Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8387–8391, September 1993
Biochemistry

Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures

(AIDS/retrovirus/aspartic protease)

LIANG TONG*†, SUSAN PAV‡, CHRISTOPHER PARGELLIS‡, FLORENCE DÔ§, DANIEL LAMARRE§, AND PAUL C. ANDERSON¶

Departments of *Medicinal Chemistry and ‡Biochemistry, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road/P.O. Box 368, Ridgefield, CT 06877; and Departments of §Biochemistry and ¶Chemistry, Bio-Méga/Boehringer Ingelheim Research, Inc., 2100 rue Cunard, Laval, PQ, Canada H7S 2G5

Communicated by Michael G. Rossmann, June 7, 1993

ABSTRACT **The crystal structure of HIV-2 protease in complex with a reduced amide inhibitor [BI-LA-398; Phe-Val-Phe-Ψ(CH_2NH)-Leu-Glu-Ile-amide] has been determined at 2.2-Å resolution and refined to a crystallographic *R* factor of 17.6%. The rms deviation from ideality in bond lengths is 0.018 Å and in bond angles is 2.8°. The largest structural differences between HIV-1 and HIV-2 proteases are located at residues 15–20, 34–40, and 65–73, away from the flap region and the substrate binding sites. The rms distance between equivalent C_α atoms of HIV-1 and HIV-2 protease structures excluding these residues is 0.5 Å. The shapes of the S_1 and S_2 pockets in the presence of this inhibitor are essentially unperturbed by the amino acid differences between HIV-1 and HIV-2 proteases. The interaction of the inhibitor with HIV-2 protease is similar to that observed in HIV-1 protease structures. The unprotected N terminus of the inhibitor interacts with the side chains of Asp-29 and Asp-30. The glutamate side chain of the inhibitor forms hydrogen bonds with the main-chain amido groups of residues 129 and 130.**

HIV-1, the causative agent of AIDS, encodes an aspartic protease that is essential for its life cycle (1). The HIV-1 protease is responsible for the processing of the gag structural polyprotein and the release of virally encoded enzymes (protease, reverse transcriptase, and integrase) from the gag–pol fusion protein (2). Consequently, the HIV-1 protease has been an important target for the design of therapeutic agents to treat HIV-1 infections (3). Much structural information is currently available for the interactions between HIV-1 protease and inhibitors due to the large number of crystal structures that have been determined (for a review, see ref. 4). The structures showed that these peptidomimetic inhibitors are bound in extended conformations, occupying the substrate binding pockets of the protease. Hydrogen bonds are made between the inhibitor and the main-chain atoms of the residues in the flap region.

Whereas HIV-1 is primarily found in the United States, Europe, and Asia, HIV-2 is mostly limited to regions of western Africa (5). The proteases of HIV-1 and HIV-2 share 50% sequence identity (6, 7). In contrast to HIV-1 protease, there have been few reports on the crystal structure of HIV-2 protease.‖ A model for HIV-2 protease has been proposed based on the crystal structures of HIV-1 protease (7). We report here the crystal structure at 2.2-Å resolution of HIV-2 protease in complex with a reduced amide inhibitor [BI-LA-398; Phe-Val-Phe-Ψ(CH_2NH)-Leu-Glu-Ile-amide].** This compound has an IC_{50} value of ≤3.1 nM against HIV-1 protease and ≤3.5 nM against HIV-2 protease, assayed at pH 5.5 in the presence of 700 mM KCl (D. Thibeault and D.L., unpublished results).

MATERIALS AND METHODS

Crystallization of HIV-2 Protease. The details of the cloning, expression, and purification of HIV-2 protease will be presented elsewhere (S.P., L.T., K. Lubbe, C.P., F.D., and D.L., unpublished data). Briefly, HIV-2 protease (rod isolate) was expressed in *Escherichia coli* BL21 cells. After lysis of the cell paste, soluble HIV-2 protease was isolated and purified by ammonium sulfate precipitation, size-exclusion, and pepstatin-agarose affinity chromatography methods. Purified HIV-2 protease was concentrated to 6 mg/ml in Dulbecco's phosphate-buffered saline (GIBCO/BRL, 310-4200) at 1:10 dilution (pH 7.4) and 0.02% NaN_3. Stock solution of the inhibitor (BI-LA-398) at 20 mg/ml in dimethyl sulfoxide (DMSO) was further diluted in 10% (vol/vol) DMSO to 40- to 60-fold molar excess of the enzyme concentration. Final dilution of the inhibitor into the enzyme solution resulted in a 4:1 to 6:1 molar excess of the inhibitor, with a final DMSO concentration of 4–6%. The enzyme/inhibitor mixture was incubated for 1 hr at room temperature in a 0.5-ml siliconized microcentrifuge tube. The incubation mixture was used for a hanging drop experiment and a 2-μl excess of the mixture was left in the tube at room temperature. It was discovered a week later that crystals had formed in the tube, directly from the incubation mixture. Subsequent experiments were able to reproduce these crystals.

Data Collection and Data Processing. The x-ray diffraction data to 2.2-Å resolution were collected at 4°C on an R-Axis (Molecular Structure, The Woodlands, TX) mounted on a Rigaku RU-200 rotating anode generator operated at 50 kV and 100 mA. The crystal-to-detector distance was 110 mm. The oscillation angle per frame was 1.5° and the exposure time was 55 min. A total of 47 frames were collected, indexed, and processed, and the reflection data were merged with the R-Axis software. The space group of the crystal was determined to be $P4_1$ or $P4_3$ with unit-cell parameters of $a = b =$ 55.1 Å and c = 138.9 Å. A total of 85,564 observations was used to produce the final data set of 18,007 reflections to 2.2-Å resolution (82% complete), with an overall R_{merge} value on I (for $I > \sigma_I$ observations; where I is reflection intensity) of 6.6% (Table 1).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HIV, human immunodeficiency virus.

†To whom reprint requests should be addressed.

‖Mulichak, A. M. & Watenpaugh, K. D., American Crystallographic Association Annual Meeting, August 9–14, 1992, Pittsburgh, PA, abstr. W03.

**The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 2MIP).

Table 1. Summary of crystallographic information

Parameter	Value
Number of frames	47
Number of observations	85,564
Number of unique reflections	18,007
R_{merge} ($I > \sigma_I$ observations)	6.6%
Number of reflections 6–2.2 Å	16,845 (80% complete)
R factor for 6- to 2.2-Å reflection data	17.6%
Number of reflections at 2.3–2.2 Å	1665 (64% complete)
R factor for 2.3- to 2.2-Å reflection data	25.0%
Number of water molecules	90
rms deviation in bond lengths	0.018 Å
rms deviation in bond angles	2.8°

Structure Determination by Molecular Replacement. There are two HIV-2 protease dimers in the asymmetric unit of this crystal form, giving a volume per unit molecular weight (V_m) value of 2.5 Å^3/dalton (8). The atomic model at 2.2-Å resolution for HIV-2 protease dimer in complex with a different inhibitor (L.T., S.P., and P.C.A., unpublished data), which was crystallized by the hanging drop method from 0.5 M NaCl/0.1 M NaOAc, pH 5.4, with one HIV-2 protease dimer in the asymmetric unit of space group $P4_32_12$ ($a = b = 62.6$ Å, $c = 115.8$ Å), was placed in a $P1$ cell with $a = b = c = 100$ Å and $\alpha = \beta = \gamma = 90°$. The inhibitor and the water molecules were removed from the model and the atomic temperature factors were set to 15 Å^2. Reflection data between 10- and 3.5-Å resolution were used in the rotation function calculation with the program GLRF (9). A total of 1036 of the 4862 observed reflections in this resolution range were saved as large terms and the radius of integration was 20 Å. Two unique peaks were found in the rotation function map, with peak heights of 999 (arbitrary scale, 9.8σ above the map average) and 870 (8.3σ), respectively. The next peak in the rotation function had a height of 521 (4.2σ). Some 992 out of the 4613 observed reflections between 8- and 3.5-Å resolution were used in the "Patterson correlation" translation function calculation (10, 11). For the first rotation function peak, a clear translation function solution was obtained in space group $P4_3$. The top peak was 11σ above the map average and the second peak was 4σ above the average. With the first dimer correctly placed in the unit cell, the position of the second dimer was then determined. The top peak in this translation function was 18σ above the average, whereas the second peak was 9σ above the average. Reasonable packing of the molecules in the unit cell confirmed the correctness of the solution.

Structure Refinement. The crystallographic refinement was carried out with the program X-PLOR (12). Rigid-body refinement using 6.0- to 3.5-Å resolution reflection data lowered the R factor from 37.5% to 31.2%. The overall atomic temperature factor was set at 15 Å^2. A simulated annealing slow cooling refinement reduced the R factor to 24.0% for 16,845 reflections between 6- and 2.2-Å resolution. The atomic model was examined against the $2F_o - F_c$ electron density map (where F_o and F_c are the observed and calculated structure factors, respectively) and minor changes were made with the program FRODO (13). The inhibitor molecules were modeled into the $2F_o - F_c$ density, as were two tetrahedrally coordinated water molecules situated between the inhibitors and the protein. The polarity of the inhibitor molecules was tentatively assigned based on comparisons of the electron density at the S_2 and S_2' pockets. A cycle of refinement in the presence of the inhibitors and the two water molecules lowered the R factor to 21.7%, with the overall B factor set to 19 Å^2 as determined from refinement. The R factor became 20.1% after restrained individual temperature factor refinement. Difference electron density maps were then examined to locate solvent molecules. Both the position and the temperature factor of the solvent molecules were subsequently refined. The alternative orientation of the inhibitor molecules was modeled by applying the local twofold axis of the protease dimer to the original orientation. Two parallel positional and temperature factor refinements were carried out, one for each orientation of the inhibitor. The temperature factor values for the $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, and C_ζ atoms of Phe-3 and the C_δ, $O_{\varepsilon 1}$, and $O_{\varepsilon 2}$ atoms of Glu-5 in one orientation are roughly 10 Å^2 higher than in the other one. An occupancy factor of 0.55 and 0.45 was assigned to the two orientations. One further cycle of positional refinement with the program TNT (14) with both orientations present produced the current atomic model.

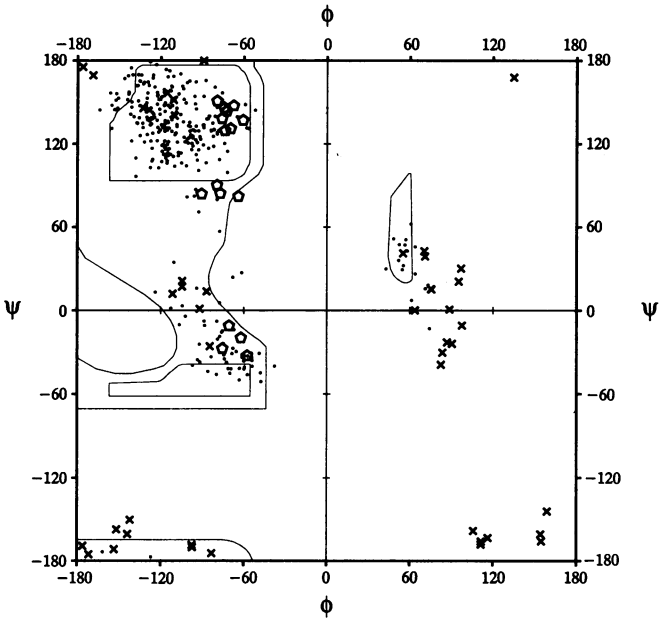

FIG. 1. Ramachandran plot for the four molecules in the current atomic model for HIV-2 protease. Glycine residues are shown as crosses, proline residues are shown as pentagons, and the others are shown as dots.

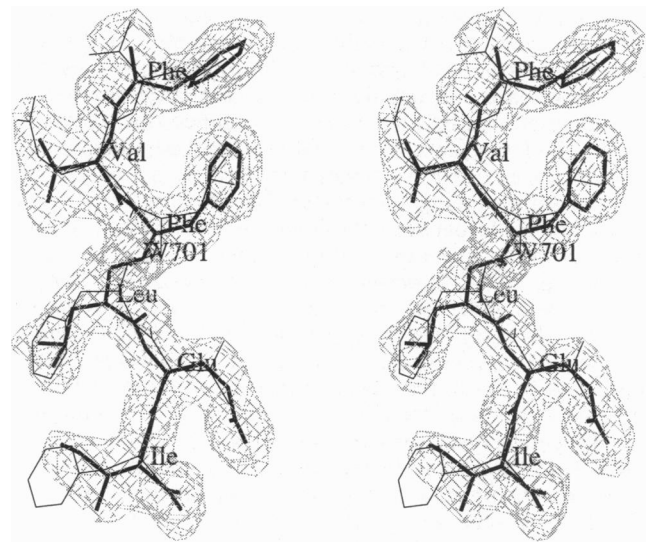

FIG. 2. Stereo diagram showing the $2F_o - F_c$ electron density map for the inhibitor molecule after refinement, contoured at 0.5σ. The tetrahedrally coordinated water molecule (labeled W701) is also shown. The alternative orientation of the inhibitor molecule is shown as thin lines.

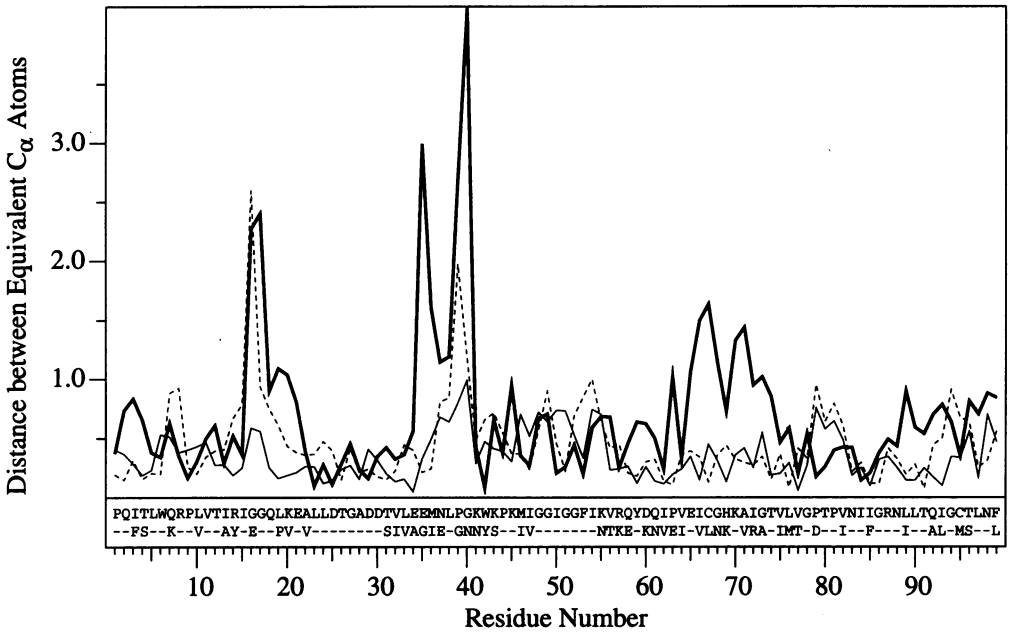

FIG. 3. Plot of the distances (in Å) between equivalent C_α atoms of residues in HIV-1 (16) and HIV-2 protease crystal structures (thick lines), between two monomers (residues 1–99 and 101–199) of the HIV-2 protease dimer (thin lines), and between two monomers of the HIV-1 protease dimer (dashed thin lines). (*Inset*) The sequence alignment of HIV-1 (sf2 isolate, upper sequence) and HIV-2 (rod isolate, lower sequence) proteases. A dash represents an amino acid identity between HIV-2 and HIV-1 proteases.

RESULTS AND DISCUSSION

The Quality of the HIV-2 Protease Structure. The current atomic model for HIV-2 protease consists of 396-aa residues from the two HIV-2 protease dimers (numbered 1–99 and 101–199 for the first dimer and 201–299 and 301–399 for the second dimer), the two inhibitor molecules (numbered 401–406 and 411–416 for the two orientations bound to the first dimer and 451–456 and 461–466 bound to the second dimer), and 90 water molecules. The rms deviation from ideality in bond lengths is 0.018 Å and in bond angles is 2.8°. The overall *R* factor is 17.6% for 16,845 reflections between 6- and 2.2-Å resolution, which accounts for 80% of the total number of reflections expected. The *R* factor is 25.0% in the resolution shell 2.3–2.2 Å, where the reflection data is 64% complete (Table 1). The overall average temperature factor for the protein atoms is 21 Å^2. The error in the atomic coordinates is estimated to be 0.25 Å, based on a Luzzati plot (ref. 15 and data not shown), though atoms with higher temperature factors should be expected to have larger positional errors. All the amino acid residues assume energetically favorable main-chain conformations (Fig. 1).

The inhibitor molecules are bound to the HIV-2 protease in the extended conformation (Fig. 2). There is good electron density for Phe-3 and Glu-5 residues in the alternative orientation, whereas the electron density for Phe-1 is much weaker beyond the C_δ atoms. The Phe-1 residue in the original orientation is involved in extensive crystal packing interaction with residue 141. This twofold disorder of the inhibitor molecules should be due largely to the disorder in packing the HIV protease dimers in either of the two orientations. A preferred orientation for the inhibitors can be expected as the protein residues that form the S_2 and S_3 pockets are also involved in crystal packing interactions.

The angle of rotation of the local twofold axis is 178.2° for the first dimer (residues 1–99 and 101–199) and is 178.7° for the second dimer. The translation elements along the axes are 0.0 Å. The rms distances between the 99 equivalent C_α atoms of any pair of the four monomers range from 0.35 to 0.4 Å, suggesting the four molecules are essentially identical. The largest difference is 1.1 Å, between residues 152 and 252.

A total of 90 solvent molecules are included in the structure model. The temperature factor values of the water molecules vary between 9 and 43 Å^2. The tetrahedrally coordinated water molecule that was first identified in HIV-1 protease inhibitor complex structures (4) is also found in this structure (numbered 501 and 701), making two hydrogen bonds each to the inhibitor and the protein (see Fig. 6).

Comparison with HIV-1 Protease Structures. The rms distance between the 99 equivalent C_α atomic positions of HIV-2 protease (residues 101–199) and HIV-1 protease complexed with the reduced amide inhibitor MVT-101 (ref. 16; Protein Data Bank entry, 4HVP) (residues B1–B99) is 1.0 Å (Figs. 3 and 4). The largest differences in C_α atomic positions are located at residues 15–20, 34–40, and 65–73 (Figs. 3 and 4), consistent with an earlier report.‖ If these residues are excluded from the calculation, the rms deviation is 0.5 Å for 77 C_α pairs. These residues form three loops on the surface of the protease and are located away from the flap region (Fig. 4). Thus it would be unlikely for the differences at these residues to have any direct effect on inhibitor binding.

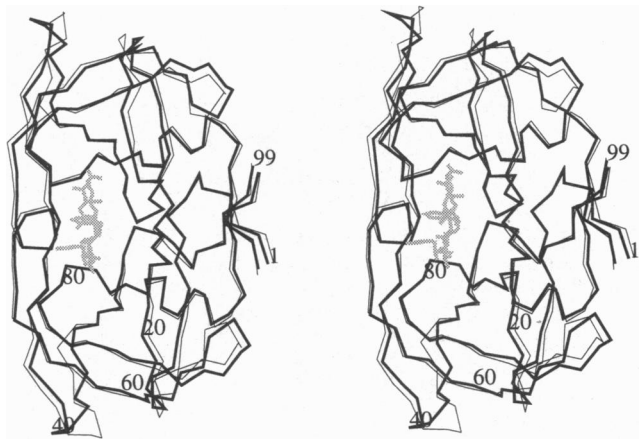

FIG. 4. Stereo diagram showing the superposition of the structures of HIV-2 (thick lines) and HIV-1 (16) (thin lines) protease dimers. The inhibitor BI-LA-398 is shown by a stippled line.

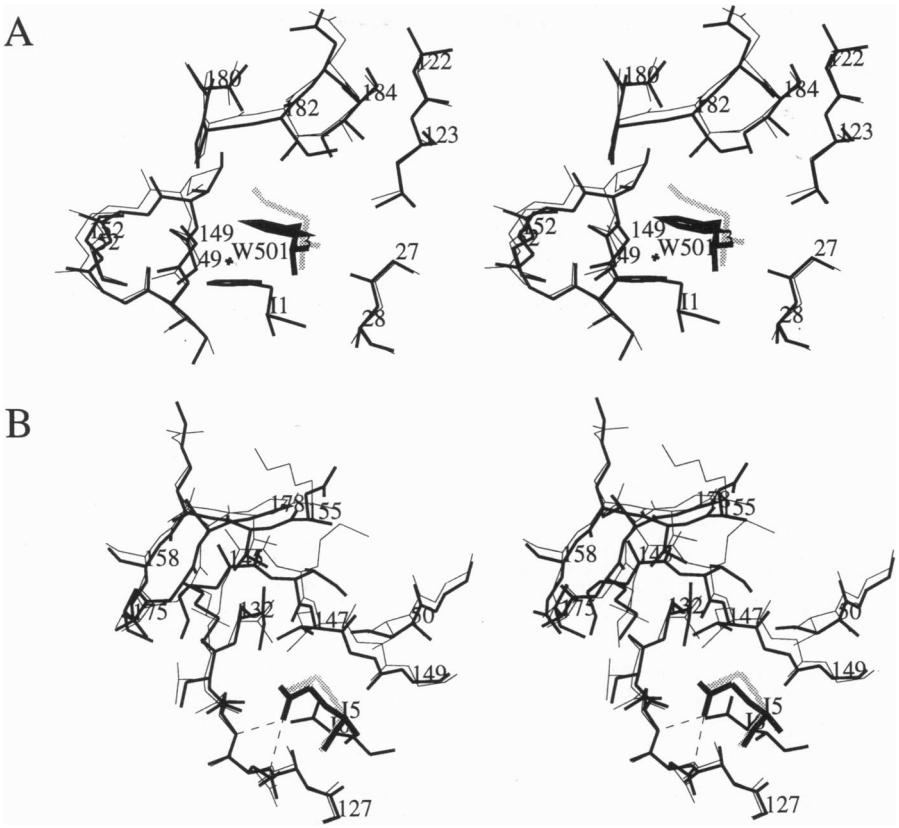

FIG. 5. Stereo diagram showing the residues forming the S_1 substrate binding pocket in HIV-2 (thick lines) and HIV-1 (16) (thin lines) protease structures (*A*) and the residues forming the S_2' pocket in HIV-2 (thick lines) and HIV-1 (thin lines) proteases (*B*). The P_1 (phenylalanine and norleucine) and P_2' (glutamate and glutamine) residues are also shown for BI-LA-398 (dark lines) and MVT-101 (stippled lines), respectively. The hydrogen bonds between the glutamate side chain of the inhibitor and the main-chain amido groups of residues 129 and 130 are shown as dashed lines. The inhibitor residues are labeled I1, I3, I5, and I6.

To identify which of these regions are genuine differences between HIV-1 and HIV-2 protease structures, the two molecules of the HIV-1 protease dimer and their temperature factors were compared. The rms distances in C_α positions of the two monomers for three HIV-1 protease structures are 0.5 Å (5HVP; ref. 17) and 0.6 Å (4HVP and 7HVP; ref. 18), with differences of up to 2 Å at residues 15–18 and 37–40 (Fig. 3). Residues 16–19 are involved in crystal packing in the HIV-2 protease crystals, whereas the corresponding residues in the HIV-1 protease structure are not involved in crystal packing and show large structural differences between the two monomers (Fig. 3). Therefore, the differences at residues 15–20 between HIV-1 and HIV-2 proteases may be a crystal packing artifact.

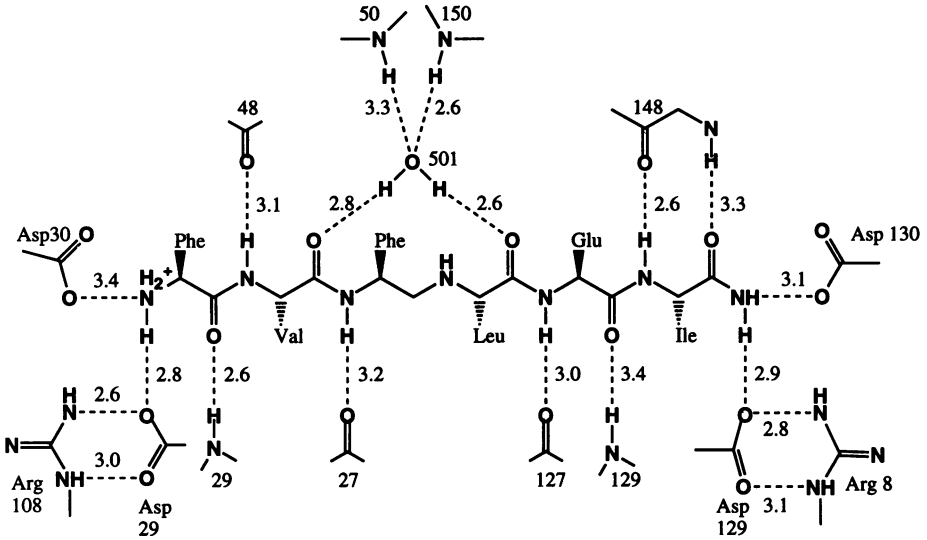

FIG. 6. Hydrogen-bonding interactions between the BI-LA-398 inhibitor and the HIV-2 protease dimer. Potential hydrogen bonds are presented as dashed lines, with the corresponding hydrogen-bonding distances (from O to N) shown in Å. The tetrahedrally coordinated water molecule is labeled 501.

Residue Glu-35 forms a salt bridge with residue Arg-57 in the HIV-1 protease structures. In HIV-2 protease, it was proposed based on a modeling study that an equivalent salt bridge is formed between residues Glu-37 and Lys-57 as residue 35 is glycine in HIV-2 protease (7). However, the distances between the Glu-37 and the Lys-57 side chains range between 4.8 and 6.0 Å in the four monomers in the crystal structure. Moreover, the extensive rearrangement in the main-chain atomic positions was not predicted by that study. Residues 37–40, 137–140, and 237–240 are not involved in crystal packing. Similarly, the differences at residues 60–73 occurred as a result of amino acid substitutions (notably C67L, A71V, I85F, and F99L) between HIV-1 and HIV-2 proteases.

The formation of the dimer leads to the burial of 1700 Å^2 surface area per monomer for HIV-2 protease, as compared to 1800 Å^2 for HIV-1 protease (19). Of the 17 residues that each contributes 40 Å^2 or more to the dimer interface, only 4 residues are different between HIV-1 and HIV-2 proteases—I3F, C95M, T96S, and F99L.

The Substrate Binding Pockets. There are 5 amino acid differences between HIV-1 and HIV-2 proteases in the residues that form the S_1 and S_2 substrate binding pockets—V82I in the S_1 and S_1' pockets and V32I, I47V, V56T, and L76M in the S_2 and S_2' pockets. These changes caused small disturbances to the shape of the binding pockets (Fig. 5). As in the HIV-1 protease structure (16), the side chains of the catalytic residues Asp-25 and Asp-125 are located close to the reduced amide bond, with the shortest distance between the side chain O_δ atoms at 2.4 Å. The flap-to-flap hydrogen bond between the amido group of residue 51 and the carbonyl group of residue 150 is maintained in HIV-2 protease structure (Fig. 5). Of special interest is the binding of the charged glutamate side chain of the inhibitor in the S_2' pocket. One of the side-chain oxygen atoms forms hydrogen bonds with the main-chain amido groups of residues 129 and 130 (Fig. 5) (20). The N terminus of this inhibitor molecule is not protected, and the positively charged ammonium group interacts with the side chains of Asp-29 and Asp-30 (Fig. 6). An equivalent interaction is found at the C terminus, between the amido nitrogen group and the side chains of Asp-129 and Asp-130. It should be pointed out that the inhibitor under study here is highly potent against both proteases. There are many inhibitors that show drastically different activity against the two proteases (6). It would be interesting to carry out comparative structural studies for such inhibitors to see whether they can induce different conformations in HIV-1 and HIV-2 proteases or their different inhibitory activity is due to other mechanisms.

Note Added in Proof. A paper on HIV-2 protease structure has recently appeared in the literature (21).

We thank Klaus Lubbe for carrying out the fermentation process, Diane Thibeault for the HIV-1 and HIV-2 protease inhibition assays, Sylvie Goulet and Dr. Neil Moss for the synthesis of BI-LA-398, and Thomas Warren and Michael Sintchak for helpful discussions. The HIV-2_{rod} phage was obtained from Dr. Ronald Desrosiers through the AIDS Research and Reference Reagent Program, National Institutes of Health. We thank Drs. Julian Adams, Peter Farina and Peter Grob for their enthusiasm and encouragement in this project.

1. Kohl, N. E., Emini, E. A., Scheif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Skolnick, E. M. & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4686–4690.
2. Peterlin, B. M. & Luciw, P. A. (1988) *AIDS* **2** Suppl. 1, S29–S40.
3. Mitsuya, H., Yarchoan, R. & Broder, S. (1990) *Science* **249**, 1533–1544.
4. Wlodawer, A. & Erickson, J. (1993) *Annu. Rev. Biochem.* **62**, 543–585.
5. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) *Nature (London)* **326**, 662–669.
6. Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. J., Bannow, C., Reardon, I. M., Howe, W. J., DeCamp, D. L., Craik, C. S. & Heinrikson, R. L. (1990) *J. Biol. Chem.* **265**, 14675–14683.
7. Gustchina, A. & Weber, I. T. (1991) *Proteins* **10**, 325–339.
8. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497.
9. Tong, L. & Rossmann, M. G. (1990) *Acta Crystallogr. Sect. A* **46**, 783–792.
10. Crowther, R. A. & Blow, D. M. (1967) *Acta Crystallogr.* **23**, 544–548.
11. Tong, L. (1993) *J. Appl. Crystallogr.*, in press.
12. Brunger, A. T. (1992) *X-Plor Manual* (Yale Univ., New Haven, CT), Version 3.0.
13. Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272.
14. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43**, 489–501.
15. Luzzati, V. (1952) *Acta Crystallogr.* **5**, 802–810.
16. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H. & Wlodawer, A. (1989) *Science* **246**, 1149–1152.
17. Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C.-T., Herber, W. K., Dixon, R. A. F. & Darke, P. L. (1990) *J. Biol. Chem.* **265**, 14209–14219.
18. Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H. & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8805–8809.
19. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
20. Griffiths, J. T., Phylip, L. H., Konvalinka, J., Strop, P., Gustchina, A., Wlodawer, A., Davenport, R. J., Briggs, R., Dunn, B. M. & Kay, J. (1992) *Biochemistry* **31**, 5193–5200.
21. Mulichak, A. M., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Curry, K. A., Tomich, C.-S, Thaisrivongs, S., Sawyer, T. K. & Watenpaugh, K. D. (1993) *J. Biol. Chem.* **268**, 13103–13109.