Three-Dimensional Structures of HIV-1 and SIV Protease Product Complexes[†,‡]

Robert B. Rose, Charles S. Craik,[§] Nancy L. Douglas,[§] and Robert M. Stroud*

Department of Biochemistry and Biophysics, S-960, The University of California, San Francisco, California 94143-0448

Received May 29, 1996; Revised Manuscript Received July 29, 1996[⊗]

ABSTRACT: Strain is eliminated as a factor in hydrolysis of the scissile peptide bond by human immunodeficiency virus (HIV)-1 and simian immunodeficiency virus (SIV), based on the first eight complexes of products of hydrolysis with the enzymes. The carboxyl group generated at the scissile bond interacts with both catalytic aspartic acids. The structures directly suggest the interactions of the gemdiol intermediate with the active site. Based on the structures, the nucleophilic water is displaced stereospecifically by substrate binding toward one catalytic aspartic acid, while the scissile carbonyl becomes hydrogen bonded to the other catalytic aspartic acid in position for hydrolysis. Crystal structures for two N-terminal (P) products and two C-terminal (Q) products provide unambiguous density for the ligands at 2.2−2.6 Å resolution and 17−21% *R* factors. The N-terminal product, Ac-S-L-N-F/, overlaps closely with the N-terminal sequences of peptidomimetic inhibitors bound to the protease. Comparison of the two C-terminal products, /F-L-E-K and /F(NO_2)-E-A-Nle-S, indicates that the P2′ residue is highly constrained, while the positioning of the P1′ and P3′ residues are sequence dependent.

HIV protease, essential to the life-cycle of the human immunodeficiency virus (HIV),[1] processes eight cleavage sites of the viral gag and gag-pol polyproteins (Henderson et al., 1988; Kohl et al., 1988; McQuade et al., 1990; Seelmeier et al., 1988). Retroviral proteases in general are homodimeric aspartyl proteases in which each monomer contributes one of the two catalytic aspartic acid residues, located inside a tunnel-shaped binding site. Binding of substrate leads to a large conformational change of the protease "flaps", flexible β-strand structures that clasp the substrate chain (James et al., 1982; Miller et al., 1989). The mechanism involves acid−base hydrolysis mediated by the two aspartic acids (James et al., 1977; Antonov et al., 1978, 1981; Polgár, 1987). The putative catalytic water hydrogen-bonded between the two aspartic acids is located in unliganded structures of monomeric and retroviral aspartyl proteases (Sielecki et al., 1990; Suguna et al., 1987a; Wlodawer et al., 1989). The nucleophilic water attacks the carbonyl carbon of the substrate to form a gemdiol intermediate. We report the configuration of the catalytic water in a modeled Michaelis complex and the configurations of the gemdiol intermediate, previously controversial.

Peptidomimetic inhibitors with tetrahedral geometry in place of the scissile carbonyl were thought to mimic the transition state (Wlodawer & Erickson, 1993), and many structures of these have been determined as part of the drug design effort against AIDS (Rutenber & Stroud, 1996). Difluoroketone inhibitors, which were also thought to mimic gemdiol intermediates, have been cocrystallized with different aspartyl proteases to interpret the mechanism (James et al., 1992; Parris et al., 1992; Silva et al., 1996; Veerapandian et al., 1992), leading to the idea that the gemdiol interacts with one catalytic aspartic acid preferentially. However, these inhibitors are not natural reactants. Thus, to probe the nature of the reaction mechanism, we determined the structures of eight complexes of the enzymes cocrystallized in the presence of their natural products.

These are the first structures of product complexes of aspartyl proteases with HIV-1 and simian immunodeficiency virus (SIV) proteases. They include two N-terminal and two C-terminal product complexes. Based on the active site interactions of the C-termini in the "N-terminal complexes", the model for the gemdiol intermediate interacts equally with both aspartic acids. Modeling the substrate from the products indicates that substrates bind to the active site without strain. A chemical mechanism for HIV protease is proposed based on models for the Michaelis complex and gemdiol intermediate.

MATERIALS AND METHODS

Crystallization. HIV-1 protease and SIV protease were expressed and purified as we described previously (Rosé et

[†] This work supported by NIH Grant GM 39552 to R.M.S.. The graduate studies of R.B.R. were supported by NIH Training Grant CA-09215 in the Graduate Group in Biophysics and the Biotechnology Resources and Education Program Training Grant.

[‡] The coordinates have been deposited with the Brookhaven National Laboratories Protein Data Bank with the assigned identification codes: HIV-1 protease complexed with Ac-S-L-N-F is designated 1YTH and R1YTHSF for structure factors, HIV-1 protease complexed with P-I-V-NH_2 is designated 1YTG and R1YTGSF for structure factors, SIV protease complexed with F-L-E-K is designated 1YTI and R1YTISF, and SIV protease complexed with F(NO_2)-E-A-Nle-S is designated 1YTJ and R1YTJSF.

* Corresponding author. E-mail: stroud@msg.ucsf.edu.

[§] Department of Pharmaceutical Chemistry.

[⊗] Abstract published in *Advance ACS Abstracts,* September 1, 1996.

[1] Abbreviations: Ac, acetyl group; AIDS, acquired immune deficiency syndrome; DMSO, dimethyl sulfoxide; DTT, dithiothreitol; F_c, calculated structure factor; F_o, observed structure factor; HIV, human immunodeficiency virus; Nle, norleucine; rms, root mean square; SIV, simian immunodeficiency virus; TFA, trifluoroacetic acid; amino acids of proteins are designated by the three-letter code; amino acids of peptide ligands are designated by the one-letter code. An arrow, as in Ac-S-L-N-F↓, denotes the end of the product oriented toward the catalytic aspartic acids in the crystal structure. A "/" as in /F-L-E-K denotes the expected cleavage site for the protease. Peptides oriented with their C-termini toward the aspartic acids are referred to as "P" products; peptides oriented with their N-termini toward the aspartic acids are referred to as "Q" products. Amino acid residues in the P product are identified as P1 to P4 outward from the catalytic site position, starting from the C-terminus (Schecter & Berger, 1967). Amino acids in the Q product are identified as P1′ to P4′ outward from the catalytic center, starting from the N-terminus.

S0006-2960(96)01273-1 CCC: $12.00

al., 1993). The primary autolysis site in each protein was altered by mutagenesis of a single site by mutation to reduce autolysis: Gln 7 to Lys in HIV-1 protease, Ser 4 to His in SIV protease. These mutations stabilized them several fold toward autolysis but did not alter the enzymatic activity of the proteases. Protein was dialyzed and concentrated for crystallization in an 8-mL collodion membrane (Rose et al., 1993). The dialysis buffer for both HIV-1 and SIV protease crystallizations contained 20 mM sodium acetate, pH 5.4, and 50 mM sodium chloride, with 1 mM DTT added for HIV-1 protease crystallizations. Protein was concentrated to about 1 mg/mL in the collodion membrane, as measured by the Bio-Rad Bradford assay. For SIV protease, immunoglobulin G was used as a standard; for HIV-1 protease, bovine serum albumin was used as a standard. Peptides were added, and the protein was further concentrated by evaporation with nitrogen until the volume was halved and quartered.

Crystals were grown by hanging drop vapor diffusion at room temperature. Conditions for SIV protease were 4−6% saturated sodium chloride and 100 mM sodium cacodylate, pH 6.5 (Rose et al., 1993). Conditions for HIV-1 protease were 42−50% saturated ammonium sulfate, 200 mM sodium acetate, pH 5.4, and 1 mM DTT (McKeever et al., 1989). High salt was chosen for HIV-1 protease crystallization to increase peptide binding (Hyland et al., 1991b; Richards et al., 1990).

Peptides for crystallization were synthesized (AnaSpec, Incorporated, San Jose, CA), purified by reverse-phase chromatography on a Vydac C-18 column using a water, 0.1% TFA−acetonitrile, 0.1% TFA gradient, and lyophilized. Peptides were dissolved in 5% (v/v) DMSO and added to protein to a final concentration of 30−50 mM. When uncleaved substrate was cocrystallized with protease, the protein was allowed to incubate in the presence of peptide for 2 h prior to evaporation.

SIV protease was crystallized in the presence of the substrate R-V-Nle/F(NO_2)-E-A-Nle-S (the slash denotes the scissile bond) or with each of the cleavage products of the cleavage reaction separately, R-V-Nle/, /F(NO_2)-E-A-Nle-S, or with the product /F-L-E-K. HIV-1 protease was crystallized in the presence of the substrate Ac-S-L-N-F/P-I-V-NH_2, or its products, with the N-terminus acetylated and the C-terminus amidated to neutralize the charge. HIV-1 protease was also crystallized in the presence of the product P-I-V and with both peptides Ac-S-L-N-F/ and /P-I-V.

Structure Determination. Diffraction intensities for the HIV-1 protease, crystallized in the presence of the peptide /P-I-V, the substrate Ac-S-L-N-F/P-I-V-NH_2, or with both peptides Ac-S-L-N-F/ and /P-I-V, and for SIV protease crystallized in the presence of the peptide /F(NO_2)-E-A-Nle-S, were collected using nickel filtered Cu Kα radiation on a 30 cm diameter image plate detector (MAR Research, Hamburg), mounted on a rotating target X-ray source (Rigaku RU-300; 3 kW). Data for HIV-1 protease crystallized in the presence of the peptide Ac-S-L-N-F/ were collected using a wavelength of 1.08 Å on a 30 cm diameter image plate detector (MAR Research, Hamburg) at the Stanford Synchrotron Radiation Laboratory. X-ray data for SIV protease crystallized in the presence of the peptides /F-L-E-K, /R-V-Nle, and R-V-Nle/F(NO_2)-E-A-Nle-S were collected using graphite monochromatized Cu Kα radiation on an image plate detector (R-axis II, Rigaku, Japan), mounted on a rotating target source (Rigaku RU-300; 9 kW). All intensities were collected at room temperature. Intensities were integrated using the program DENZO (Otwinowski, 1993).

The HIV-1 protease crystal form was close to isomorphous with our previously determined structures in space group $P2_12_12_1$, having a protease dimer in the asymmetric unit (Table 2A) (Rutenber et al., 1993). The SIV protease crystals were close to being isomorphous with our previously determined structures crystallized in space group $C222_1$ (Table 2B) (Rose et al., 1993). Intensities were reduced in a lower symmetry space group $C2$, $\beta = 90°$, with a dimer in the asymmetric unit, to allow for any preferential ordering of the asymmetric ligand-bound species. When the density in the active site around the dimer two-fold axis was symmetrical, the structure was further refined in space group $C222_1$ with a monomer in the asymmetric unit.

Protease cores were subject to least-squares rigid body positional refinement within the unit cell, omitting the flaps, waters, and ligand from the protease, and allowing the two monomer positions to refine independently. Positional parameters and thermal factors for atoms of the protease were refined prior to calculation of difference density maps. Structures for the ligands and their interactions were determined from $F_o - F_c$ difference electron density maps (Chambers & Stroud, 1977). Occupancy was determined for bound ligands and water molecules by integration of difference electron density. Integrated density was standardized against the density in a difference map calculated for the catalytic aspartic acid side chains, after omitting them from previous refinement cycles, and from the integrated density for well ordered water molecules seen in $F_o - F_c$ difference electron density maps (Chambers & Stroud, 1977). Water molecules were used to place-hold regions of significant density during structure determination. Omit maps were used to improve reliability and test alternate interpretations. Density maps were interpreted using molecular graphics with CHAIN (Sack, 1988). Crystallographic refinement was carried out by difference density refinement (Chambers & Stroud, 1977) and least-squares positional and temperature factor refinement with restrained molecular mechanics using X-PLOR (Brünger et al., 1987). The free R factor was used to validate additions to the interpretation and only accepted if R_{free} was reduced by their inclusion (Brünger, 1992); any atoms which increased the R_{free} were removed. Crystallographic statistics are summarized in Table 1.

Comparison of Structures. C_α atoms of protease monomers, or dimers were superimposed by least-squares minimization using GEM (available through http://util.ucsf.edu/). To compare the configurations of the bound products with those of peptidomimetic inhibitors, the product structures were superimposed on three previously reported protease−inhibitor structures, HIV-1 protease−U-85548E (8hvp) (Jaskólski et al., 1991), HIV-1 protease−JG365 (7hvp) (Swain et al., 1990), and SIV protease−SKF107457 (1siv) (Zhao et al., 1993) The peptide-based inhibitor sequences are as follows: U-85548E, VSQNL-Ψ[CH(OH)CH_2]-VIV; JG365, Ac-SLNF-Ψ[CH(OH)CH_2N]-PIV-OMe; SKF107457, AAF-Ψ[(S)-CH(OH)CH_2]-GVV-OMe. Ψ denotes replacement of the peptide -CONH- by the group in square brackets, which thereby modifies the scissile peptide bond. These hydroxyethylene and hydroxyethylamine isosteres are all tetrahedral at the P1 carbon in place of the peptide carbonyl. JG365 has the same sequence as one set of substrate and products used in the present report: Ac-S-L-N-F/P-I-V.

Table 1

	(A) Statistics for HIV-1 Protease Structures			
peptide in structure[a]	Ac-SLNF↓	PIV-NH_2↓	ambiguous	Ac-SLNF↓
peptide added during crystallization[b]	Ac-SLNF/	Ac-SLNF/PIV-NH_2	/PIV	Ac-SLNF/ + /PIV
resolution (Å)	2.2	2.3	2.5	2.3
no. of reflections [$I \geq 1\sigma(I)$]	29 000	42 000	15 000	14 000
no. unique reflections	9700	8700	6400	7800
completeness (%)	96	97	93	89
completeness in highest resolution bin (%)	93	97	97	90
R_{sym} (%)	8	7	9	8
average *B* factor (Å^2)	21	22	14	17
R factor (from 7 Å)	19.8	17.9	17.5	17.0
R_{free} (from 7 Å)	28.3	28.5	28.0	26.2
disordered side chains[c]	7, 41, 43, 55, 70, 7′, 14′, 18′, 21′, 41′, 43′, 55′, 65′, 70′	7, 8, 14, 41, 55, 70, 7′, 41′, 43′, 55′	7, 18, 41, 43, 55, 7′, 14′, 41′, 70′, 80′	7, 41, 43, 55, 70, 7′, 14′, 41′, 43′, 55′
no. waters	31	42	19	36
rmsd from ideal				
bond length (Å)	0.014	0.015	0.018	0.014
bond angles (deg)	2.9	3.1	3.1	3.0
dihedral angles (deg)	27.0	27.3	27.0	27.0
	(B) Statistics for SIV Protease Structures			
peptide in structure[a]	↓FLEK	↓F(NO_2)EAA	AAA↓	AAA↓
peptide added during crystallization[b]	/FLEK	/F(NO_2)EANleS	RVNle/	RVNle/F(NO_2)EANleS
resolution (Å)	2.2	2.5	2.6	2.2
no. of reflections [$I \geq 1\sigma(I)$]	21 000	13 000	7000	15 000
no. unique reflections	4900	3500	4300	7750
completeness (%)	93	98	85	80.5
completeness in highest resolution bin (%)	83	100	86	72
R_{sym} (%)	5.3	7.9	5.1	7.4
average *B* factor (Å^2)	29	14	20	21
R factor (from 7 Å)	21.1	17.4	18.2	17.0
R_{free} (from 7 Å)	29.5	28.5	28.0	26.0
disordered side chains[c]	7, 37, 70, 72	37	7, 70, 72, 79, 7′, 61′, 70′, 72′	7, 37, 70, 72, 7′, 13′, 37′, 70′, 72′
no. waters	19	15	24	42
rmsd from ideal				
bond length (Å)	0.016	0.014	0.008	0.016
bond angles (deg)	3.1	3.1	1.55	3.1
dihedral angles (deg)	27.0	26.7	26.6	27.5

[a] An arrow denotes peptide residue closest to the catalytic aspartic acids in the structure. An arrow to the right of the sequence denotes a P product, an arrow to the left denotes a Q product. [b] A slash denotes location of the scissile bond in the substrate or, for the products, the substrate from which the product was derived. [c] These residues were built as alanines in the structure. For structures with a dimer in the asymmetric unit, unprimed residues belong to the monomer more closely associated with the P product; primed residues belong to the monomer more closely associated with the Q product.

Superimposed on the same species of protease, the differences in structures have an rms deviation of 0.5 Å or less. Thus the backbones remain very similar between all the HIV-1 structures, or between all the SIV structures. When SIV structures were superimposed on HIV-1 protease structures the differences have an rms deviation of about 1.2 Å. This larger value may in part reflect the different molecular arrangements in the two crystal space groups wherein the SIV protease dimer lies on a crystallographic two-fold axis, where the HIV-1 protease dimer does not. To determine and compare a bound product with a peptidomimetic inhibitor conformation, ligand-bound structures were superimposed by optimizing overlap of the C_α atoms of the protease monomer most closely associated with the product. This facilitates the optimal comparison of the ligands on the product side, minimizing the impact of any structural differences that may result on the other side since one of the structures has no bound ligand there.

Generating the Substrate Model. The substrate was modeled from product structures by superimposing the protease dimers from structures bound to N- and C-terminal products and reconnecting the scissile bond. The HIV-1 protease dimer bound to the N-terminal product Ac-S-L-N-F/ was superimposed on that of SIV protease bound to the C-terminal product /F-L-E-K. One of the C-terminal carboxyl oxygens of Ac-S-L-N-F/ was removed, and the terminal carbon was reconnected to the terminal nitrogen of /F-L-E-K. The geometry of the peptide was regularized, and the degree of variation within the energy minimum conformation was assessed.

RESULTS

Structure of HIV-1 Protease with N-Terminal Product Ac-S-L-N-F↓. The N-terminal product Ac-S-L-N-F↓ was cocrystallized with HIV-1 protease. Figure 1a shows the final $2F_o - F_c$ density in the binding pocket for HIV-1 protease bound to the product Ac-S-L-N-F↓, which unambiguously defines the conformation of the P product. Electron density for the catalytic aspartic acids, the peptide, and the conserved water 401 is displayed. Only the N-terminal acetyl group of the peptide is without density. The density in the unliganded half of the binding pocket contains three ordered

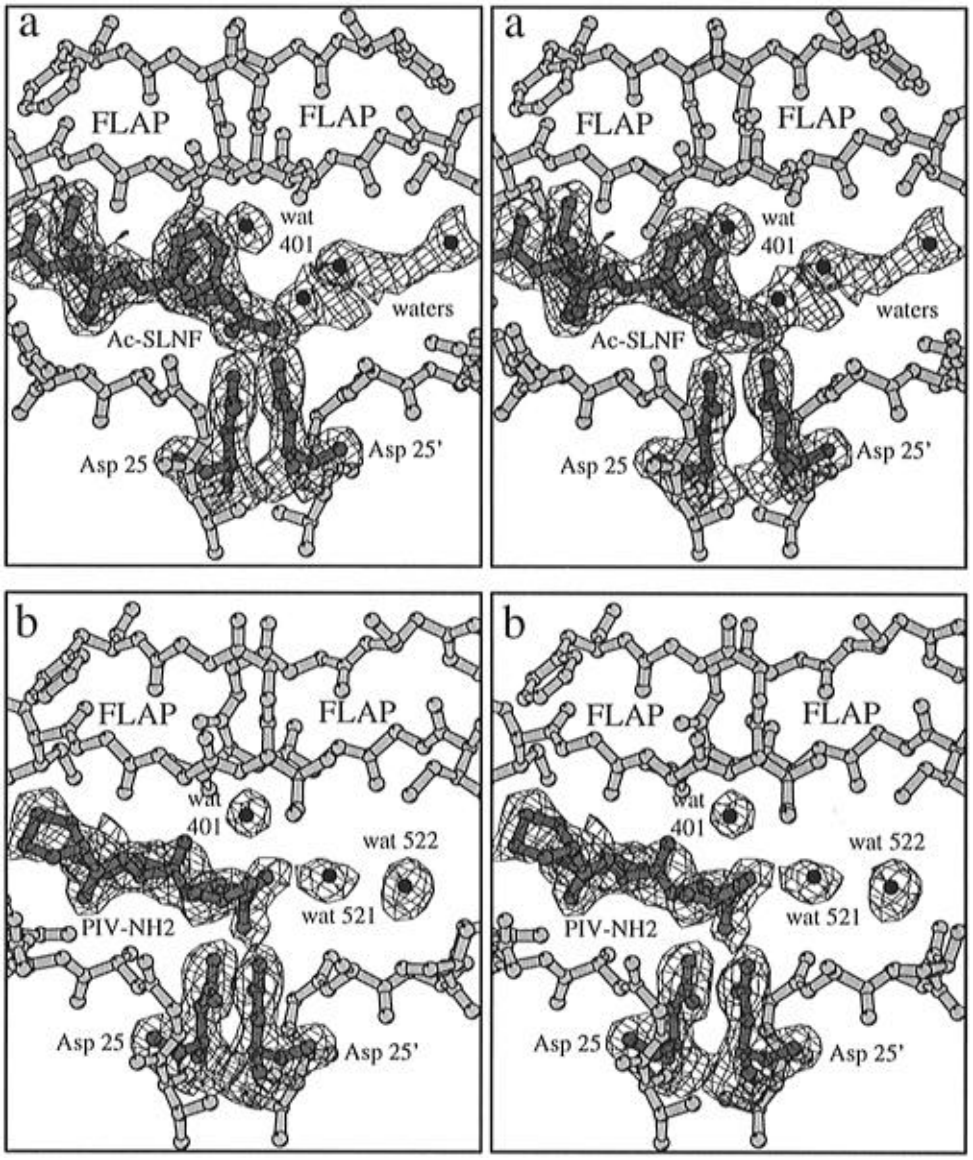

FIGURE 1: Final $2F_o - F_c$ density (contoured at 1σ) in the binding pocket of HIV-1 protease bound to the peptide products (a) Ac-S-L-N-F↓ and (b) P-I-V-NH_2↓. Density is also shown for the active site aspartic acids and water 401. Density for residues 46−53 and 46′−53′ (the flaps) and residues 26−30 and 26′−30′ (the base of the active site) is not displayed. (a) The peptide and three water molecules have been built into the binding pocket. These waters do not explain all of the density. (b) The peptide is seen to bind "backwards" from the sequence in the substrate Ac-S-L-N-F↓P-I-V-NH_2, with the valine in the P1 position. The unliganded half of the binding pocket contains density for three water molecules: waters 521 and 522 (displayed) and water 520 not displayed because it lies directly in front of the C-terminus (cross-eyed stereo).

water molecules. They are not hydrogen-bonded to the protein or other ordered water molecules and do not explain all of the difference density. They have thermal *B* factors between 43 and 53 Å^2, compared with the average *B* factor for the peptide C_α's of 20 Å^2.

The C-terminal carboxyl group oxygens are each within hydrogen-bonding distance of one of the catalytic aspartic acids, forming an unusual three-carboxyl triad interaction (Figure 2). Two tautomeric forms are represented in Figure 3a, though the geometry clearly favors the first configuration. Two oxygens, one from each of the catalytic carboxyls, share a hydrogen bond as well as being hydrogen-bond acceptors from the amide nitrogens of Gly 27 and Gly 27′.

All hydrogen bonds between enzyme and product are made with one monomer of the protease dimer (identified as the unprimed monomer for P products) except for the hydrogen bonds involving the catalytic aspartic acids. The backbone hydrogen bonds are the same as those formed by peptidomimetic inhibitors (Swain et al., 1990). Two hydrogen bonds are formed from the nitrogen and to the oxygen of Gly 48 in the flap, and two hydrogen bonds are formed to the oxygen of Gly 27 and from the nitrogen of Asp 29 at the base of the binding pocket. A fifth hydrogen bond is formed between the side chain oxygen of Asp 29 and the backbone oxygen of the P3 Leu of the peptide. Water 401 is hydrogen-bonded to the carbonyl of the peptide at P2 Asn. This conserved water is hydrogen-bonded from the amide nitrogens of Ile 50 and Ile 50′, located on the flaps of both protease monomers, and also seen in all peptidomimetic inhibitor structures (Wlodawer & Erickson, 1993). Two hydrogen bonds are also formed between side chains of Ac-S-L-N-F↓ and the protease. The P2 Asn is hydrogen-bonded with Asp 29 and 30 of the protease, and the P4 Ser is hydrogen-bonded with Asp 30 of the protease.

Table 3 lists the protease residues forming the binding pockets of the product complex structures (defined as being

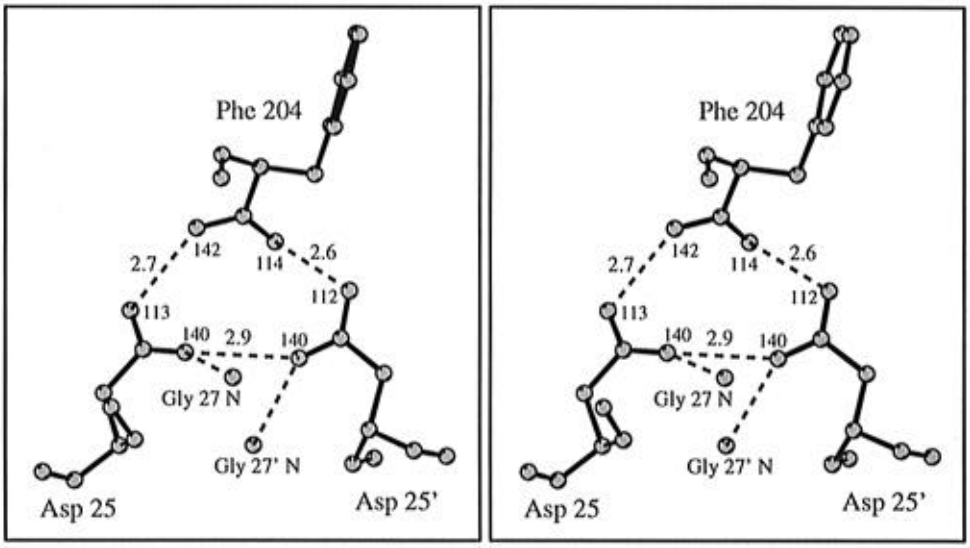

FIGURE 2: Diagram of the "triangular interaction" among the three carboxyl groups: the active site aspartic acid residues and the C-terminus of the phenylalanine residue of the peptide product Ac-S-L-N-F↓. The distances between carboxyl oxygens and the angles between carbon−oxygen−oxygen atoms are displayed (cross-eyed stereo).

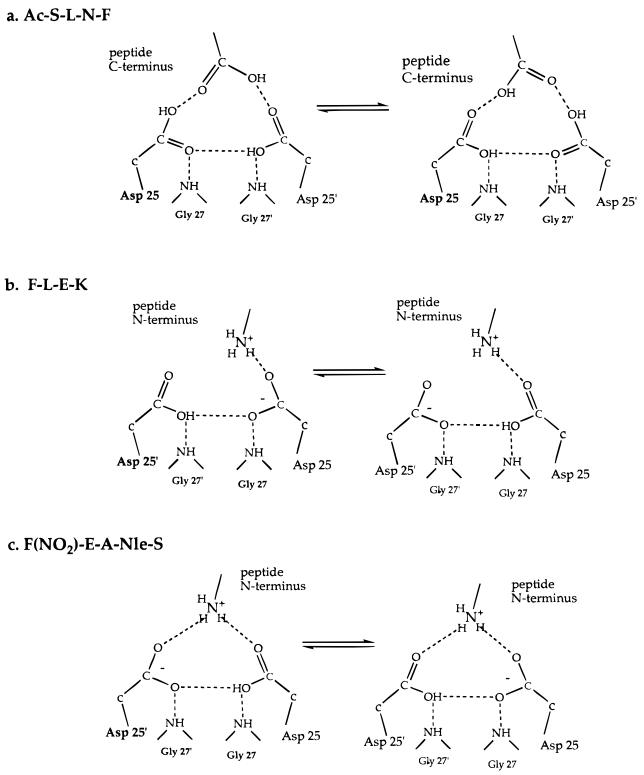

FIGURE 3: Tautomers of the N- and C-termini of products interacting with the catalytic aspartic acids. (a) Two tautomeric forms of the carboxy terminus of the P product Ac-S-L-N-F↓. The left-hand tautomer is geometrically favored. (b) Two tautomers of the amino terminus of the Q product ↓F-L-E-K. The amino terminus forms a hydrogen bond with one of the aspartic acids. At pH 5.4 the amino terminus would be protonated. (c) Two tautomers of the amino terminus of the Q product ↓F(NO_2)-E-A-Nle-S. The amino terminus forms hydrogen bonds with both aspartic acids.

within 4.2 Å of the residue). The side chains of the product peptides protrude on alternate sides of the peptide backbone in a β-sheet conformation. Most of the interactions are with residues of the unprimed monomer except at the P1 subsite.

Figure 4a shows the superposition of the Ac-S-L-N-F↓ product peptide with JG365 and U-85548E. The backbone atoms of Ac-S-L-N-F↓ and JG365 for residues P1−P3 overlap with an rms deviation of 0.4 Å, and diverge at the P1 carboxyl carbon of Ac-S-L-N-F↓, which is 0.8 Å from the corresponding carbon of JG365. The conformation of the N-terminal acetyl group and the P4 Ser differ slightly. The position of the hydroxyl of the hydroxyethylene isostere is oriented in a direction between the carboxyl oxygens of product terminus in Ac-S-L-N-F↓. The backbone atoms of Ac-S-L-N-F↓ and U-85548E for residues P1 to P3 overlap closely with an rms deviation of 0.4 Å. The C-terminal carbon of Ac-S-L-N-F↓ is 0.3 Å from the corresponding carbon of U-85548E. According to the formula for estimation of expected accuracy in atomic positions as a function of *B* factor (Stroud & Fauman, 1995), the atomic coordinates for the backbone atoms of Ac-S-L-N-F↓ at residues P1−P3 are accurate to within 0.2 Å. The atomic coordinates for the backbone atoms of JG365 at residues P1−P3 are accurate to within 0.3 Å. Therefore, changes of the order of 0.4 Å between Ac-S-L-N-F↓ and JG365 are significant.

Monomers of the protease dimer are no longer identical once the peptide product binds. C_α's from one monomer optimally superimpose on those of the other monomer with a rotation of 178° and an rms deviation of 0.53 Å. Differences between corresponding atoms are as great as 1.8 Å, at residue 49 of the flaps. The liganded monomer is slightly more compact than the unbound monomer, with the mainchain atoms shifted inward toward the bound peptide. This is particularly true of residues 27−30 (the β-strand forming the base of the binding pocket), residues 33−56 (the flap residues), residues 78−83 (the loop which partially forms the P1 binding pocket), and residues 86−95 (the helix). The N-terminal β-strand of the primed monomer, residues 1′−8′, are shifted in toward the peptide; Arg 8′ forms part of the P3 binding pocket. Several side chains differ in conformation between monomers, mostly on the surface of the protein. Within the binding channel, the side chain of Ile 84′ is rotated to interact with the peptide P1 Phe. And the side chain of Asp 30 is hydrogen-bonded to the peptide of Ser P4 in the liganded monomer and is turned toward solvent in the unliganded monomer.

From a comparison of the temperature factors between the C_α atoms of the monomers, the liganded monomer (residues 1−99) is more rigidly constrained than the unliganded monomer (residues 1′−99′). The average *B* factors

Table 2

(A) Dimensions of HIV-1 Protease Unit Cells

peptide	Ac-SLNF/	Ac-SLNF/PIV-NH_2	/PIV	Ac-SLNF/ +/PIV
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	51.1	51.3	51.2	51.5
b (Å)	59.3	59.1	58.9	59.3
c (Å)	61.4	62.3	58.9	62.4

(B) Dimensions of SIV Protease Unit Cells

peptide	/FLEK	/F(NO_2)EANleS	RVNle/	RVNle/ + F(NO_2)EANleS
space group	$C222_1$	$C222_1$	$C2$	$C2$
a (Å)	63.7	62.3	62.8	62.6
b (Å)	32.5	32.1	32.0	32.0
c (Å)	97.7	96.3	96.4	96.3
β			90°	90°

Table 3: Protease Side Chains Forming the Peptide Binding Subsites

subsite	protease	peptide[a]	protease contacts[b]
S4	HIV-1	Ac-**S**-L-N-F↓	D29, D30, K45, I47
S3	HIV-1	Ac-S-**L**-N-F↓	**G48, D29**, R8′
S2	HIV-1	Ac-S-L-**N**-F↓	**G27, A28, D29, D30, I47, G48, I84**, I50′
S1	HIV-1	Ac-S-L-N-**F**↓	**G49, I50**, L23′, P81′, V82′, I84′
S1′	SIV	↓**F**-L-E-K	**G27′**, L23, I82, D25, I84
S2′	SIV	↓F-**L**-E-K	**A28′, D30′, I32′, I84′, G48′**, I50
S3′	SIV	↓F-L-**E**-K	**D29′**, R8
S4′	SIV	↓F-L-E-**K**	**I46′, D30′, E58′, K45′, M76′**

[a] In bold and underlined are the residues of the peptide occupying the respective subsites. [b] Unprimed residues belong to the monomer more closely associated with the P product; primed residues belong to the monomer more closely associated with the Q product. Residues belonging to the more closely associated monomer are in bold for emphasis.

for the C_α atoms of the liganded and unliganded monomers are 18.5 and 22.8 Å^2, respectively. The average *B* factors of both flaps are 17 Å^2, indicating that both flaps are stabilized in this structure.

Structures of Other N-Terminal Product Complexes. Crystallization of HIV-1 protease in the presence of both products Ac-S-L-N-F/ and /P-I-V resulted in a structure similar to that of HIV-1 protease cocrystallized with Ac-S-L-N-F/ alone, without P-I-V bound. The rms deviation between the peptides from these two structures was 0.23 Å, essentially identical to one another.

Crystallization of HIV-1 protease in the presence of the (uncleaved) substrate, Ac-S-L-N-F/P-I-V-NH_2, resulted in a structure of the amidated product P-I-V-NH_2↓ bound in the catalytic site as a P product. The peptide bound with the Val in the P1 position, instead of the Pro in P1′ as would be expected from the substrate cleavage site, due to the interaction of the C-terminal amide with the catalytic carboxyls. Figure 1b shows the final $2F_o - F_c$ density in the binding channel. The C-terminal amide was pointed perpendicular to the orientation of the C-terminal carboxyl group of Ac-S-L-N-F↓, similar to the orientation of the hydroxyl group of the hydroxyethylene isosteres U-85548E and SKF107457. It was not possible to distinguish the amino group from the oxygen from the density. The amino group lies toward the active site aspartic acids, 3.1 and 3.5 Å away. The oxygen lies within hydrogen-bonding distance of waters 401 and 521. With only three residues, the peptide P-I-V-NH_2↓ makes three of the four backbone interactions observed in the Ac-S-L-N-F↓ peptide, omitting the fourth possible hydrogen bond to the Gly 48 oxygen of the flap. This hydrogen bond is formed by the nitrogen of proline P3.

The peptide /P-I-V was crystallized with HIV-1 protease to determine if the unamidated peptide would bind as a Q product. There was good difference density for the isoleucine in one half of the binding pocket, but inclusion of the proline or valine on either side of the isoleucine increased the R_{free}. We conclude that the peptide bound in multiple conformations.

The structure of SIV protease with the product R-V-Nle↓ showed asymmetrical difference electron density in omit maps of the active site region. The structure was therefore refined in the lower symmetry space group *C*2 with a dimer in the asymmetric unit, allowing for different contents within the two halves of the active site. The P-product A-A-A↓ fitted the density better than the Q-product ↓A-A-A because of the density defining the carboxyl group near the catalytic aspartic acids. Unlike in the Ac-S-L-N-F↓ structure, the C-terminal carboxyl group in this structure was oriented toward one of the aspartic acids.

The structure of SIV protease crystallized in the presence of the complete (uncleaved) substrate R-V-Nle/F(NO_2)-E-A-Nle-S resembled that of SIV protease crystallized with R-V-Nle/ alone, indicating the substrate had been cleaved and only the P product was bound. The electron density was best described by the peptide sequence A-A-A↓.

Structure of the C-Terminal Products ↓F-L-E-K and ↓F-(NO_2)-E-A-Nle-S. The structure of SIV protease crystallized with the product peptides ↓F-L-E-K, or ↓F(NO_2)-E-A-Nle-S, was refined in space group $C222_1$ with one monomer (half of the protease) in the asymmetric unit. Electron density corresponding to bound peptide is therefore seen on both sides of the binding channel. The interpretation of the peptide conformations, ~50% on each side, is unambiguous. The stoichiometry of binding is one peptide per protease dimer, as determined by integration of the standardized difference electron density (Chambers & Stroud, 1977). Therefore, each protease dimer binds only one product peptide and in a statistical sense such that neither side is favored. Furthermore, the N-termini of the two half-occupied, symmetry-related images of ↓F(NO_2)-E-A-Nle-S peptides were only 1.5 Å apart, excluding the possibility that some molecules had two peptides bound, while other dimers might have had only one. In the ↓F-L-E-K peptide-bound structure the occupancy of the peptide refined to 60% ± 10%, essentially 50% occupancy corresponding to a stoichiometry of one peptide per dimer. The symmetry related positions for the N-termini of the Q product were 2.9 Å apart. These product complex structures therefore contain one peptide product per dimer, randomly distributed in the crystal with

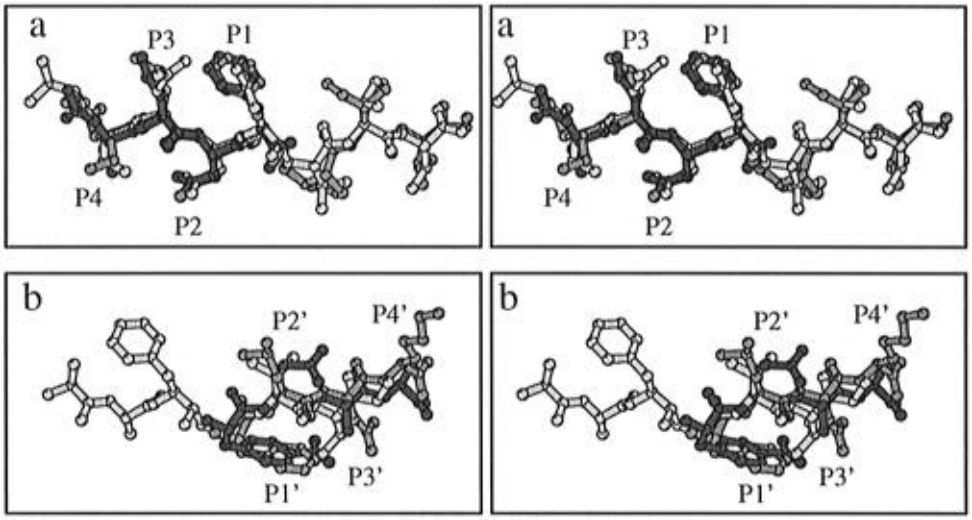

FIGURE 4: Conformations of the bound product peptides are compared with those of peptidomimetic inhibitors. The peptides were overlapped by superimposing the C_α atoms of the proteases. (a) The P product Ac-S-L-N-F↓ (black) is overlapped with the inhibitor JG365 (gray) of the same N-terminal sequence. The peptides differ the most at the P1 carbonyl of Ac-S-L-N-F↓. The third peptide, U-85548E (white), adopts the same backbone conformation. (b) The Q products ↓F(NO_2)-E-A-Nle-S (black) and ↓F-L-E-K (gray) overlap closely at the P2′ C_α, but less well at P1′ and P3′. The same is true of the peptidomimetic inhibitor SKF107457 (white). The P2′ residue is usually a glutamate or glutamine in substrates (cross-eyed stereo).

no preference for binding either of the two monomers. Any differences between monomers that resulted from ligand binding are therefore averaged in the electron density.

Figure 5a shows the final $2F_o - F_c$ electron density in the active site of the ↓F-L-E-K structure. Density for the catalytic aspartic acids, the peptide, and the conserved water 401 is displayed. The peptide is in good density except for the terminal three atoms of the Leu side chain in the P2′ binding site.

Figure 5b shows the final $2F_o - F_c$ density in the active site of the ↓F(NO_2)-E-A-Nle-S structure. The sequence ↓F-(NO_2)-E-A-A was built into the active site since there was no density for the P4′ norleucine (Nle) side chain or for the P5′ Ser. The phenyl ring of the *para*-nitrophenylalanine residue was not completely in density, suggesting disorder. The refined thermal *B* factors for the ring were nevertheless low, ranging from 3 to 6 Å^2. Unlike the structure with ↓F-L-E-K, the $2F_o - F_c$ map (contoured at 0.8σ) showed continuous density between the aspartic acids. A water molecule built between the aspartic acids clashes sterically with the N-terminus of the peptide.

All hydrogen bonds are made with one monomer of the protease dimer (identified as the primed monomer for the Q products), except for the hydrogen bonds to the catalytic aspartic acids. The same four backbone hydrogen bonds formed by the P product are formed to Gly 48′ of the flaps and Gly 27′ and Asp 29′ at the base of the binding pocket. The P1′ carbonyl forms a hydrogen bond to the conserved water 401. The N-terminus of ↓F-L-E-K is hydrogen-bonded with the catalytic aspartic acid of the unprimed monomer, 3.3 Å away. The N-terminus of ↓F(NO_2)-E-A-Nle-S is hydrogen-bonded with both active site aspartic acids, both 3.1 Å away. Two tautomeric forms for each configuration are represented in Figure 3b,c.

One side chain from each of the products ↓F-L-E-K and ↓F(NO_2)-E-A-Nle-S forms hydrogen bonds with the protease: the P4′ Lys of ↓F-L-E-K hydrogen-bonds with the side chain of Asp 58, and the P2′ Glu of ↓F(NO_2)-E-A-Nle-S hydrogen-bonds with the backbone nitrogen of Asp 29 and Asp 30. In addition, Arg 8 forms a salt bridge with the P3′ Glu of ↓F-L-E-K. The P1′ nitro group of ↓F(NO_2)-E-A-Nle-S interacts with Arg 8 electrostatically.

Table 3 lists the protease residues forming the binding pockets for ↓F-L-E-K. Most of the interactions are with residues of the primed monomer, the same monomer forming the hydrogen bonds to the product, except for the S1′ subsite. The S3' subsite, on the same side of the peptide as the S1' subsite, is formed partly by Arg 8 of the unprimed monomer.

Figure 4b compares the conformation of ↓F-L-E-K with ↓F(NO_2)-E-A-Nle-S. The rms deviation for the C_α atoms is 0.4 Å. The C_α atoms at the P2′ position are only 0.2 Å apart. The peptides overlap less well at the other residues; the C_α's of P3′ are 0.4 Å apart, and the C_α's at P1′ are 1.3 Å apart. A similar trend holds for the overlap with the peptidomimetic inhibitor SKF107457. The N-terminal nitrogen atoms of the peptide products are 1.6 Å apart. Water 401 is shifted toward the aspartic acids in the ↓F(NO_2)-E-A-Nle-S structure by about 0.5 Å, and the flaps are slightly more closed by as much as 0.6 Å. According to an estimation of the errors in the atomic coordinates as a function of *B* factor (Stroud & Fauman, 1995), the atomic coordinates for the backbone atoms of ↓F-L-E-K, and ↓F-(NO_2)-E-A-Nle-S at residues P1′−P3′ are accurate to within 0.2 Å. Adding these errors in quadrature indicates that changes of the order of 0.3 Å between ↓F-L-E-K, and ↓F-(NO_2)-E-A-Nle-S are significant.

Model of Substrate. The substrate was modeled by superimposing the HIV-1 protease structure bound to Ac-S-L-N-F↓ onto the SIV protease structure bound to ↓F-L-E-K. The distance between the scissile carbon of the P product Ac-S-L-N-F↓ and nitrogen of the Q-products ↓F-L-E-K or ↓F(NO_2)-E-A-Nle-S, was 2.2 or 1.3Å, respectively, to be compared with the idealized peptide bond in which it is 1.45 Å and planar. The ϕ and ψ angles for the resulting peptide were within the allowed region on a Ramachandran plot indicating that the geometry was favorable and likely to represent the substrate conformation prior to cleavage. Figure 6 shows the carbonyl after regularization, and to eliminate the possibility of bias in favor of one aspartate, following removal of each of the carboxyl oxygens in turn.

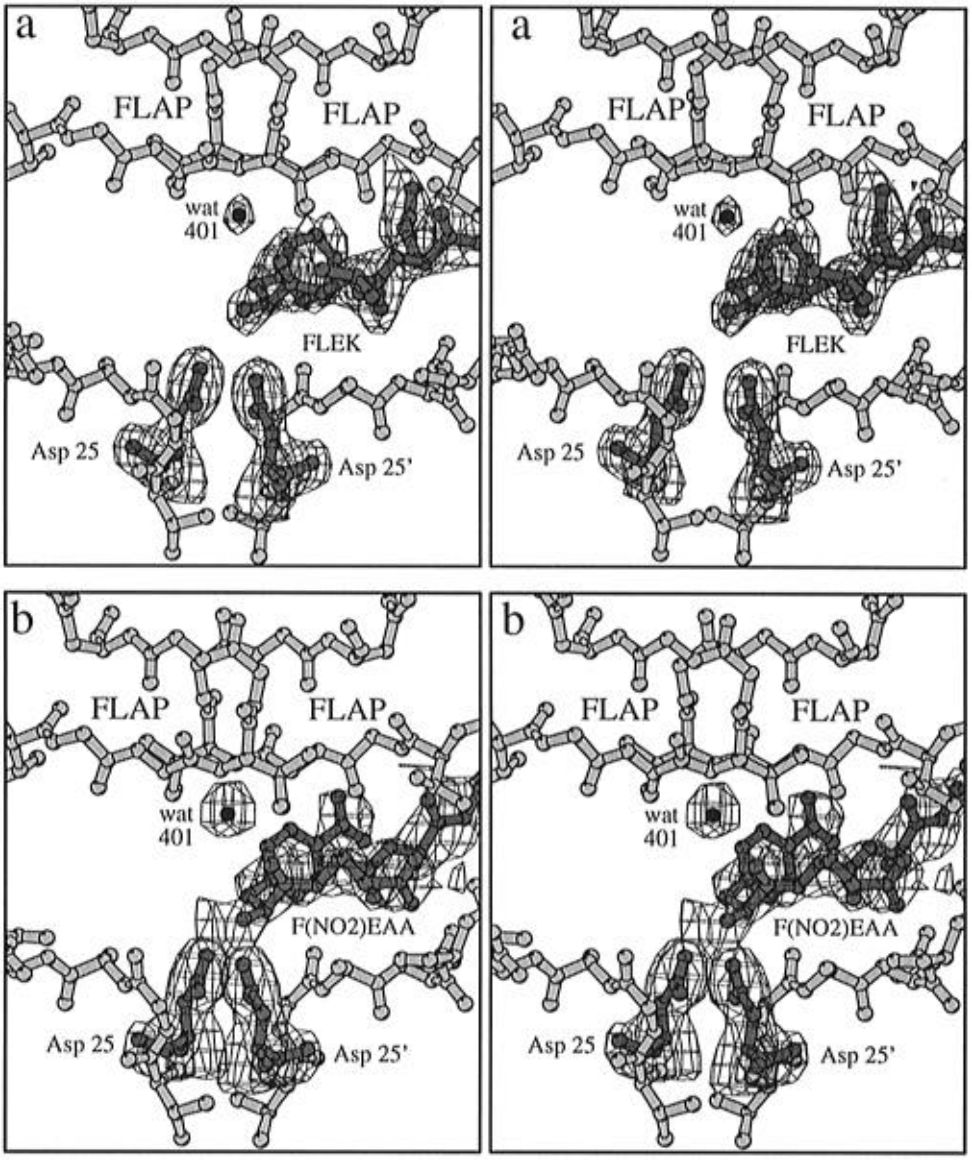

FIGURE 5: Final $2F_o - F_c$ density in the binding pocket of SIV protease bound with the peptide products (a) ↓F-L-E-K (contoured at 0.9σ) and (b) ↓F(NO_2)-E-A-Nle-S (contoured at 0.8σ). The monomers of the dimer are related by a crystallographic two-fold axis; density is shown only for one of the peptides in the active site. Density is also shown for the active site aspartic acids and for water 401, between the flaps. Density for residues 46−53 and 46′−53′ (the flaps), and residues 26−30 and 26′−30′ (the base of the active site) is not displayed. (b) Density between the aspartic acids is unaccounted for (cross-eyed stereo).

Removal of OT1, near Asp 25′, led to much less movement of the carbonyl than removal of OT2, near Asp 25. The black line connecting these positions represents conformations of the peptide bond of the same minimal energy. Regularization with the oxygen at any of these positions did not shift the position of the oxygen. Varying the atoms fixed during regularization did not affect the outcome of regularization; fixing P2 to P4 and P2′ to P4′ gave the same results as removing all constraints, indicating that the conjugation of P and Q products involves no strain.

Re-forming the peptide bond was achieved with minimal movements of atoms: 0.7 Å for the carbon and 0.8 Å for the nitrogen of the scissile bond, respectively, 0.3 and 0.5 Å for the C_α atoms of P1 and P1′, respectively, and less than 0.2 Å for the backbone atoms more distant from the scissile bond. The distance between the C_α atoms of the P1 and P1′ residues in the products was 4.6 Å, and in the final substrate was 3.9 Å. The ϕ and ψ angles rotated 15° or less, also demonstrating that strain, or difference between substrate and products, is not involved in substrate cleavage, as assessed by comparison of products and substrate.

DISCUSSION

Product Complexes Are Reaction Intermediates. These are the first structures of noncovalent product complexes of aspartyl proteases. Two peptide substrates and their proteolyzed products were crystallized in the presence of HIV-1 and SIV proteases. The substrate Ac-S-L-N-F/P-I-V-NH_2 is a modification of the p17/p24 cleavage site of the viral polyprotein (Rich et al., 1990). The second substrate, R-V-Nle/F(NO_2)-E-A-Nle-S, is a nonnatural chromogenic substrate used in spectrophotometric assays. An additional product, ↓F-L-E-K, was studied, derived from the reverse transcriptase/integrase cleavage site for SIV protease: R-Q-V-L↓F-L-E-K (Grant et al., 1991). The structures of two amino-terminal products (the P products, Ac-S-L-N-F↓ and P-I-V-NH_2↓) with HIV-1 protease and two carboxy-terminal products (the Q products, ↓F(NO_2)-E-A-Nle-S and ↓F-L-E-

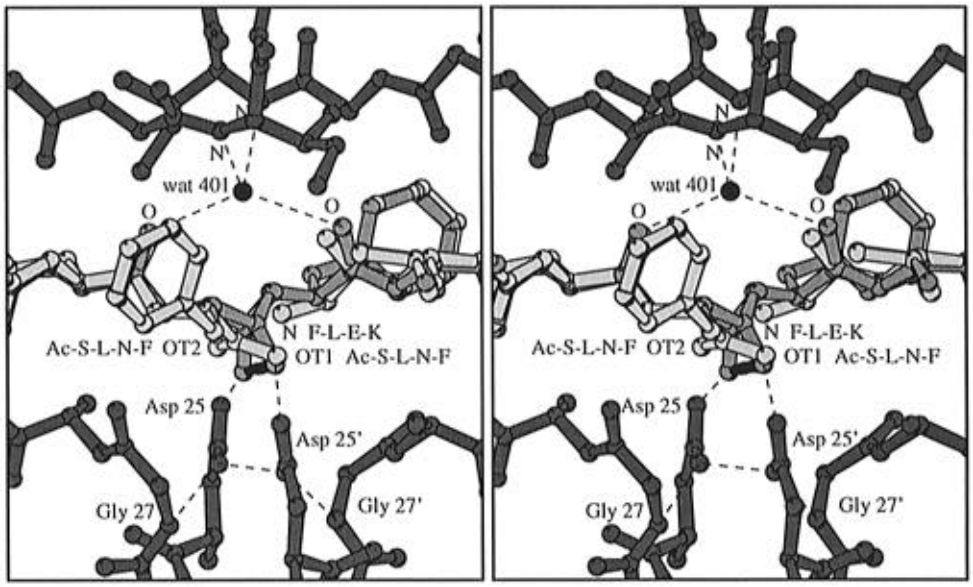

FIGURE 6: Model of the substrate (gray) overlapped with the two products (white): Ac-S-L-N-F↓ and ↓F-L-E-K. The scissile carbonyl of the substrate adopts a range of equal-energy conformations (black line) without strain the two extreme conformations of which are displayed. The position closest to OT1 of the C-terminal carboxyl group of Ac-S-L-N-F↓ is within hydrogen-bonding distance (dotted line) of Asp 25′ (black). Hydrogen bonds are also shown to water 401 (cross-eyed stereo).

K) with SIV protease showed well ordered density for the peptides.

No structures with both products bound simultaneously were obtained, presumably due to the close interactions that would be forced by the observed locations of products on either side of the catalytic carboxyls. Crystallization of protease in the presence of substrate or both products resulted in structures with one product bound to the enzyme. We expected the substrate to be cleaved by the protease during crystallization, but we considered it possible that the resulting structure would contain both products if crystal formation preceded product release.

HIV-1 and SIV proteases are structurally and functionally similar. Their sequences are 50% identical. Their structures differ by an rms deviation of 1.2 Å between the C_α atoms when the dimers are superimposed (Rose et al., 1993). SIV protease is capable of cleaving the HIV-1 polyprotein cleavage sites with varying efficiencies (Debouck, 1991; Grant et al., 1991). The product complex structures with each of these proteins are therefore comparable.

Kinetics show formation of product complexes. Competitive and noncompetitive inhibition patterns for product inhibitors of HIV-1 protease with millimolar inhibition constants have been determined (Hyland et al., 1991b). From the peptide /P-I-V-NH_2, short peptide products may bind the protease in either orientation, complicating the interpretation of such inhibition studies, since the substrate may not be binding in the presumed mode. The two-fold symmetry of the binding channel of the dimer requires that the channel in the monomer accommodates the backbone of a peptide ligand oriented as a P product or as a Q product. Amidation led to the reversed orientation of this peptide with respect to the substrate. It positioned the C-terminus perpendicular to that of the carboxyl group of Ac-S-L-N-F↓, similar to the conformation of the hydroxyethylene hydroxyl group of U85548-E. Since the sequences studied by Hyland et al. (1991a,b) were all acetylated and amidated, the assumption that the products all bind in one orientation is probably incorrect. Despite this caveat, these inhibition studies indicated random order of release of products (Meek et al., 1994). Our structures with P and Q products are consistent with a random release mechanism.

The P product Ac-S-L-N-F↓ adopts a conformation essentially identical to the "P" portion of the peptidomimetic inhibitor JG365, except for the carboxy-terminal carbon. The repositioning of this carbon may reflect both the flexibility required to cleave substrates with a P1′ proline and the extra bond of the hydroxyethylamine isostere JG365. The "P" portion of U-85548E also superimposes closely with Ac-S-L-N-F↓, despite the difference in sequence. The P product Ac-S-Q-N-Y↓ has been shown to bind to HIV-1 protease in an enzymatically competent conformation in solution (Hyland et al., 1991b). The exchange rate of ^{18}O from $H_2{}^{18}O$ into the C-terminal carboxyl group is similar to the exchange of ^{18}O into substrate. We assume that the C-terminal carboxyl group of the bound product Ac-S-L-N-F↓ reflects this conformation.

The carboxy terminus of the P-product, Ac-S-L-N-F↓, forms a "triangular" interaction with the two catalytic aspartic acids as shown in Figure 2. Each carboxyl group is protonated to form the three hydrogen bonds of the triangle. In order to form hydrogen bonds at pH 5.4, the pH of the crystals, the pK_a's of the three carboxyl groups are raised from the solution value of about 3 in the P-product complex. The pK_a of the proton shared between the two catalytic aspartic acids has been measured to be between 5.5 and 6.8 (Hyland et al., 1991b; Ido et al., 1991). Raising the pK_a of the other protons of the triad is consistent with the P product binding to the protease more tightly at low pH, where these bonds can be maintained, than at high pH, where the hydrogen bonds would be broken. The peptide Ac-S-Q-N-Y↓ binds 5 times tighter at pH 3.5 (K_i = 0.2 mm) than at pH 6 (K_i = 1.0 mm) (Hyland et al., 1991b). A deuterium isotope effect which decreased between pH 6 and 3.2 was also observed (Hyland et al., 1991a). This was interpreted as a noncatalytic step becoming rate-limiting at low pH, possibly involving release of the P product.

Unlike the P product, the Q product's mainchain atoms adopt a sequence-dependent conformation when bound to the protease, except at P2′ (Figure 4b). The more rigidly

held P2′ side chain is often a glutamate or glutamine in naturally occurring substrates, and the site is less tolerant of mutations than are other subsites (Poorman et al., 1991; Tomasselli & Heinrikson, 1994; Wlodawer & Erickson, 1993). One oxygen of the glutamate side chain in the P2′ site of ↓F(NO_2)-E-A-Nle-S forms a bifurcated hydrogen bond to the backbone nitrogens of Asp 29 and Asp 30, as in HIV-2 protease (Griffiths et al., 1992; Tong et al., 1993), providing some of the constraints at this side chain. A glutamine side chain could satisfy the same interactions.

The positions of the N-terminal nitrogens of ↓F-L-E-K and ↓F(NO_2)-E-A-Nle-S differ by 1.6 Å. The *para*-nitrophenylalanine residue of ↓F(NO_2)-E-A-Nle-S may cause the amino terminus to be closer to both aspartic acids. The N-terminal nitrogen of ↓F-L-E-K, with the smaller P1′ Phe residue, hydrogen bonds with only one aspartic acid.

Substrate Binds without Strain. The idea that substrate strain can assist in bond cleavage has been discussed in the context of aspartyl proteases for some time (Davies, 1990; Pearl & Blundell, 1984). Fruton proposed that the strain in the substrate upon binding to pepsin could contribute to bond cleavage (Fruton, 1976). Monomeric aspartyl proteases and HIV-1 protease complexed with peptidomimetic inhibitors with tetrahedral geometry at the scissile carbon have been interpreted in terms of strain of the planar peptide bond of a substrate (Jaskólski et al., 1991; Suguna et al., 1987b). Modeling the substrate from the product complex structures indicates the absence of strain in the reconnected peptide bond.

In modeling the substrate from the products, we assume that the products bind to the protease in the most energetically favorable conformation for that sequence. Because the product complex structures contain only one of the products, their positioning is independent of the repulsive forces that would occur in the newly cleaved substrate. A similar analysis has been carried out with dihydrofolate reductase in which a ternary complex was modeled by overlapping two binary complexes (Davies et al., 1990).

The scissile carbonyl of the modeled substrate can occupy a range of positions of equal energy without strain, as shown in Figure 6. These conformations are attained by minimal movements of the P1 and P1′ residues, indicating that a substrate bound to the protease utilizing similar interactions to the products would not deform the scissile peptide bond. The movements in the P1 and P1′ residues are small enough to be accommodated by the protein without much energetic cost. This argues against the development of strain as a result of substrate binding.

A planar peptide bond has been modeled from a hydroxyethylene peptide isostere using a similar regularization procedure (Jaskólski et al., 1991). The carbonyl oxygen of the hydroxyethylene isostere is located between the two carboxyl oxygens of Ac-S-L-N-F↓. Positioning of the scissile carbonyl was sensitive to the atoms held fixed while the geometry was regularized, unlike our model. If the P1 and P1′ atoms were rigidly fixed (except for the P1 carbonyl oxygen) to their conformation in the peptidomimetic inhibitor, then the planar peptide bond was deformed toward the tetrahedral transition state. This lead to the conclusion that strain developed in the scissile bond upon substrate binding. Based on our structural results, we contend that the protease accommodates the products readily and the substrate residues are not so rigidly held that they would strain the scissile bond.

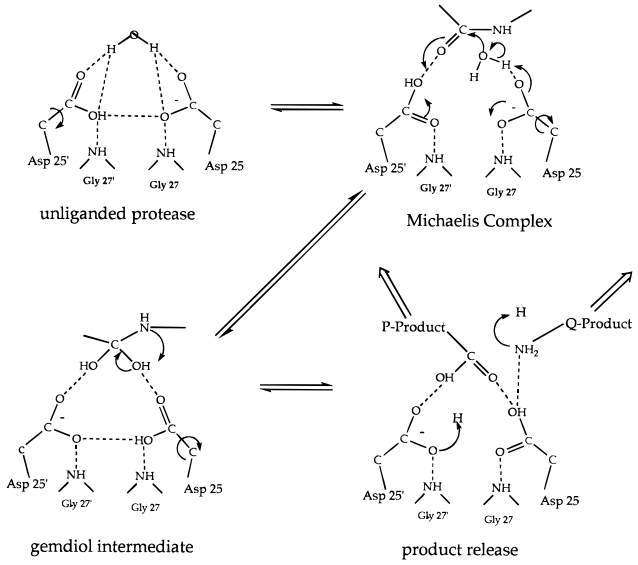

FIGURE 7: Proposed chemical mechanism with the four primary reaction intermediates represented. The unliganded protease shows the catalytic water hydrogen-bonded between the two aspartic acids. Rotations around the $C_\beta-C_\alpha$ bond are represented by an arched arrow. The structure of the Michaelis complex shows the catalytic water displaced by the scissile carbonyl, in position for nucleophilic attack. The hydroxyls of the gemdiol intermediate hydrogen bond with both catalytic aspartic acids equally. The nitrogen of the scissile bond is shown being protonated by one of the gemdiol hydrogens. Following cleavage of the peptide bond, the products are released randomly (double arrows). The amino group of the Q product becomes protonated. The carboxyl group of the P product can form the triangular interaction with the two aspartic acids.

Differences between the monomers of HIV-1 protease bound to Ac-S-L-N-F↓ may reflect conformational changes and increased thermal stability of the protein accompanying ligand binding. Some of these differences may also result from different crystal contacts made by each monomer. The "liganded" monomer is more compact than the "unliganded" monomer, reflecting contacts made between the protease and the peptide. The flaps, which close in response to ligand binding, differ conformationally by up to 1.8 Å in the liganded and unliganded monomers. Differences also occur in the base of the binding pocket and the S1 subsite. The *B* factors of the unliganded monomer are 5−10 $Å^2$ greater than for the liganded monomer. The flaps have similar *B* factors in the two monomers.

The Proteolytic Mechanism. Biochemical (Hyland et al., 1991a,b) and structural (Jaskólski et al., 1991; Silva et al., 1996) studies of HIV protease have led to proposals for its mechanism based on the mechanism of pepsin and other monomeric aspartyl proteases. We propose structures for the major steps of proteolysis derived from the product complexes, the Michaelis complex, and gemdiol intermediate. Figure 7 shows a chemical mechanism based on these structures.

There is general agreement that the ordered water molecule hydrogen-bonded between the aspartic acids in the unliganded structure becomes the catalytic nucleophile (James et al., 1992; Jaskólski et al., 1991; Suguna et al., 1987b). In Figure 7 the water is depicted forming bifurcated hydrogen bonds to the two catalytic aspartic acids. This water is displaced as a result of substrate binding and formation of a hydrogen bond between the scissile carbonyl oxygen and the protonated carboxylate of Asp 25′.

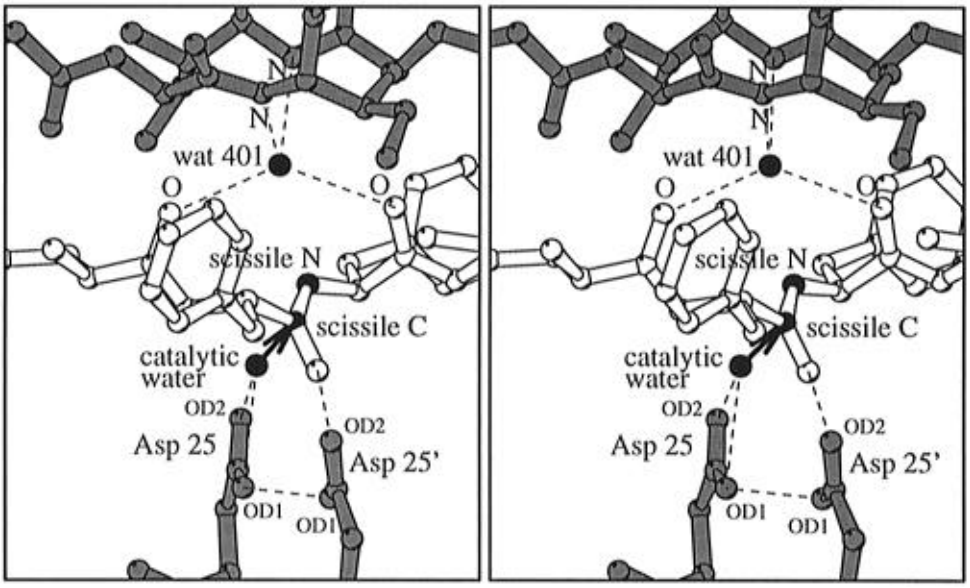

FIGURE 8: Model of the Michaelis complex. The scissile carbonyl of the substrate (white) is hydrogen-bonded (dotted lines) to Asp 25′ (gray). The catalytic water (black) forms a bifurcated hydrogen bond to the two oxygens of Asp 25. Hydrogen bonds to water 401 are shown, which also forms hydrogen bonds with the flaps (black) (cross-eyed stereo).

The position of the nucleophilic water during catalysis is usually modeled from the position of an oxygen of the putative transition state or intermediate structure, for instance from pepstatin (Bott et al., 1982; James & Sielecki, 1985) or difluoroketone (James et al., 1992; Parris et al., 1992; Veerapandian et al., 1992). We model the water near one carboxy-terminal oxygen of the P-product. Assuming that the nuclei do not move significantly during the time of the reaction, the oxygen of the attacking water will be positioned similarly before and after hydrolysis. We also assume that the carboxy terminus of the P-product Ac-S-L-N-F↓ has not shifted significantly after hydrolysis. The positions of the carbonyl and catalytic water are similar to those proposed by Suguna et al. (1987b) for the reduced isostere inhibitor and by the difluoroketone inhibitors except that the water forms hydrogen bonds with only one of the aspartic acids. The alternative to repositioning the water with substrate binding is to rotate the peptide carbonyl to attain a good angle of attack, as proposed by Jaskólski et al. (1991).

Figure 8 depicts the catalytic water modeled with the substrate. The carbonyl adopts the position within hydrogen-bonding distance of Asp 25′ (the primed monomer associated with the Q product), close to the position of one C-terminal carboxyl oxygen of the product Ac-S-L-N-F↓. The peptide carbonyl displaces the catalytic water, which forms a bifurcated hydrogen bond with Asp 25, 2.1 Å from OD1 and 3.3 Å from OD2. In this position, close to the position of the other C-terminal carboxyl oxygen of the product Ac-S-L-N-F↓, the water is hydrogen-bonded to only one of the aspartic acids and is 1.6 Å from the carbon of the scissile bond. The proximity of the water to the carbon may mean that nucleophilic attack begins as the substrate binds. The water molecule is depicted attacking the carbonyl stereospecifically at an 80° angle. This angle is close to the optimal 107° angle for nucleophilic attack of a carbonyl (Burgi et al., 1973). The reaction is stereospecific with the direction of attack being the same as predicted by Jaskólski et al. (1991).

Crystal structures with pepstatin (James & Sielecki, 1985) reduced peptide isosteres (Suguna et al., 1987b), hydroxyethylene isosteres (Veerapandian et al., 1990), and most recently difluoroketone (James et al., 1992; Parris et al., 1992; Silva et al., 1996; Veerapandian et al., 1992) have been used as models of the gemdiol intermediate. We propose that the conformation of the gemdiol intermediate is similar to that of the carboxy terminus of Ac-S-L-N-F↓, with each hydroxyl interacting with one of the catalytic aspartic acids. This is in contrast to the conformation of the difluoroketone gemdiol structures and the concerted mechanism proposed by Jaskólski et al. (1991).

In the substrate model, the scissile bond nitrogen is 4 Å from Asp 25, too far for efficient donation of a proton. The source of this proton is depicted in Figure 7 as one of the gemdiol hydroxyls. Silva et al. (1996) have proposed an *anti* to *gauche* transition of the gemdiol modeled from an HIV-1−difluoroketone structure by rotation around the C−N bond. This rotation would bring the nitrogen closer to the catalytic aspartic acids, although the associated conformational change of the backbone and side chains is not evident in our model of the substrate from the products.

These modeled structures provide a framework for the proteolytic mechanism. The major differences between the mechanism proposed by these structures and that proposed by Hyland et al. (1991a,b) are the positioning of the catalytic water and the symmetrical disposition of the gemdiol intermediate relative to the catalytic aspartic acids. Our model of the substrate therefore shows that the two aspartic acids are not equivalent. It suggests that Asp 25, associated with the P product, acts as the general base extracting a proton from the water, while Asp 25′ acts as the general acid, donating a proton to the carbonyl of the scissile bond 2.6 Å away. This would echo the in equivalence of the catalytic aspartic acids in the monomeric aspartyl proteases; Asp 213 is thought to act as the general base in penicillopepsin, while Asp 33 acts as the general acid (James et al., 1992).

CONCLUSIONS

Clearly, much can be learned regarding the mechanism of catalysis from detailed analyses of protease−product complexes. Subsequent structural studies will address factors

other than substrate strain contributing to catalysis. Particularly, the role of conformational changes in the protein will be investigated. Ultimately these studies will contribute to an explanation of the complex specificity patterns displayed by many aspartyl proteases.

ACKNOWLEDGMENT

We thank Tom Meek for helpful discussions on HIV protease function and product interactions. We also thank Earl Rutenber for advice on structure determination and interpretation and Earl Rutenber, Carleton Sage, and Peter Sayre for critical review of the manuscript.

REFERENCES

Antonov, V. K., Ginodman, L. M., Kapitannikov, Y. V., Barshevskaya, T. N., Gurova, A. G., & Rumsh, L. D. (1978) *FEBS Lett. 88*, 87−90.

Antonov, V. K., Ginodman, L. M., Rumsh, L. D., Kapitannikov, Y. V., Barshevskaya, T. N., Yavashev, L. P., Gurova, A. G., & Volkova, L. I. (1981) *Eur. J. Biochem. 117*, 195−200.

Bott, R., Subramanian, E., & Davies, D. R. (1982) *Biochemistry 21*, 6956−6962.

Brünger, A. T. (1992) *Nature 355*, 472−474.

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

Burgi, H. B., Dunitz, J. D., & Shefter, E. (1973) *J. Am. Chem. Soc. 95*, 5065−5067.

Chambers, J. L., & Stroud, R. M. (1977) *Acta Crystallogr. B33*, 1824−1837.

Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 189−215.

Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., & Kraut, J. (1990) *Biochemistry 29*, 9467−9479.

Debouck, C. (1991) *Adv. Exp. Med. Biol. 306*, 407−415.

Fruton, J. S. (1976) in *Methods in Enzymology* (Meister, A., Ed.) pp 1-35, John Wiley and Sons, New York.

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.

Griffiths, J. T., et al. (1992) *Biochemistry 31*, 5193−5200.

Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., & Oroszlan, S. (1988) *Human Retrovirus, Cancer and AIDS: Approaches to Prevention and Therapy*, Liss, New York.

Hyland, L. J., Tomaszek, T. J., & Meek, T. D. (1991a) *Biochemistry 30*, 8454−8463.

Hyland, L. J., Tomaszek, T. J., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., Culp, J. S., et al. (1991b) *Biochemistry 30*, 8441−8453.

Ido, E., Han, H. P., Kezdy, F. J., & Tang, J. (1991) *J. Biol. Chem. 266*, 24359−24366.

James, M. N. G., & Sielecki, A. R. (1985) *Biochemistry 24*, 3701−3713.

James, M. N. G., Hsu, I.-N., & Delbaere, L. T. J. (1977) *Nature 267*, 808−813.

James, M. N. G., Sielecki, A. R., Salituro, F., Rich, D., & Hofmann, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6137−6141.

James, M. N. G., Sielecki, A. R., Hayakawa, K., & Gelb, M. H. (1992) *Biochemistry 31*, 3872−3886.

Jaskólski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., & Wlodawer, A. (1991) *Biochemistry 30*, 1600−1609.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686−4690.

McKeever, B. M., Navia, M. A., Fitzgerald, P. M., Springer, J. P., Leu, C. T., Heimbach, J. C., Herbert, W. K., Sigal, I. S., & Darke, P. L. (1989) *J. Biol. Chem. 264*, 1919−1921.

McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrikson, R. L., & Tarpley, W. G. (1990) *Science 247*, 454−456.

Meek, T. D., Rodriguez, E. J., & Angeles, T. S. (1994) *Methods Enzymol. 241*, 127−156.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B., & Wlodawer, A. (1989) *Science 246*, 1149−1152.

Otwinowski, Z. (1993) in *CCP4 Study Weekend: Data Collection and Processing*, 29−30 Jan 1993 (Sawyer, L., Isaacs, N., & Bailey, S., Eds.) pp 56−62, SERC Daresbury Laboratory, England.

Parris, K. D., Hoover, D. J., Damon, D. B., & Davies, D. R. (1992) *Biochemistry 31*, 8125−8141.

Pearl, L., & Blundell, T. (1984) *FEBS Lett. 174*, 96−101.

Polgár, L. (1987) *FEBS Lett. 219*, 1−4.

Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., & Kézdy, F. J. (1991) *J. Biol. Chem. 266*, 14554−14561.

Rich, D. H., Green, J., Toth, M. V., Marshall, G. R., & Kent, S. B. H. (1990) *J. Med. Chem. 33*, 1285−1288.

Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. M., Hirel, P.-H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., & Kay, J. (1990) *J. Biol. Chem. 265*, 7733−7736.

Rosé, J. R., Salto, R., & Craik, C. S. (1993) *J. Biol. Chem. 268*, 11939−11945.

Rose, R. B., Rosé, J. R., Salto, R., Craik, C. S., & Stroud, R. M. (1993) *Biochemistry 32*, 12498−12507.

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., & Stroud, R. M. (1993) *J. Biol. Chem. 268*, 15343−15346.

Sack, J. (1988) *J. Mol. Graphics 6*, 225.

Schecter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.

Seelmeier, S., Schmidt, H., Turk, B., & von der Helm, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6612−6616.

Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., & James, M. N. G. (1990) *J. Mol. Biol. 214*, 143−170.

Silva, A. M., Cachau, R. E., Sham, H. L., & Erickson, J. W. (1996) *J. Mol. Biol. 255*, 321−346.

Stroud, R. M., & Fauman, E. B. (1995) *Protein Sci. 4*, 2392−2404.

Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., & Davies, D. R. (1987a) *J. Mol. Biol. 196*, 877−900.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., & Davies, D. R. (1987b) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7009−7013.

Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805−8809.

Tomasselli, A. G., & Heinrikson, R. L. (1994) *Methods Enzymol. 241*, 279−301.

Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., & Anderson, P. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8387−8391.

Veerapandian, B., Cooper, J. B., Sali, A., & Blundell, T. L. (1990) *J. Mol. Biol. 216*, 1017−1029.

Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., & Hoover, D. J. (1992) *Protein Sci. 1*, 322−328.

Wlodawer, A., & Erickson, J. W. (1993) *Annu. Rev. Biochem. 62*, 543−585.

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.

Zhao, B., Winborne, E., Minnich, M. D., Culp, J. S., Debouck, C., & Abdel-Meguid, S. S. (1993) *Biochemistry 32*, 13054−13060.

BI9612733