Structure of the Protease from Simian Immunodeficiency Virus: Complex with an Irreversible Nonpeptide Inhibitor†,‡

Robert B. Rose,§ Jason R. Rosé,‖ Rafael Salto,⊥ Charles S. Craik,§,⊥ and Robert M. Stroud*,§,⊥

Departments of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448

Received June 9, 1993; Revised Manuscript Received August 27, 1993®

ABSTRACT: A variant of the simian immunodeficiency virus protease (SIV PR), covalently bound to the inhibitor 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP), was crystallized. The structure of the inhibited complex was determined by X-ray crystallography to a resolution of 2.4 Å and refined to an *R* factor of 19%. The variant, SIV PR S4H, was shown to diminish the rate of autolysis by at least 4-fold without affecting enzymatic parameters. The overall root mean square (rms) deviation of the α-carbons from the structure of HIV-1PR complexed with a peptidomimetic inhibitor (7HVP) was 1.16 Å. The major differences are concentrated in three surface loops with rms differences between 1.2 and 2.1 Å. For 60% of the molecule the rms deviation was only 0.6 Å. The structure reveals one molecule of EPNP bound per protease dimer, a stoichiometry confirmed by mass spectral analysis. The epoxide moiety forms a covalent bond with either of the active site aspartic acids of the dimer, and the phenyl moiety occupies the P1 binding site. The EPNP nitro group interacts with Arg 8. This structure suggests a starting template for the design of nonpeptide-based irreversible inhibitors of the SIV and related HIV-1 and HIV-2 PRs.

Simian immunodeficiency virus (SIV)[1] is a retrovirus closely related to the type 2 human immunodeficiency virus (HIV-2) and more distantly related to human immunodeficiency virus type 1 (HIV-1). Infectious clones of a strain isolated from macaque monkeys ($SIV_{mac}239$) also produce an AIDS-like disease in rhesus monkeys (Kestler et al., 1990). These infected monkeys provide an animal model for testing therapeutic agents targeting HIV-1 or HIV-2. The structure of SIV_{mac} protease (SIV PR) was determined to facilitate the incorporation of data from *in vivo* testing with efforts to improve the design of drugs targeting the HIV-2 and HIV-1 PRs.

The residues in the binding pocket of SIV PR differ from those of HIV-1 PR in 3 of 13 positions identified as major peptide binding determinants in HIV-1 PR (Miller et al., 1989). Despite these differences, SIV PR is capable of authentically processing the HIV-1 $p53^{gag}$ polyprotein *in vitro* (Grant et al., 1991). The sequence of HIV-2 PR is identical to SIV PR at these 13 positions.

It has been shown *in vivo* that the virus can develop resistance to drugs targeting reverse transcriptase (Richman, 1993). It has been proposed that the virus could develop resistance to antiprotease drugs as well (Cameron et al., 1993). A comparison of the HIV-1, HIV-2, and SIV protease structures will identify regions that are structurally conserved and may represent binding sites for inhibitors that are less susceptible to the development of drug resistance.

A variant of SIV_{mac} protease was constructed to stabilize the enzyme to autolysis, by replacing serine 4 by histidine (SIV PR S4H), at the major autolytic site (Rosé et al., 1993). This mutation was found to diminish the rate of autolysis by a factor of 4 without affecting enzymatic parameters.

1,2-Epoxy-3-(*p*-nitrophenoxy)propane (EPNP) is a covalent inhibitor of aspartyl proteases, requiring the presence of two neighboring carboxyl groups for the reaction to take place at the active site. This is the first structure of a retroviral protease bound to a covalent inhibitor. The structure reveals aspects that bear directly on the reaction mechanism of inhibition.

MATERIALS AND METHODS

Purification of Recombinant Proteins. The expression of $SIV_{mac}239$ from plasmid SOD/SIV has been described elsewhere (Rosé et al., 1993). Ser 4 of SIV PR was replaced with His to increase the stability of the protease to autoproteolysis. Soluble mature SIV PR was purified from lysates generated by sonication of induced cells harboring the plasmid. Pellets were resuspended in 14 volumes of lysis buffer [50 mM Tris, pH 7.5, 5 mM EDTA, 10% glycerol, 0.2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (ICN)] and lysed by sonication. The lysate was clarified by centrifugation at 12 500 rpm in a SS34 rotor for 30 min and the supernatant precipitated overnight by treatment with 57.5% ammonium sulfate. The pellet was resuspended in 5 volumes of 50 mM HEPES, pH 7.5, and 1 mM EDTA and loaded on a column of pepstatin A agarose. The protease was eluted with a buffer containing 250 mM ε-aminocaproic acid, pH 10.5, 5% glycerol, 5% ethylene glycol, and 1 mM EDTA. Active fractions were pooled and loaded onto a DEAE-Sepharose column equilibrated in the pepstatin A agarose column elution buffer. The protease was eluted from the DEAE column with 50 mM HEPES, pH 7.0, and 1 mM EDTA. Fractions after the pH change were analyzed by SDS–PAGE on 15% polyacrylamide gels, and fractions of single-band purity were used for crystallography (Figure 1). Typical yields were 6–8 mg of

† This work was supported by NIH Grant GM 39552. R.B.R. was supported by a graduate training grant from the University of California Systemwide Biotechnology Research and Education Program, Project Title "Biotechnology Program in Program Crystallization". J.R.R. was supported by Graduate Training Grant GM 09175.

‡ The atomic coordinates in this paper have been submitted to the Brookhaven Protein Data Bank (1Sam).

* Author to whom correspondence should be addressed.

§ Graduate Group in Biophysics.

‖ Department of Pharmacology.

⊥ Department of Pharmaceutical Chemistry.

® Abstract published in *Advance ACS Abstracts,* October 15, 1993.

[1] Abbreviations: DMSO, dimethylsulfoxide; EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; SIV, simian immunodeficiency virus; HIV, human immunodeficiency virus; PR, protease; DTT, dithiothreitol; HPLC, high-performance liquid chromatography; EDTA, ethylenediaminetetraacetic acid.

0006-2960/93/0432-12498$04.00/0

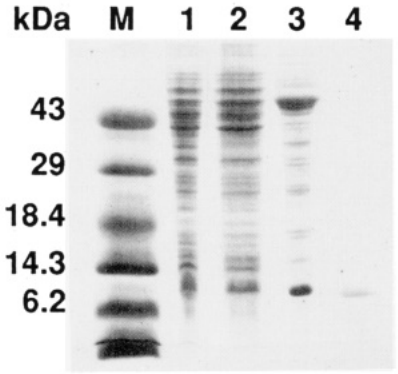

FIGURE 1: Purification of SIV PR. (Lane M) Low molecular weight markers; (lane 1) crude cell lysate; (lane 2) ammonium sulfate precipitate; (lane 3) pepstatin column elution pool; (lane 4) DEAE elution pool.

purified SIV PR from 6.5 g of *Escherichia coli*. The protease was frozen over dry ice without glycerol and stored at −70 °C.

SIV PR Inactivation with EPNP. EPNP, an active site directed irreversible inhibitor of aspartyl proteases (Tang, 1971), was purchased from Sigma and was purified by recrystallization in methanol. HIV-1 PR (15 μg/mL final concentration) and HIV-2 or SIV PR (24 μg/mL final concentration) were preincubated at 25 °C in buffer containing 250 mM sodium acetate, pH 5.4, 500 mM NaCl, 1 mM EDTA, and 10% DMSO in the presence of 1–10 mM EPNP; 500 μM DTT was added for the HIV-1 enzyme. Baseline measurements were carried out with 10% DMSO in the absence of inhibitor. At eight different times, aliquots were removed and assayed for activity. Proteases were assayed with the fluorescent substrate ABZ-Thr-Ile-Nle-Phe(p-NO_2)-Gln-Arg-NH_2 under conditions described previously (Toth & Marshall, 1990). For the calculation of the rates of inactivation, kinetic data were fit to a pseudo-first-order equation. Eight different inhibitor concentrations were used to calculate the maximum inactivation rate and $K_{inactivation}$ (K_{inact}) for each protease. K_{inact} is the concentration of inhibitor that results in half the maximal rate of inactivation. Each determination of K_{inact} was carried out in duplicate.

SIV Labeling with EPNP. SIV PR (300 μg) was treated with 10 mM EPNP at 25 °C in 1 mL of 250 mM sodium acetate, pH 5.4, containing 500 mM NaCl, 1 mM EDTA, and 10% DMSO. After 2 h, the fully inhibited enzyme was isolated by reversed-phase HPLC on a Vydac C_3 column using a 100-min linear gradient from 5 to 65% acetonitrile in 0.1% TFA. The peak of protein identified by absorbance at 220 nm was collected, concentrated under vacuum, and dissolved in a small volume of 50% acetonitrile and 1% acetic acid. Molecular weights of protease and protease–EPNP complexes formed in solution were determined by mass spectrometry, using a model Bio-Q VG electrospray mass spectrometer. Crystals of SIV PR S4H inhibited with EPNP were dissolved in pH 5.5 sodium acetate buffer (50 mM) and analyzed in an identical fashion.

Crystallization. Previously stored samples were thawed rapidly to room temperature using a water bath, and the pH was adjusted by dilution with 1 M sodium acetate, pH 5.4, to a final concentration of 300 mM sodium acetate; 10 mM EPNP was added and incubated for 3 h. Precipitated EPNP was removed by centrifugation in a microfuge. The protein was dialyzed and concentrated in an 8-mL collodion membrane (Schleicher and Schuell), using a dialysis buffer of 20 mM sodium acetate, pH 5.4, and 50 mM NaCl. Final protein concentration for crystallization was 2.5–3 mg/mL as measured by the Bio-Rad Bradford assay using IgG as a standard.

Table I: Statistics on Data Collection and Refinement

resolution (Å)	2.4
R_{sym}[a] (%)	5.4
completeness to 2.5 Å (%)	92
completeness of 2.5–2.25 Å resolution shell (%)	77
total no. of reflections $>1\sigma$	21 880
no. of independent reflections	8 824
search model based on Swain et al. (1990)	7HVP
final R factor (7–2.4 Å) (%)	19
rms deviation of bond lengths (Å)	0.015
rms deviation of bond angles (deg)	3.2
rms deviation of dihedral angles (deg)	28.3
Wilson B factor	23

[a] $R_{sym} = \sum_i\sum_j|\langle I_i\rangle - I_{ij}|/\sum_i\sum_j I_{ij}$. Data reported for $I > 1\sigma$.

Crystals were grown using the hanging drop vapor diffusion method. Four microliters of protein solution was mixed with 4 μL of well buffer: 4–6% saturated NaCl (230–330 mM) and 100 mM sodium cacodylate, pH 6.5. Large crystals (0.4 × 0.15 × 0.05 mm) were grown consistently by macroseeding pre-equilibrated drops. Crystals were plates with rhomboid faces with two opposite corners truncated by parallel sides.

Structure Determination. The space group and cell dimensions of the SIV PR S4H/EPNP crystals were determined by precession photography. I_{hkl} data were collected with an R-axis II image plate detector using X-rays generated from a Rigaku 18 KW generator. Data were reduced using software provided by Rigaku. A summary of the data collection and refinement statistics is provided in Table I. Initial phases were determined by molecular replacement using a model that was built from an HIV-1 PR structure, 7HVP (Swain et al., 1990). All nonidentical residues between SIV and HIV-1 protease were changed to alanines to generate the search model (41 of 99 residues). Positions in the search model corresponding to glycines in either the HIV-1 or SIV PR structure were defined as glycines (there are three nonconserved glycines). I62 and I66 in the HIV-1 PR structure were changed to V62 and V66 as in the SIV PR structure. The 10 flap residues, 46–55, were removed because the conformation of the flaps can be highly variable (Navia et al., 1989; Wlodawer et al., 1989; Swain et al., 1990; Rutenber et al., 1993). Because the asymmetric unit contained one protease monomer, a monomer was used as the initial search model.

A rotation search using the above model was carried out using I_{hkl} data between 15- and 3-Å resolution and a maximum vector length of 30 Å. Patterson correlation refinement distinguished the correct solution from others. In fact, the correct solution corresponded to the highest peak from the initial rotation search. The solution obtained was confirmed in part since it reconstructed the expected association of monomers in the protease dimer of HIV-1 by orientation relative to one of the crystallographic twofold axes. A translation search was carried out using F_{hkl} data between 8- and 3-Å resolution. A grid size of 1 Å was used. After rigid body minimization of the top 1000 translation search solutions, the R factor for the top solution was 47%. The R factor for the next best search solution was 50%. Molecular replacement calculations were carried out using the program XPLOR (Brunger et al., 1987).

The search model was refined using XPLOR (Brunger et al., 1987) with manual rebuilding into $(2|F_o| - |F_c|),\alpha_c$ and $(|F_o| - |F_c|),\alpha_c$ electron density maps displayed using the molecular graphics programs FRODO (Jones, 1985) and CHAIN (Sack, 1988). After refinement to an R factor of 0.28, individual isotropic B factors were added. EPNP was

added to the model at an R factor of 0.27. Individual water molecules were added if density appeared in difference maps calculated using terms ($|F_o| - |F_c|$),α_c and if they formed reasonable hydrogen bonds to the protein (Finer-Moore et al., 1992). In the final structure there was no density for four side chains, all of which are on the protein surface and are probably disordered.

Electron density maps computed using terms ($2|F_o| - |F_c|$),α_c showed equivalent density for EPNP at each of the two active site aspartic acids, which corresponded to half-occupancy at each carboxyl. This could correspond either to only one molecule of EPNP bound per dimer—bound statistically to either site—or to statistical occupation of each site in the dimer independently. In either case, the dimeric molecule becomes *asymmetric*. However, molecules are statistically distributed, 50:50, in the crystal; thus, *statistically*, the space group reflected in the data is $C222_1$, though any particular single unit cell may not present this symmetry. To allow for the refinement of two equivalent half-occupied EPNP sites that might overlap during refinement if there were only one per dimer, interactions between monomers were not included during refinement. To assess whether there was some asymmetry in the protein due to a single bound EPNP, possibly expected in the region of the EPNP, refinement of the dimer as the asymmetric unit in a lower symmetry space group ($C2$) was carried out. At this resolution no asymmetry is yet apparent; thus, all final refinement was carried out using the monomer plus 0.5 EPNP as the asymmetric unit in space group $C222_1$.

RESULTS

Production of SIV PR. Expression and characterization of $SIV_{mac}BK28$ was as described by Grant et al. (1991). The protease clone used in this study was of $SIV_{mac}239$, which differs from $SIV_{mac}BK28$ by the substitution of Glu for Lys at position 63. This results in a significant change in the isoelectric point of the enzyme, which is predicted to be 5.1 rather than the observed value of 8.2 for $SIV_{mac}BK28$. The pI calculation used the program PREDICT (available on request). This predicted pI of 5.1 is essentially the same as that predicted for HIV-2 PR, and a similar purification protocol to that described previously for HIV-2 was effective (Rosé et al., 1993).

SIV PR S4H Mutant. Crystals were initially grown using wild-type SIV PR. However, out of concern that autoproteolysis prior to inhibition with EPNP might affect sample homogeneity, a variant of SIV PR was made to prevent cleavage at the primary autolysis site. This variant, SIV PR S4H, did not significantly affect the kinetic parameters of the enzyme but stabilized it approximately 4-fold to autolysis (Figure 2). The loss of activity seen with the variant did not correlate with the formation of autoproteolytic fragments as monitored by SDS–PAGE (data not shown). In addition, the decay curve for SIV PR S4H is fit better by a first-order than a second-order rate equation, suggesting that a process other than autoproteolysis is responsible for the observed loss of activity. It is possible that this is due to oxidation of methionine residues or thermal denaturation during the course of the assay. No difference density for the oxygens was seen for the methionines (residues 76 and 95). In crystallization trials the SIV PR S4H variant produced large crystals consistently, which were used for data collection. These crystals appeared to have the same morphology as those obtained from wild-type SIV PR.

EPNP Inactivation. The maximum rates of inactivation (V_{max}) of the SIV, HIV-1, and HIV-2 PRs by EPNP were similar: approximately 0.05 min^{-1} (Table II). The similarity indicates that the mechanism of inactivation is the same for the three proteins. The stoichiometry of protease modification was determined by electrospray mass spectrometry. Samples of inactivated SIV PR in solution and of inactivated SIV PR S4H from crystals contained two major species which could be isolated by HPLC (Figure 3). The areas under peaks I and II were roughly equal for both SIV PR and SIV PR S4H, indicating that half of the protease monomers had been modified in each case. Material from peaks I and II was

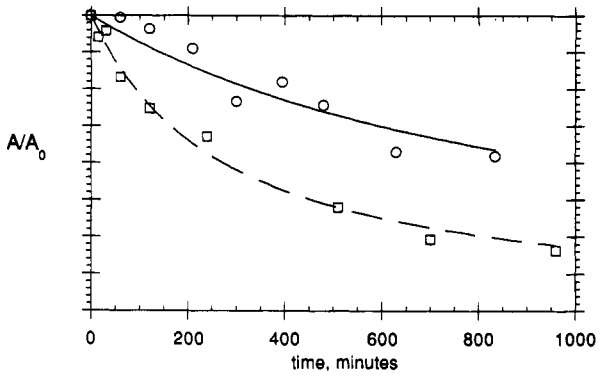

FIGURE 2: Comparison of decay curves for SIV PR (□) and SIV PR S4H (○). Starting protease concentration was 40 ng/μL. A/A_0 represents the ratio of active protein at a given time (A) to the total active protein at time = 0 (A_0). Decay curves were generated by assaying remaining protease activity at various time points and fit to either a first-order $A = A_0e^{-kt}$ or a second-order $A = A_0/(1 + A_0k_2t)$ rate equation, where t is time and k and k_2 are the rate constants for inactivation.

Table II: Inactivation of HIV and SIV PRs by EPNP

protease	V_{max}[a] (min^{-1})	K_{inact}[b] (mM)
HIV-1	0.060 ± 0.006	9.85 ± 0.98
HIV-2	0.048 ± 0.004	6.71 ± 0.58
SIV	0.060 ± 0.002	8.03 ± 0.24

[a] Maximum inactivation rate. [b] EPNP concentration at half-maximal inactivation rate, $V_{max}/2$.

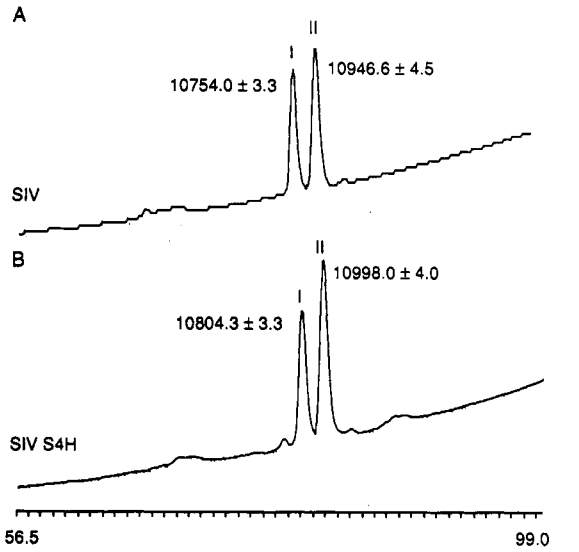

FIGURE 3: Stoichiometry of EPNP modification. SIV PR that had been inactivated in solution (A) or inactivated and crystallized SIV S4H (B) was analyzed by HPLC at 220 nm on a Vydac C_3 column using a 100-min linear gradient from 5 to 65% acetonitrile in 0.1% TFA. Molecular weights for peaks I and II were determined by electrospray mass spectrometry. The areas under the peaks were 3.92×10^5 (I) and 4.91×10^5 (II) for (A) and 4.22×10^5 (I) and 6.00×10^5 (II) for (B).

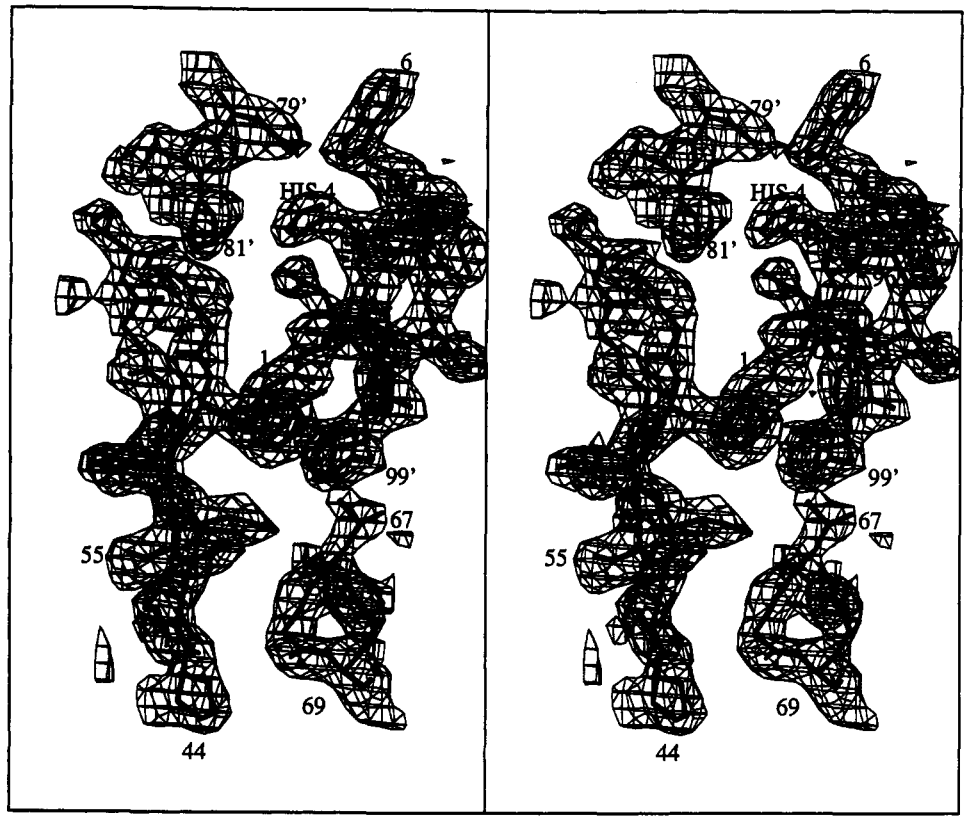

FIGURE 4: Electron density in a $(2|F_o| - |F_c|),\alpha_c$ map showing the crystal contacts formed near the flap residues, 44–55. Included in the figure are the dimer-related residues, 79′–81′, and residues from the C-centered related dimer: 1–6, 67–69, and 96′–99′. Density for the substitution S4H can be seen. Density for the side chains is well-defined, and main-chain carbonyl oxygens are visible. The density is contoured at 0.9 σ from the XPLOR output.

analyzed by electrospray mass spectrometry to determine the molecular masses of the two species. The masses of the two species differed by 194 daltons, the predicted molecular mass of EPNP. Thus, one molecule of EPNP was covalently added to each functional dimer.

Overall Structure. The crystal contained one monomer of the protease per asymmetric unit, with the twofold axis of the PR dimer coincident with the crystallographic *a*-axis. The space group was determined to be $C222_1$ with unit cell dimensions of a = 62.7 Å, b = 32.2 Å, and c = 96.1 Å, with an error of ±0.1 Å. In addition to the protease, the final structure contained one molecule of EPNP per protease dimer and 25 water molecules. The atomic coordinates and temperature factors have been submitted to the Brookhaven Protein Data Bank (1Sam). Electron density computed using terms $(2|F_o| - |F_c|),\alpha_c$ for the "flaps" (residues 46–55) and the C- and N-terminal residues is shown in Figure 4. Density for the His 4 side chain, the point mutation, is well-defined in the figure.

The overall fold of SIV PR follows that of HIV-1 PR, as can be seen in the superposition of SIV PR and 7HVP in Figure 5. A core of 53 residues was identified in the SIV PR dimer using difference distance matrices (Perry et al., 1990) to identify regions of the structure whose α-carbon positions overlap most closely with the HIV-1 PR structure, 7HVP. These 53 residues were used to superimpose the two structures using the program Gem [written by E. B. Fauman (University of California at San Francisco, 1993) available on request], resulting in an rms deviation of 0.24 Å for these 53 residues. With this alignment, the overall rms deviation for the α-carbon backbones of the two protease dimers is 1.16 Å. Figure 5 shows the dimers divided into three regions. The backbones of the central region are the most conserved. Excluding the three amino-terminal residues from each monomer, which have different conformations, the rms deviation for this central region is 0.64 Å. This core of the protein contains the active site aspartic acids, D25 and D25′ (primes indicate the symmetry-related monomer of the dimer), and the flaps comprising residues 46–55 and 46′–55′. The outer regions show the most deviation: β-strand residues 11–22, rms deviation of 1.2 Å; random coil residues 33–45, rms deviation of 2.1 Å; β-strand residues 63–75, rms deviation of 1.7 Å.

The positional difference of the 13 residues in contact with the peptidomimetic inhibitor in the HIV-1 structure, 7HVP (R8, L23, D25, G27, A28, D29, V32, I47, G48, I50, P81, V82, I84) in the HIV-1 structure (Miller et al., 1989), and the corresponding residues in SIV PR S4H is recorded in Table III. Three of these residues differ between these two sequences: at residues 32, 47, and 82. To assess the significance of these positional differences, Table III also shows the standard deviation of the position expected when atomic coordinates of two independently refined but otherwise identical structures are compared (Fauman, 1993), as an extension of the methods developed in Chambers and Stroud (1979) and Perry et al. (1990). This error is an empirically derived function of the B factor (B): $p_1 + p_2 \exp(B/p_3)$, with p_1, p_2, and p_3 functions of the number of atoms refined in the structure divided by the number of reflections measured in the data. In this structure determination the standard deviation of atomic position, SD, is calculated

$$\mathrm{SD} = 0.15 + 0.012 \exp(B/10.99)$$

Density for the flaps was clearly seen in the original maps after being omitted in the molecular replacement search model. The flaps were found to be in a "closed" conformation, similar to the conformation in the 7HVP complex which contained

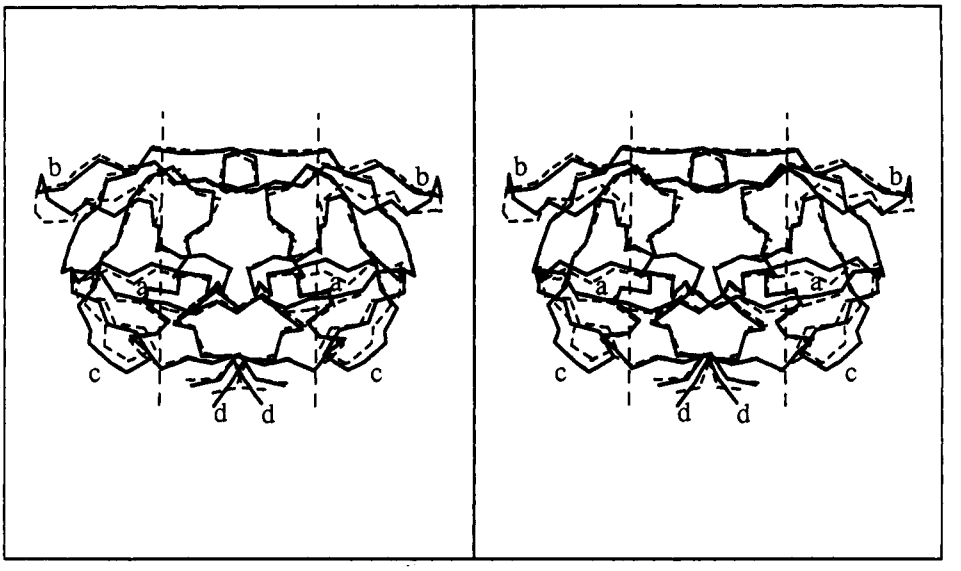

FIGURE 5: Backbone trace of SIV PR (solid lines) superimposed on that of HIV-1 PR (dashed lines). Superposition was carried out using a core of 53 residues as described in the text. The dotted vertical lines delineate the central region of the alignment in which the backbone structures are almost identical. The loops labeled in the outer regions show the most variation in position: (a) residues 11–22; (b) residues 33–45; (c) residues 63–75. The conformation of the N-terminal three residues (d) is also noticeably different.

Table III: Comparison of the α-Carbon Positions of the Substrate Binding Pocket Residues in SIV S4H PR and HIV-1 PR, 7HVP

residues[a]	Δr (1–99)[b] (Å)	Δr_ϵ[c] (1–99)[b] (Å)	Δr (101–199)[d] (Å)	Δr_ϵ (101–199)[c] (Å)
R8	0.2	0.4	0.4	0.3
L23	0.5	0.3	0.3	0.3
D25	0.2	0.3	0.6	0.3
G27	0.3	0.3	0.4	0.3
A28	0.2	0.3	0.5	0.3
D29	0.2	0.4	0.6	0.4
I32(V)	0.5	0.3	0.5	0.3
V47(I)	0.6	0.5	0.5	0.5
G48	0.2	0.5	0.7	0.5
I50	0.9	0.7	0.4	0.7
P81	**1.2**	**0.5**	**1.2**	**0.5**
I82(V)	**1.0**	**0.4**	**1.3**	**0.4**
I84	0.1	0.4	0.1	0.4

[a] Residues are from the SIV PR sequence, with the HIV-1 PR residue in parentheses when different from the SIV PR sequence. [b] Deviations between SIV PR and the monomer of 7HVP with residues numbered between 1 and 99. [c] Δr_ϵ (in columns 3 and 5) is an estimate of the standard deviation in atomic position if the two structures were identical. [d] Deviations between the symmetry-related monomer of the dimer, 1′–99′ in SIV PR, and the monomer of 7HVP with residues numbered between 101 and 199. In bold are the residues whose positions vary significantly relative to the estimated error.

a peptidomimetic inhibitor. The flaps form a β-strand which is initially parallel to the protein surface and which turns 90° with a right-handed twist. The tips of the dimer-related flaps are parallel to each other and are within 3.2 Å. The rms deivation of the flaps between the SIV PR and 7HVP is 0.78 Å. The largest positional difference, 1 Å, is between the α-carbon atoms of G49. The phenyl ring of the F53 side chain in the flaps of SIV PR has a significantly different conformation from that found in the 7HVP structure, facing into solvent instead of parallel to the protein surface.

The flaps participate in a crystallographic contact with other protease dimers related by C-centering in the unit cell: one dimer contacting each of the flaps. Three crystallographically related N- and C-terminal side chains, F99, F3, and P1, are within van der Waals contact of F53 and I46 side chains in the flaps (Figure 4). The four rings of F99, F3, P1, and F53 pack together, within 3.4–4.1 Å. The I46 side chain in the flaps is 2.9 Å from the carbonyl oxygen of L67. This crystallographic contact extends beyond the flaps to the side chain of L67, which is in van der Waals contact with the crystallographically related phenyl ring of F99. The carboxyl oxygens of D79 are 3.7 Å from a nitrogen in the histidine ring of H4.

The side chains of the active site aspartic acids, D25 and D25′, are more skewed in the SIV PR S4H structure than in the 7HVP structure. If a plane is defined by the Cα and Cγ atoms of each of the aspartic acid side chains, then the angle made by the C$_\gamma$–Oδ bond with the plane describes their deviation from planarity (Figure 6c). In the 7HVP structure, this angle is +13° for Oδ1 and –4° for Oδ2. In the SIV structure these angle are +36° for Oδ1 and –16° for Oδ2. In the unliganded form, a hydrogen bond exists between D25-Oδ1 and D25′-Oδ1. The unfavorable geometry in the SIV PR S4H structure suggests that the hydrogen bond between the active site aspartic acids has been disrupted as a result of covalent modification with EPNP. A hydrogen bound to Oδ1 of one of the aspartic acids is 3.1 Å from the other aspartic acid's Oδ1, making an angle of 89° for Oδ1–H–Oδ1.

Water Molecules. A total of 25 ordered water molecules have been resolved in the structure. Water 301 in the 7HVP structure is conserved in the SIV PR structure, hydrogen-bonded to the amide nitrogens of I50 and I50′ of both flaps of the dimer (Figure 6a). This water is 0.5 Å from the site occupied by water 301 in 7HVP and is 3.05 Å from the amide nitrogens to which it is bound. The *B* factor for this water is low, 9.25, while the *B* factor for the backbone nitrogens is 40. Water 305, which appears in the unliganded structure between the two catalytic aspartic acids (Navia et al., 1989; Wlodawer et al., 1989), has been displaced by EPNP. Two other well-ordered water molecules appear in the binding cavity of the SIV structure: water 307 and water 313. Water 307 is 3.0 Å from the backbone amide of D29. Water 313 is 3.2 Å from the backbone amide of G48. Both waters occupy positions near where the peptidomimetic inhibitor in 7HVP forms hydrogen bonds (the backbone nitrogen of G48 is hydrogen-bonded to the carbonyl oxygen of S$_4$ or S$_3$′ in 7HVP, and the backbone nitrogen of D29 is hydrogen-bonded to the carbonyl oxygen of S$_3$ or S$_2$′ in 7HVP). The remaining ordered

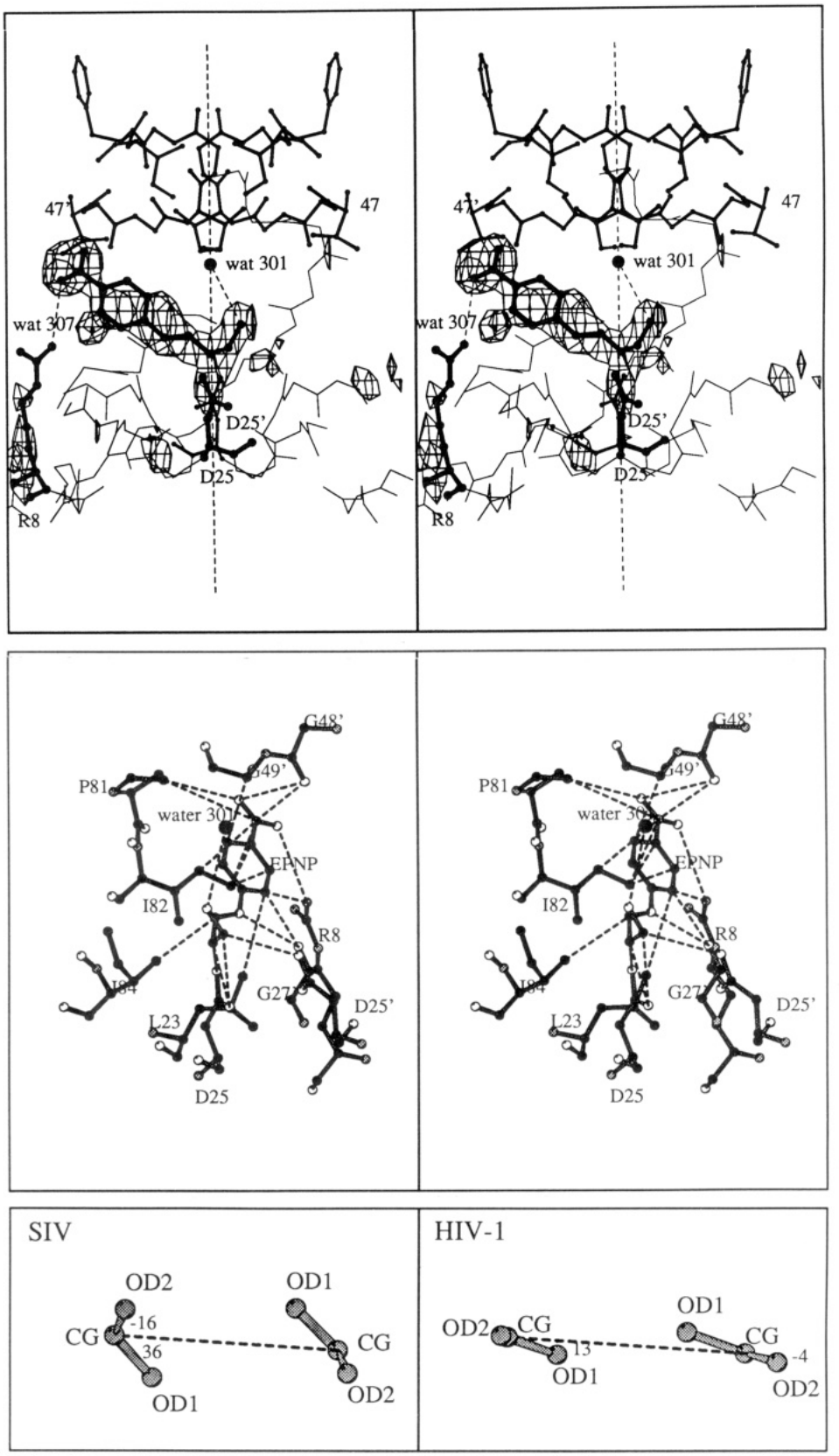

FIGURE 6: (a, top) Active site residues of SIV PR showing the inhibitor, EPNP, covalently bound to ASP 25. Difference density contoured at 1.5σ is shown after the EPNP and water 307 were omitted from the model and a "slow cool" was performed from 500 to 300 K. Drawn in thick lines are EPNP, the active site aspartic acids, and R8, which contacts the nitro group of EPNP. In thinner dark lines are the two flaps of the dimer. Also displayed is water 301, which is hydrogen-bonded to the amide backbone of I50 and I50′ of the flaps. Dotted lines to EPNP show interactions: water 301 is 3.7 Å from the EPNP hydroxyl and the guanidinium group of R8 is 3.6 Å from the nitro group of EPNP. The thin lines show the backbone trace of nearby residues. Because of the low contour level, some noise also appears in the difference map. The vertical dotted line is along the dimer axis. (b, middle) Residues within van der Waals contact (dotted lines) of EPNP in the SIV PR active site. White atoms are oxygens, gray atoms are nitrogens, and black atoms are carbons. (c, bottom) Angles formed by the carboxyl oxygens of the active site aspartic acids with a plane joining the other active site aspartic acid. These angles are formed by the Cγ–Oδ bonds and a plane through the four atoms D25 Cα, D25 Cγ, D25′ Cα, and D25′ Cγ. The angles in the SIV PR structure with EPNP bound are much greater than in the HIV-1 peptidomimetic inhibitor bound structure (7HVP).

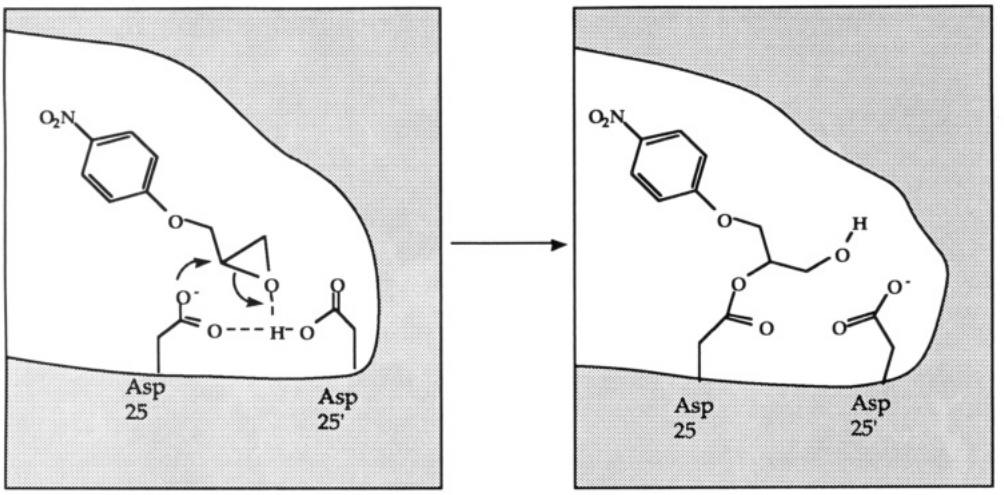

FIGURE 7: Proposed reaction mechanism of EPNP with one active site aspartic acid. The two active site aspartic acids initially share a proton. This proton is donated to the epoxide oxygen while the aspartyl oxygen attacks the secondary carbon of the epoxide ring (the C2 carbon of the propane moiety of EPNP). No reaction is observed at the primary carbon of the epoxide ring (the C1 carbon of the propane moiety of EPNP).

waters in the SIV structure are not in the binding cavity.

Inhibitor Binding Mode. Figure 6a shows difference density in the active site of SIV PR with EPNP and water 307 omitted from the model and after a "slow cool" from 500 to 300 K (Kirkpatrick et al., 1983). EPNP is shown built back into the model. With the occupancy of each orientation of EPNP set to 50%, a difference map showed no density. In addition, when the occupancy was refined, it refined to a value of 40% for each EPNP site. This is consistent with an occupancy of 50% for EPNP bound to each of the active site aspartic acids. Similar density is observed near the other active site aspartic acid of the dimer, though this is not shown in Figure 6a. Chromatographic and mass spectrometric studies of crystallized material confirmed that only one molecule of EPNP was present per dimer.

A proposed mechanism of the covalent attachment of EPNP to aspartic acid 25 is shown in Figure 7. The negatively charged oxygen of the active site aspartate attacks the C2 carbon of the propane moiety of the epoxide ring, forming an ester linkage. This mechanism is consistent with the difference density observed in Figure 6a.

Figure 8 shows both possible positions of EPNP in the active site of SIV protease, viewed down the twofold axis. The position where the peptidomimetic inhibitor from the 7HVP structure would be is also shown for comparison. From this figure, the EPNP is seen to be oriented parallel to the backbone of a bound peptide, with the phenyl group near the P1 (or P1′) sites. This figure also shows that the protease binding cavity would also accommodate EPNP if it were rotated 180 ° so that the phenyl ring occupied the position of the S2 or S2′ subsites. When two EPNP molecules were built into the model in these additional conformations and the occupancy was allowed to refine, the occupancy of the alternate binding mode refined to zero.

Figure 6b is a more detailed view of the contacts between EPNP and the protein. Nonspecific van der Waals interactions are made with residues 23, 27, 48, 49, 81, 82, and 84. The carbonyl oxygen of G27 forms a weak hydrogen bond with the oxygen of the phenoxy moiety of EPNP, 3.1 Å away. The terminal hydroxyl of EPNP, generated by opening of the epoxide ring, is oriented toward water 301, 3.7 Å away. In addition, the guanidinium group of R8 also contacts the nitro group of EPNP, 3.6 Å away. This interaction is too far for a hydrogen bond, but at the pH used for crystallization, pH 6.5, the guanidinium group is protonated and forms an electrostatic interaction with the nitro group oxygens.

DISCUSSION

Characteristics of the SIV PR S4H Structure. As anticipated, the overall fold of SIV PR is identical to that of HIV-1 PR. Gustchina et al. modeled the structure of HIV-2 PR by building from the structure of HIV-1 PR (Gustchina et al., 1991). The SIV PR S4H sequence is 87% identical to that of HIV-2 PR and 51% identical to HIV-1 PR. The model predicted an rms deviation for 185 of 198 residues in the dimer to be 0.28 Å. The rms deviation for the Cα positions between SIV PR and 7HVP is actually 1.16 Å. Most of this difference derives from two β-strands and a random coil loop farthest from the binding pocket and on the surface of the protein.

When the positions of the Cα atoms of the 13 residues forming the peptide binding pocket in HIV-1 PR (7HVP) were compared with their positions in SIV-1 PR, there were significant differences for residues P81 and I82 (see Table III). I82 is a Val in HIV-1 PR, which may contribute to its positional difference. Binding of the different inhibitors in the SIV structure and in the HIV-1 structure may account for some differences.

Unlike the peptidomimetic-bound HIV-1 structure 7HVP, the asymmetric unit of the SIV PR crystals contained one monomer of the functional dimer. The symmetry of the protease dimer allows the dimer twofold axis to be a crystallographic twofold axis. Although EPNP binding to one of the active site aspartic acids introduces local asymmetry in the dimer, the symmetry of the crystal is not affected because each aspartic acid of the dimer is modified with equal probability. A larger asymmetric inhibitor could presumably influence the crystal packing and reduce the symmetry of the crystal.

The monomer subunits of the 7HVP structure are not identical. When the α-carbon backbones are optimally superimposed, the rms deviation of the Cα atoms is 0.6 Å with a rotation of 178.6° instead of the 180° required by a crystallographic twofold axis. (If rotated by 180°, the rms deviation between the Cα atoms of the two monomers is 3.9 Å). Much of this asymmetry can be attributed to the peptide-based inhibitor, which is asymmetrical. In addition, an

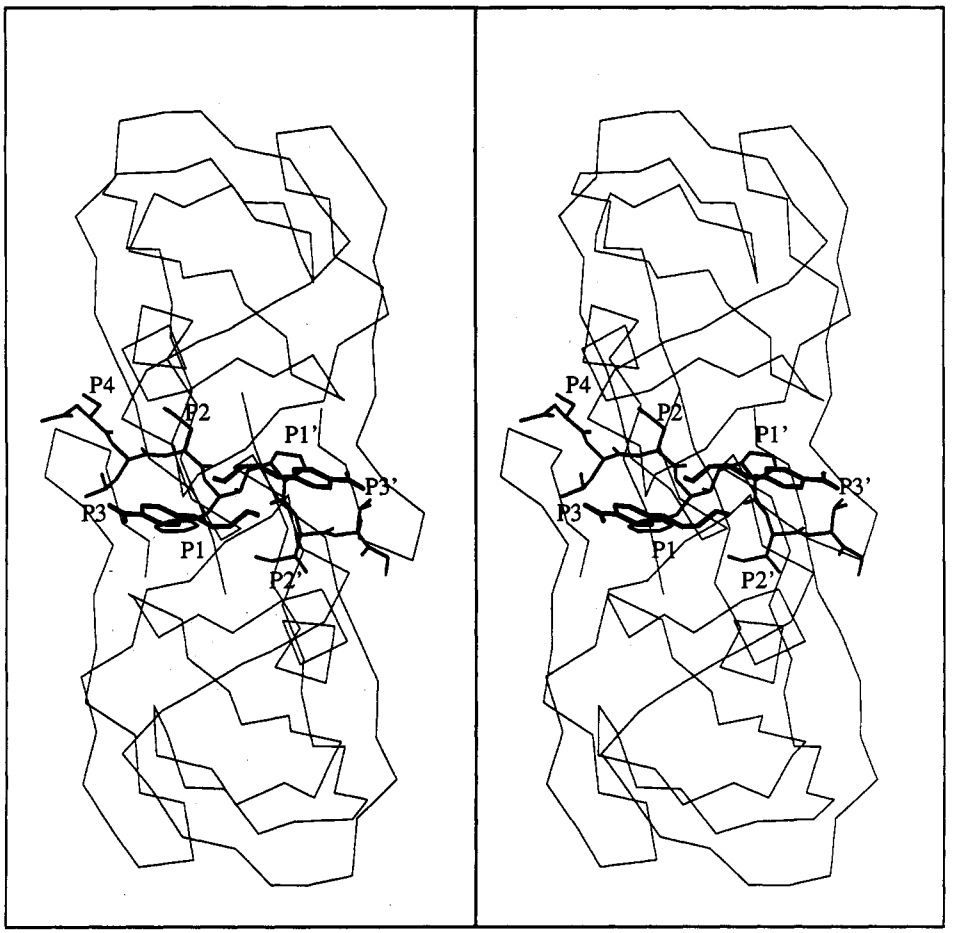

FIGURE 8: α-Carbon trace of the SIV PR dimer looking down the twofold axis. Two EPNP molecules are shown as they are oriented in the active site when bound to each of the two aspartic acids (white in black outline). Superimposed is the peptidomimetic inhibitor that was crystallized with HIV-1 in the 7HVP structure (in bold black). (The superposition uses the Cα postions of the 53 core residues, as described in the text). The P4 to P3′ peptide binding sites are labeled.

asymmetrical hydrogen bond is formed between the flaps of the dimer: the amide nitrogen of G51 donating the hydrogen to the carbonyl oxygen of I50′ (I150 according to the 7HVP numbering). This is only possible because the carbonyl of I50′ points toward its dimer-related flap, while the carbonyl of I50 points away from its dimer-related flap. In the SIV structure both of these carbonyl oxygens point away from their dimer-related flaps. No density can be seen for the other conformation of the carbonyl, even if it is built into the model.

The histidine side chain present in SIV PR S4H may have had two effects on the crystallization of the protease. The mutation clearly stabilizes the protein to autolysis, thereby increasing sample homogeneity which may facilitate better crystal formation. H4 also makes up part of the surface involved in a crystal contact. The His side chain points into solvent and participates in the crystallographic contact surface formed by the N- and C-terminal β-sheet. It is therefore possible that the S4H mutation influenced the ease of crystallization of SIV PR.

The flaps of the EPNP-bound SIV structure were in their fully closed position, as in the 7HVP structure which has a peptidomimetic inhibitor bound (Swain et al., 1990). Interactions between EPNP and the flaps are weak, consisting of van der Waals interactions, and yet the conformation of the flaps is closed: the Cα of I50 shifts by 0.4–0.9 Å from the 7HVP position. It is possible that the conformation of the flaps is influenced by crystallographic contacts. Though the present crystal packing arrangement would accommodate the flaps in the open conformation, the close van der Waals interactions in the present structure may form more favorably in the flaps-closed conformation, thus favoring this conformation during crystal growth. The conformation of the flaps in the HIV-2 PR structures with peptide-based inhibitors appears also to be closed, very similar to the SIV PR structure (Mulichak et al., 1993).

Another difference in structure is the conformation of the loop preceding the flaps, residues 33–45, which may influence the conformation of the flaps. The position of the prolines in SIV PR differs in this loop from HIV-1 and HIV-2 PRs: residue 40 is a Pro in SIV PR, residue 39 is a Pro in HIV-1 PR, and there is no Pro at either of these positions in HIV-2 PR. All three sequences have a conserved Pro at residue 44. Pro 40 is responsible for the large difference in position of this loop in the SIV PR structure relative to the HIV-1 structure.

The position of the tightly bound water 301 (*B* factor of 9), hydrogen-bonded between the flaps, is the same as that seen in the HIV structures. The other waters in the SIV PR S4H binding pocket, 307 and 313, occupy positions near where a bound peptide would form hydrogen bonds. We would expect the carbonyl oxygens and amide nitrogens of the protein to be hydrogen-bonded to waters if possible when they are not interacting with a peptide, though no ordered water is observed near most of these in the EPNP-bound structure.

The conformation of the N-terminal three residues of SIV PR is shifted away from the body of the protein relative to the conformation in HIV-1 PR (7HVP): the Cα of the N-terminal proline is 1.9 Å from its position in 7HVP, and the Cγ position is 3.5 Å from its position in 7HVP. These

three residues participate in a four-stranded β-sheet with the other N-terminal and C-terminal residues of the dimer. Protein packing may influence the conformation of these residues. Hydrogen bonding is not disrupted in the β-sheet configuration in the SIV structure. In contrast, two hydrogen bonds are disrupted with the backbone amide and carbonyl of F3 in the HIV-2 PR structure (Mulichak et al., 1993). The N-terminal residues also participate in crystal contacts in this structure.

Binding Mode of EPNP. EPNP is a specific inhibitor of aspartyl proteases which labels the active site aspartic acids (Tang, 1971). Epoxides undergo ring opening by nucleophilic attack at either the primary or the secondary carbon. At neutral or basic pH, the less substituted carbon is attacked more readily. At acidic pH, the epoxide is protonated and the secondary carbon forms the more stable carbocation and is preferentially attacked (March, 1985). Nucleophilic attack on 1,2-epoxides in acidic solutions is known to occur preferentially at the 2-position of the propane moiety (Parker & Isaacs, 1959).

The only other crystal structure of EPNP complexed with an aspartyl protease is of penicillopepsin (James et al., 1977). This structure revealed three sites where EPNP bound to the protein: one noncovalently associated with the protein outside the active site and one bound to each of the active site aspartic acid residues. From the electron density map, James et al. concluded that one aspartic acid reacted at the C2 position of the propane moiety, while the other aspartic acid reacted at the C1 position. The difference density in the SIV PR S4H structure shows EPNP bound at the C2 position of the propane moiety. The density around the EPNP is consistent with there being one EPNP molecule bound per protease dimer.

To catalyze the ring opening of the epoxide at the C2 position, the protease must satisfy three conditions: A strong nucleophile is required; this is provided by the active site aspartic acid. The EPNP must be positioned favorably for a reaction to occur (Tang, 1971), and a proton donor must be available.

At pH 5.4 one of the SIV PR active site aspartic acids is protonated and the other is unprotonated (Grant et al., 1991). A hydrogen shared between the two aspartic acids raises the pK_a of the second proton from this pair of aspartic acids. Therefore, initially only one aspartic acid is available for a reaction with EPNP. Because the secondary carbon of the epoxide ring becomes covalently linked to the aspartic acid, the epoxide oxygen must be protonated. An appealing source for this proton is the hydrogen that was shared between the active site aspartic acids (see Figure 7). This hydrogen bond is clearly disrupted after the EPNP reacts, as demonstrated in Figure 6c. After the initial EPNP reacts, the second aspartic acid is accessible to solvent, and its pK_a probably reduces to its solvent-exposed value of 4.0. At pH 5.4 it is therefore predominantly unprotonated and could therefore act as a nucleophile. A second EPNP does not react at this site, perhaps because no proton is available to donate to the epoxide oxygen. Also, the positioning of the C2 carbon of a second EPNP at this aspartic acid is sterically unfavorable: the C1 carbon of the first EPNP would then be only 2.5 Å from the C1 carbon of the second EPNP. Reaction of the second EPNP at the C1 position of the propane moiety is not sterically prohibited but is not observed. An aspartic acid is not generally a strong enough nucleophile to perform this reaction, and it is unclear why the reaction occurs in the case of penicillopepsin.

EPNP orients itself in the binding pocket of the protease parallel to where a peptide backbone would lie, with the phenyl ring reaching into the S1 (or S1′) site. Tang proposed that EPNP reacts with pepsin by occupying the hydrophobic binding site which would be occupied by a substrate (Tang, 1971). HIV and SIV proteases prefer hydrophobic residues in the S1 and S1′ subsites (Poorman et al., 1991). The S1 and S1′ binding pockets are composed of residues L23, D25, G27, I50, P81, V82, and I84 in HIV-1 (Miller et al., 1989). All of these except I50 form van der Waals contacts with EPNP, with distances very similar to those found for the P1 and P1′ sites in the 7HVP peptidomimetic-bound structure. The side chain of I50 is out of reach of the EPNP. Residue 82 is an Ile in the SIV sequence and makes a closer contact with EPNP than there would be if it were a Val, as in HIV-1.

Other than the reactive aspartic acid, the only specific interactions with EPNP are weak: the phenoxy oxygen of EPNP and the carbonyl oxygen of G27, and electrostatic attraction between the nitro group of EPNP and the guanidinium group of R8. R8 is associated with the S3 binding pocket (Miller et al., 1989). Water 301 is too far from the EPNP hydroxyl group to form a hydrogen bond. Its hydrogens are probably oriented toward the EPNP hydroxyl since its lone-pair electrons are hydrogen-bonded to the amide nitrogens of I50 and I50′ of the flaps. This would allow van der Waals interactions between the hydrogens of water 301 and the hydrogen of the hydroxyl of EPNP.

For a covalent inhibitor like EPNP, the K_{inact} reflects the noncovalent interactions with the protein. K_{inact} is the concentration of inhibitor that results in half the maximal rate of inactivation. The weakness of these interactions explains why K_{inact} is 8 mM.

This structure represents the first structure of a covalently modified protease closely related to the HIV-1 PR and can be used as a guide for drug design efforts. Starting from the known position of EPNP, derivatives may be designed in which further substituents are added onto the EPNP scaffold to take advantage of binding pockets in the enzyme. Enhancing the hydrophobic and van der Waals interactions which already exist, or building specific hydrogen-bonding interactions, should decrease the K_{inact}. Such substituents may increase both the affinity and specificity of these compounds of retroviral enzymes relative to cellular aspartyl proteases. In addition, a comparison of the conserved surfaces between binding pockets of the SIV and HIV-1 PRs can provide a template for the design of irreversible inhibitors with specificity for the HIV-1, HIV-2, and SIV proteases. Such inhibitors may be less susceptible to the development of resistance by the virus.

ACKNOWLEDGMENT

We acknowledge Earl Rutenber and Janet Finer-Moore for helpful discussions on crystallography and Julie Newdoll for assistance with graphics programs. Mass spectrometry was performed by Zhoughua Xu in the University of California San Francisco Mass Spectrometry Facility (A. L. Burlingame, Director), which is supported by National Institute of Environmental Health Sciences Grant ES 04705.

REFERENCES

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Cameron, C. E., Grindes, B., Jacques, P., Jentoft, J., Leis, J., Wlodawer, A., & Weber, I. T. (1993) *J. Biol. Chem. 268*, 11711–11720.

Chambers, J., & Stroud, R. M. (1979) *Acta Crystallogr. B35*, 1861–1874.
Fauman, E. B. (1993) Ph.D. Thesis, University of California, San Francisco.
Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T., & Stroud, R. M. (1992) *Proteins 12*, 203–222.
Grant, S. K., Deckman, I. C., Minnich, M. D., Culp, J., Franklin, S., Dreyer, G. B., Tomaszek, T. A., Jr., Debouck, C., & Meek, T. D. (1991) *Biochemistry 30*, 8424–8434.
Gustchina, A., Weber, I. T., & Wlodawer, A. (1991) in *Structure and Function of the Aspartic Proteinases* (Dunn, B. M., Ed.) pp 549–553, Plenum Press, New York.
James, M. N. G., Hsu, I., & Delbaere, L. T. J. (1977) *Nature 267*, 808–813.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Kestler, H., Kodama, T., Ringler, D., Marthas, M., Pedersen, N., Lackner, A., Regier, D., Sehgal, P., Daniel, M., King, N., & Desrosiers, R. (1990) *Science 248*, 1109–1112.
Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science 220*, 671–680.
March, J. (1985) in *Advanced Organic Chemistry*, Wiley, New York.
Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989) *Nature 337*, 576–579.
Mulichak, A. M., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Curry, K. A., Tomich, C., Thaisrivong, S., Sawyer, T. K., & Watenpaugh, K. D. (1993) *J. Biol. Chem. 268*, 13103–13109.
Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 337*, 615–620.
Parker, R. E., & Isaacs, N. S. (1959) *Chem. Rev. 59*, 7373–7399.
Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F., & Stroud, R. M. (1990) *Proteins 8*, 315–333.
Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., & Kézdy, F. J. (1991) *J. Biol. Chem. 266*, 14554–14561.
Richman, D. D. (1993) *Antimicrob. Agents Chemother. 37*, 1207–1213.
Rosé, J. R., Salto, R., & Craik, C. S. (1993) *J. Biol. Chem. 268*, 11939–11945.
Rutenber, E., Fauman, E. B., Keenan, R. J., Fong, S., Furth, P. S., Ortiz de Montellano, P. R., Meng, E., Kuntz, I. D., DeCamp, D. L., Salto, R., Rosé, J. R., Craik, C. S., & Stroud, R. M. (1993) *J. Biol. Chem. 268*, 15343–15346.
Sack, J. (1988) *J. Mol. Graphics 6*, 225.
Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805–8809.
Tang, J. (1971) *J. Biol. Chem. 246*, 4510–4517.
Toth, M. V., & Marshall, G. R. (1990) *Int. J. Pept. Protein Res. 36*, 544–550.
Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.