L. K., Yamazaki, T., Stahl, S. J., Eyermann, C. J., Torchia, D. A., Hodge, C. N., Lam, P. Y., Jadhav, P. K., and Chang, C.-H. (1994) *J. Am. Chem. Soc 116*, 1581−1582.
53. Rose, R. B., Craik, C. S., Douglas, N. L., and Stroud, R. M. (1996) *Biochemistry 35*, 12933−12944.
54. Suguna, K., Bott, R., Podlan, E., Subramanian, E., Sheriff, S., Cohen, G., and Davies, D. (1987) *J. Mol. Biol. 196*, 877−900.
55. Hyland, L. J., Tomaszek, T. A., Jr., and Meek, T. D. (1991) *Biochemistry 30*, 8454−8463.
56. James, M. N., Sielecki, A. R., Hayakawa, K., and Gelb, M. H. (1992) *Biochemistry 31*, 3872−3886.
57. Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) *Protein Sci. 1*, 322−328.
58. Silva, A. M., Cachau, R. E., Sham, H. L., and Erickson, J. W. (1996) *J. Mol. Biol. 255*, 321−346.
59. Burgi, H. B., Dunitz, J. D., and Shefter, E. (1973) *J. Am. Chem. Soc 95*, 5065−5067.
60. Kocovsky, P., Turecek, F., and Hajicek, J. (1986) in *Synthesis of Natural Products: Problems of Stereoselectivity*, pp 16−21, CRC Press, Boca Raton, FL.
61. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., and Davies, D. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7009−7013.
62. Lai, M. H., Tang, J., Wroblewski, V., Dee, A. G., Margolin, N., Vlahos, C., Bowdon, B., Buckheit, R., Colacino, J., and Hui, K. Y. (1993) *J. Acquired Immune Defic. Syndr. 6*, 24−31.
63. Parris, K. D., Hoover, D. J., Damon, D. B., and Davies, D. R. (1992) *Biochemistry 31*, 8125−8141.
64. Oie, T., Loew, G. H., Burt, S. K., Binkley, S., and MacElroy, R. D. (1982) *J. Am. Chem. Soc 104*, 6169−6174.
65. Engh, R., and Huber, R. (1991) *Acta Crystallogr. A 47*, 392−400.

BI9707436