Protein Science (1992), *1*, 1061-1072. Cambridge University Press. Printed in the USA.

Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling

NARMADA THANKI,[1] J.K. MOHANA RAO,[1] STEPHEN I. FOUNDLING,[1,3] W. JEFFREY HOWE,[2] JOSEPH B. MOON,[2] JOHN O. HUI,[2] ALFREDO G. TOMASSELLI,[2] ROBERT L. HEINRIKSON,[2] SUVIT THAISRIVONGS,[2] AND ALEXANDER WLODAWER[1]

[1] Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21702
[2] Upjohn Laboratories, The Upjohn Company, Kalamazoo, Michigan 49001

(RECEIVED February 24, 1992; REVISED MANUSCRIPT RECEIVED March 27, 1992)

Abstract

The structure of a crystal complex of recombinant human immunodeficiency virus type 1 (HIV-1) protease with a peptide-mimetic inhibitor containing a dihydroxyethylene isostere insert replacing the scissile bond has been determined. The inhibitor is Noa-His-HchΨ[CH(OH)CH(OH)]Vam-Ile-Amp (U-75875), and its K_i for inhibition of the HIV-1 protease is <1.0 nM (Noa = 1-naphthoxyacetyl, Hch = a hydroxy-modified form of cyclohexylalanine, Vam = a hydroxy-modified form of valine, Amp = 2-pyridylmethylamine). The structure of the complex has been refined to a crystallographic *R* factor of 0.169 at 2.0 Å resolution by using restrained least-squares procedures. Root mean square deviations from ideality are 0.02 Å and 2.4°, for bond lengths and angles, respectively. The bound inhibitor diastereomer has the R configurations at both of the hydroxyl chiral carbon atoms. One of the diol hydroxyl groups is positioned such that it forms hydrogen bonds with both the active site aspartates, whereas the other interacts with only one of them. Comparison of this X-ray structure with a model-built structure of the inhibitor, published earlier, reveals similar positioning of the backbone atoms and of the side-chain atoms in the P2-P2′ region, where the interaction with the protein is strongest. However, the X-ray structure and the model differ considerably in the location of the P3 and P3′ end groups, and also in the positioning of the second of the two central hydroxyl groups. Reconstruction of the central portion of the model revealed the source of the hydroxyl discrepancy, which, when corrected, provided a P1-P1′ geometry very close to that seen in the X-ray structure.

Keywords: human immunodeficiency virus; inhibitor; molecular modeling; protease

Human immunodeficiency virus type 1 protease (HIV-1 PR) has now become one of the best characterized enzymes in terms of both structure and function. It serves as a prototype template for de novo design of inhibitors based solely upon structural principles. The structures of a fairly large number of inhibitor complexes of HIV-1 PR are now available (for review, see Wlodawer et al., 1992). These include nonspecific aspartic protease inhibitors such as acetyl pepstatin and pepstatin; peptides patterned after cleavage sites found in the HIV-1 virion and modified by insertions of nonscissile peptide bond replacements, blocked termini, and unnatural amino acids; and quasisymmetric peptide-based inhibitors. All of these inhibitors bind to the protease in a similar manner. Due to the symmetric nature of the enzyme, some of them are able to bind in two possible orientations (Fitzgerald et al., 1990; Bone et al., 1991; M. Miller, pers. comm.; J.K.M. Rao, unpubl.).

An inhibitor of considerable importance relative to its high binding affinity for the HIV-1 protease and its potent antiviral activity (Ashorn et al., 1990; Thaisrivongs et al., 1991) is U-75875, the structure of which is shown in Figure 1 and Kinemage 1. One interesting structural feature of U-75875 is its dihydroxyethylene insert mim-

Reprint requests to: Alexander Wlodawer, Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21702.
[3] Present address: Crystallography Laboratory, Protein Studies Program, Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, Oklahoma 73104.

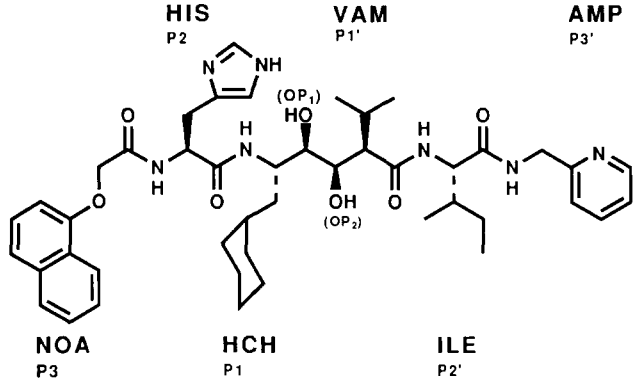

Fig. 1. Schematic diagram of the inhibitor, U-75875, showing the P3 to P3′ positions. NOA, 1-naphthoxyacetyl; HCH, a hydroxy-modified form of cyclohexylalanine (Cha); VAM, a hydroxy-modified form of valine; AMP, 2-pyridylmethylamine.

icking the transition state of substrate for the enzyme. None of the inhibitor crystal structures published to date have included a diol insert, and it was of interest to determine how the pair of hydroxyl groups, referred to herein as OP1 and OP2, would be positioned relative to the carboxyl groups of the catalytic aspartyl residues. In addition, the mode of binding of this inhibitor had already been studied by molecular modeling (Thaisrivongs et al., 1991) and the accuracy of predictions needed to be tested in order to ascertain the reliability of such studies. Potential binding modes for U-75875 derived from these modeling studies resembled other inhibitor X-ray structures in terms of side-chain positioning and hydrogen-bonding patterns, but the hydroxyl groups of the diol insert were disposed asymmetrically with respect to the active site (Thaisrivongs et al., 1991). Because of the uniqueness of the diol insert and the tightness with which the inhibitor binds to the enzyme, we undertook an X-ray crystallographic study of the enzyme/inhibitor complex. This paper reports the crystal structure of the HIV-1 protease/U-75875 complex and provides a comparison of the inhibitor crystal structure with the structure previously developed by model-building methods.

Results

Description of the refined structure of the complex

The final refined structure of the protease–inhibitor complex includes 1,516 nonhydrogen enzyme atoms with $\langle B_{iso} \rangle = 19.0$ Å^2, 59 nonhydrogen inhibitor atoms with $\langle B_{iso} \rangle = 15.6$ Å^2, and 90 water molecules with $\langle B_{iso} \rangle = 32.4$ Å^2. Geometry of the structure is summarized in Table 1. The solvent molecules are numbered such that they are sorted according to increasing B values, with the water found between the inhibitor and the flaps of the enzyme retaining the designation of Wat 301.

Table 1. *Final statistics of the protease–inhibitor structure*[a]

R factor	0.169
Weights	σ_F^{-2}
With	$\sigma_F = 18.0 + (-40.0) * (s - 1/6)$
Resolution	10.0–2.0 Å
No. of reflections	10,110
No. of atoms	1,665
	Root mean square deviations from ideality (target restraints in parentheses)
Distance restraints	
Bond distance	0.020 (0.020) Å
Angle distance	0.049 (0.035) Å
Planar 1–4 distance	0.056 (0.050) Å
Plane restraints	0.017 (0.020) Å
Chiral center restraints	0.201 (0.150) Å^3
Nonbonded restraints	
Single-torsion contact	0.191 (0.300) Å
Multiple-torsion contact	0.201 (0.300) Å
Possible ($X \ldots Y$) H bond	0.185 (0.300) Å
Conformational torsion angles	
Planar	2.8 (3.0)°
Staggered	17.6 (10.0)°
Orthonormal	17.2 (20.0)°
B_{iso} restraints	
Main-chain bond	1.5 (1.5) Å^2
Main-chain angle	2.2 (2.0) Å^2
Side-chain bond	4.2 (3.0) Å^2
Side-chain angle	5.9 (4.0) Å^2
H bond	9.5 (15.0) Å^2

[a] *Definitions:*

$$R = \frac{\Sigma ||F_o| - |F_c||}{\Sigma |F_o|},$$

$s = \sin\theta/\lambda$, and σ_F represents weights used in refinement in lieu of the standard deviations (σ) of the structure amplitudes.

The protein is made up of two 99-residue polypeptide chains (Kinemage 2); monomer 1 is numbered 1–99 and monomer 2 is numbered 101–199. The diol inhibitor, U-75875, is numbered 201–206. Like the other protease-inhibitor complexes (Wlodawer et al., 1992), this inhibitor is also bound in an extended chain in a well-defined and extensive active site cleft. This 23-Å-long groove runs across the dimer interface on one side of the molecule such that the two "flaps," one from each monomer (as described previously: e.g., Jaskólski et al., 1991), are closed over part of the inhibitor, thus protecting the inhibitor from the bulk solvent. The catalytically essential Asp 25 and Asp 125 side chains are found near the scissile bond isostere, on the other side of the inhibitor, away from the flaps. These side chains interact directly with the OH groups of the scissile bond isostere.

The ϕ,ψ angles for the non-glycine residues of the protease main chain are clustered in the allowed regions of the Ramachandran plot. The main-chain torsion angles of the inhibitor (listed in Table 2) correspond to an extended conformation, except for the torsion angles be-

Table 2. *Torsion angles for inhibitor U-75875*[a]

Residue	Backbone dihedral angles (degrees)		
	ϕ	ψ	ω
201	−170	11	177
202	−127	77	−174
203	−105	63	146
204	−68	126	177
205	−121	108	177
206	108	–	–

[a] Backbone dihedral angles for U-75875 in the enzyme–inhibitor complex are listed.

tween the P1 and P1′ positions, for which ω is 146° (i.e., torsion around the RCα-C(OH)-C(OH)-CαR′). The distance between OP1(203) and OP2(204) is 2.6 Å and the torsion around (HO)C-C(OH) is 50° (Kinemage 1). The two asymmetric centers (C * 203, 204) at the scissile bond both have the R configuration, unlike some previously reported inhibitors with one OH substituent at the scissile bond, such as hydroxyethylamine (Swain et al., 1990) and hydroxyethylene (Jaskólski et al., 1991), which have the S diastereomers preferentially bound by the enzyme. However, the R configuration reported here is equivalent to S in the inhibitors mentioned above, with the difference being the result of the nomenclature only. It should also be kept in mind that the R configuration results in better binding for some short inhibitors with a hydroxyethylene insert, with a different binding mode (Kröhn et al., 1991; Roberts et al., 1990).

Figure 2 represents the inhibitor in the $2|F_o| - |F_c|$ map. The electron density is well defined along all main-chain and side-chain atoms, except for a small fragment of the naphthalene side chain (residue 201, site P3). This residue has higher thermal vibrations than the rest of the inhibitor, and it projects slightly outside the enzyme boundary, without having any significant interactions with the protease.

The hydrogen-bonding interactions between the enzyme main-chain and side-chain atoms and the inhibitor are shown in Table 3 and Kinemages 5 and 6. The hydroxyl group in the P1 position (OP1) interacts with both the active site aspartates, whereas that at the P1′ position (OP2) only interacts with one. Wat 301, which has been observed in all other inhibitor complexes, makes contacts with the carbonyl oxygens of residues at the P2 and P1′ positions, and with the NH groups of residues Ile 50 and Ile 150 on the flaps, thus completely satisfying its hydrogen-bonding capacity. This water molecule possibly plays a crucial role in inducing the fit of the flaps over the inhibitor (Gustchina & Weber, 1990). The ND1 of histidine in the P2 position acts as a hydrogen-bond donor to the main-chain carbonyl oxygen of Gly 48 on the flap. The main-chain N of the residue at P3′ interacts with the main-chain O of Gly 148. These latter interactions possibly contribute further to stronger binding of the flaps over the inhibitor. The other interactions of the inhibitor involve the main-chain nitrogen atoms of residues at positions P1 and P2′ with main-chain carbonyl oxygens of glycine residues 27 and 127.

The two aspartic acid side chains in the active site show much more pronounced deviation from coplanarity than what was seen in other structures of inhibitor complexes of HIV-1 PR or, indeed, all other aspartic proteases. The root mean square (rms) deviation for the eight atoms of the aspartates (CB, CG, OD1, OD2 of Asp 25 and Asp 125) from the mean plane is 0.25 Å, with the shifts much larger for Asp 25 than Asp 125, reaching as much as 0.55 Å for the OD1 atom of the former (Fig. 5A; Kinemage 2).

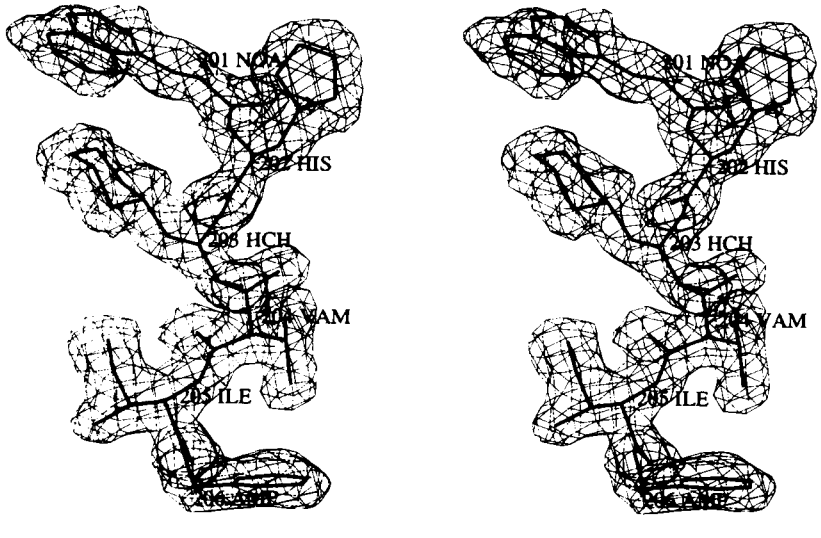

Fig. 2. Stereo view of the inhibitor in its $2|F_o| - |F_c|$ electron density map contoured at the 1.0σ level.

Table 3. *Close contacts at the active site*

201	Noa P3	Main chain	O...O (Wat 303)	3.2 Å
			O...N (Asp 29)	2.7 Å
202	His P2	Main chain	O...O (Wat 301)	2.6 Å
		Side chain	ND1...O (Gly 48)	2.9 Å
203	Cha P1	Main chain	N...O (Gly 27)	3.1 Å
		Hydroxyl	OP1...OD1 (Asp 125)	2.8 Å
		Hydroxyl	OP1...OD2 (Asp 125)	3.1 Å
		Hydroxyl	OP1...OD1 (Asp 25)	2.6 Å
204	Vam P1′	Main chain	O...O (Wat 301)	2.7 Å
		Hydroxyl	OP2...OD2 (Asp 25)	3.0 Å
205	Ile P2′	Main chain	N...O (Gly 127)	3.0 Å
			O...N (Asp 129)	2.9 Å
206	Amp P3′	Main chain	N...O (Gly 148)	3.0 Å
		Ring	N1...O (Wat 339)	2.7 Å

Symmetry of the enzyme and order/disorder of the inhibitor

Although the apoenzyme of HIV-1 PR is strictly symmetric under conditions of the investigation of its crystals reported previously (Lapatto et al., 1989; Navia et al., 1989; Wlodawer et al., 1989), most of the inhibitors studied so far are asymmetric, and thus the two molecules in the protein dimer of the resulting complex are nonequivalent. It was reported that this asymmetry is present even in the complexes of protease with quasisymmetric inhibitors (Erickson et al., 1990; Bone et al., 1991). The most telling indication of the nonequivalency of the two molecules is found in the ϕ and ψ torsion angles of residues 50 and 51. Wlodawer et al. (1992) reported that for nine structures of such complexes, ψ_{50} ranged from $-20°$ to $-45°$ and ϕ_{51} between $-65°$ and $-90°$ for molecule 1, whereas the corresponding values for molecule 2 were 130–140° and 90–115°. Thus, by this definition, molecules 1 and 2 can be distinguished even without reference to the directionality of the inhibitor. For the protease complexed with U-75875 the values for molecule 1 are $-20°$ and $-91°$, and for molecule 2, 129° and 97°, in good agreement with the previous data. Of course, in the orthorhombic space group to which these crystals belong, molecules 1 and 2 are distinct crystallographically and make different lattice contacts. As judged by the definition above, these two molecules are similar to those previously observed in the same space group for MVT-101 (Miller et al., 1989), JG-365 (Swain et al., 1990), and U-85548e (Jaskólski et al., 1991). Although this result might indicate strict correlation between crystal packing forces and orientation of the flaps, recent results from this and other laboratories have indicated that this is not the case, because in some other isomorphous structures the opposite arrangement of the two molecules was also observed (J.K.M. Rao, unpubl.; J. Erickson, pers. comm.).

A related phenomenon is the order/disorder of the inhibitor and its orientation with respect to the two protease molecules. A schematic representation of such interactions was presented as Figure 6 in Jaskólski et al. (1991), and the orientation of the inhibitor observed by us is identical to that reported there in terms of the N to C direction. The 2-Å electron density for the inhibitor U-75875 is very clear (Fig. 2), particularly for bulky end groups such as Noa, and shows beyond any doubt that only a single conformation is present. Previously, MVT-101 was considered to be present in only one orientation after 2.3-Å refinement (Miller et al., 1989), but that interpretation was changed after further refinement at 2.0 Å (Swain et al., 1992; M. Miller, pers. comm.). No indication of disorder was reported for the complex of JG-365 at 2.4 Å resolution (Swain et al., 1990), and for the 2.5-Å structure of U-85548e, it was estimated that the disorder, if present at all, does not exceed 30% (Jaskólski et al., 1991). On the other hand, pepstatin was found to be completely disordered in another crystal form at 2-Å resolution, although monomers 1 and 2 of the enzyme could also be distinguished in that case using the definitions above (Fitzgerald et al., 1990). Thus, although the picture seen in the electron-density maps of the current complex is very clear and unambiguously correlates the two molecules with the direction of the ordered inhibitor, we cannot conclude that the crystallization process has led to the unique selection of oriented molecules, and that similar selection will be present in all other structures.

Oxidation of Cys 67

Early in the refinement process electron-density maps showed unexplained densities adjacent to the sulfur atoms of Cys 67 in both subunits, whereas no equivalent densities were present adjacent to Cys 95 (Fig. 3). The distance from the center of that density to the sulfur was approximately 1.6–1.7 Å, much shorter than would be allowed for a water molecule. After about 50 water molecules were included in the model, these densities became the highest in the $|F_o| - |F_c|$ maps and could not be disregarded. Ultimately, we modeled these densities by water molecules but applied distance restraints to attach them to the respective sulfurs (see Kinemage 4). When this procedure was followed, no further unexplained electron density was noticed around the two cysteines, and the maps were in complete agreement with the model. Because only one oxygen appears to be attached to each cysteine, we must conclude tentatively that the chemical species present is sulfenic acid. Although the presence of such covalent modification of free cysteines has been rather seldom mentioned in the literature, it is not unprecedented. For example, Wilke et al. (1991) reported that catalytic sulfur of an S195C mutant of trypsin was modified in a similar fashion, leading to considerable changes of the active site of that enzyme.

It is not clear whether this modification is due to chemical processes that took place before or after crystalliza-

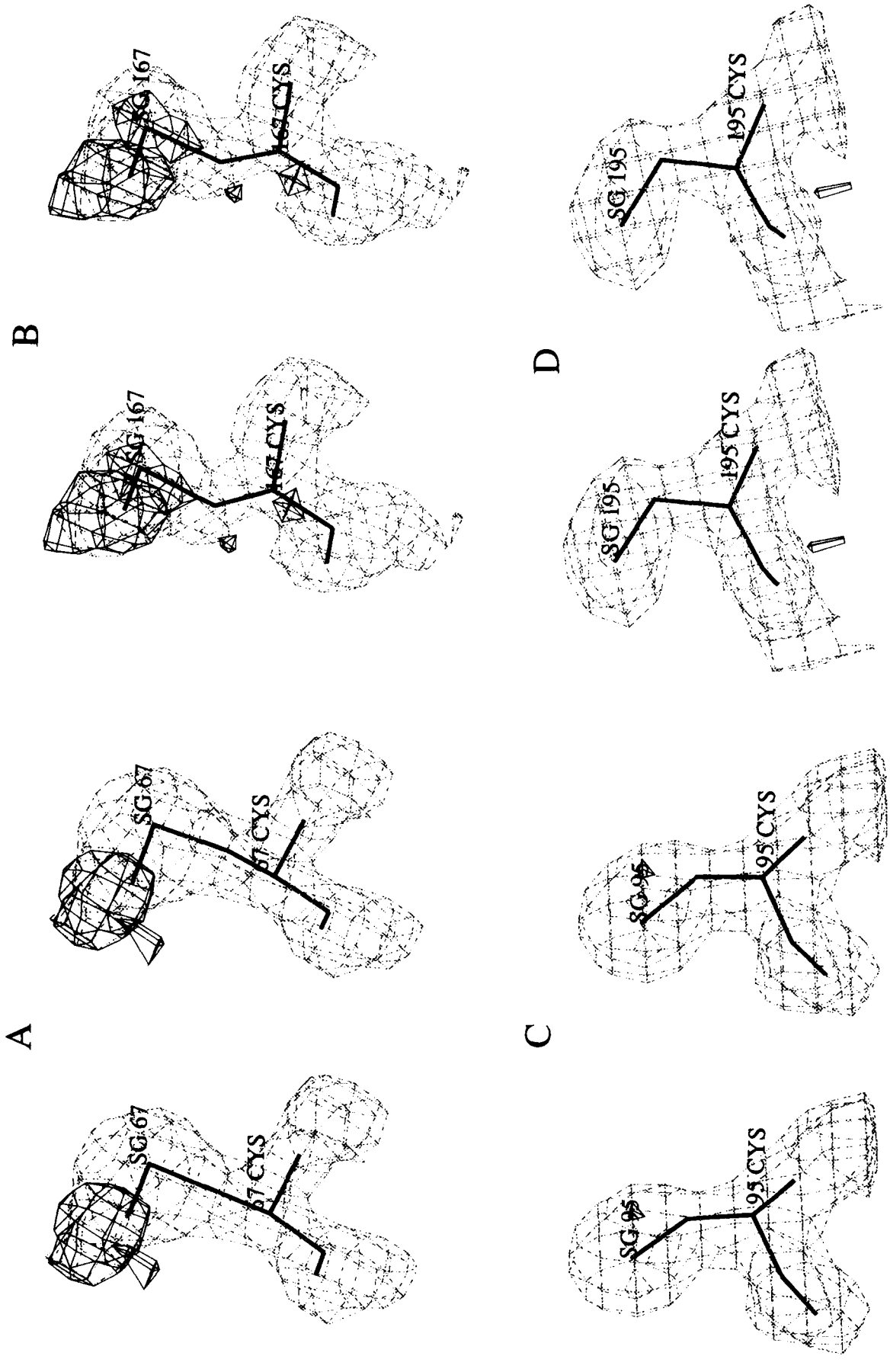

Fig. 3. Stereo views of electron-density maps around **(A)** Cys 67, **(B)** Cys 167, **(C)** Cys 95, and **(D)** Cys 195. A and B clearly show the extra densities in the $|F_o| - |F_c|$ density maps, contoured at the 3σ level. The oxygen atoms bound to SG were not used in phase calculations.

tion of the enzyme, or whether it is a radiation-induced phenomenon that happened during data collection. It does appear, however, that the oxidation was essentially complete, since the temperature factors of the oxygen are similar to those of the adjacent sulfurs. Because both Cys 67 and 95 are quite exposed, why only Cys 67 in each molecule is affected, in this fashion, remains a mystery. Perhaps, because Cys 95 of either monomer is involved in a crucial secondary structure element, namely the intermolecular β-sheet, alterations of these residues would be harder to achieve. The presence of two molecules of the protein in the crystallographic asymmetric unit makes the analysis of this phenomenon quite certain, since peaks due to any crystallographic artifacts would not be expected to affect both molecules equally. It is interesting to speculate whether oxidation of Cys 67 might lead to inactivation of the enzyme, although this residue is not close to the active site, and also if similar phenomena were indeed present in other crystallographic investigations, but were somehow overlooked due to the lack of internal check, which was available here.

Comparison of crystal structure of U-75875 with the previously published model

Correspondences between the modeled inhibitor and its crystal structure

A visual examination of the previously published model (Thaisrivongs et al., 1991) and the present crystal structure (Fig. 4A) reveals a number of similarities, the

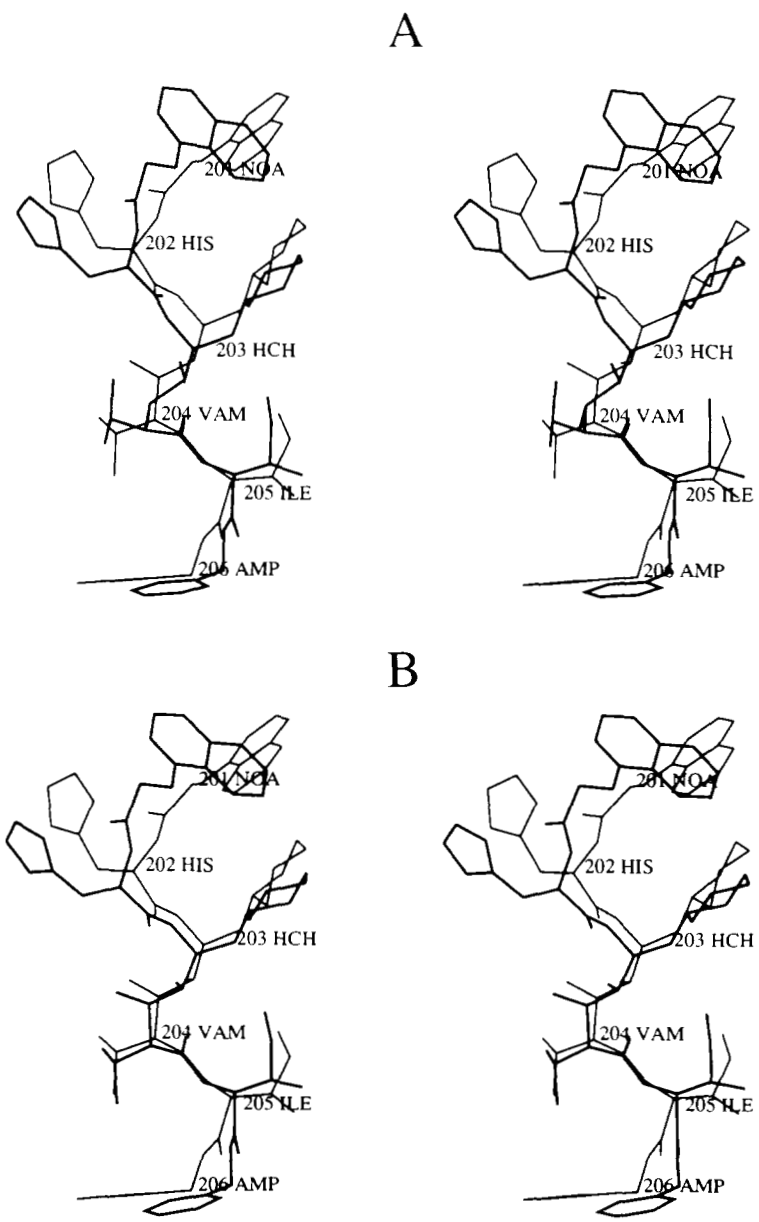

Fig. 4. Stereo views of the superpositions of the modeled inhibitor (darker lines) and the X-ray crystal structure (thin lines). **A**: Original model. **B**: Model reconstructed as described in the text. The reconstructed model was energy minimized in the active site of the MVT-101 version of the protein to a conformation 9.6 kcal lower in energy than was the original model.

most immediately noticeable being the close matching of the respective backbones. Hydrogen bonding between inhibitor backbone atoms and the enzyme is identical for the model and the crystal structure, except for the contact between the P1 main-chain N and the O of Gly 27, which was missed in the model. With the exception of the Noa placement, all side chains occupy roughly the same positions in the model and the crystal structure; the modeled side chains are within the variation seen in the crystal structures of other peptidic HIV-1 protease inhibitors (Tomasselli et al., 1991). Table 4 provides a residue-by-residue comparison of the two structures. Overall, the model correspondence with the crystal structure has an rms difference between corresponding atom positions of 1.9 Å. Most of the deviation occurred in the end groups Noa (1-naphthoxyacetyl) and Amp (2-pyridylmethylamine); these are the two regions found in the modeling to be able to accommodate multiple orientations of the groups and observed to have the highest mobility in the crystal structure. If these two groups are removed from the comparison, the rms correspondence becomes 1.3 Å.

When a model such as this one is used for synthetic optimization of side chains, the predicted overall conformation of the inhibitor is less important than is the identification of contacts between inhibitor side chains and the enzyme. For example, the correspondence between the His residues of the two structures is 1.6 Å rms. In Figure 4A this appears as a translation, plus a slight rotation, of the crystal structure relative to the model. However, when one then examines the enzyme atoms in contact with the His side chain, they too follow the same relocation. Just as the His side chain is shifted in the model relative to the inhibitor crystal structure, the enzyme pocket, into which the His fits also, shifts correspondingly from the MVT-101 version of the enzyme's structure (based on which the modeling studies were performed) to the current version. This observation holds for all six residues, and is detailed in Table 4. This synthetically relevant contact distance comparison of the model and the crystal structure is consistently better than the absolute comparison (1.3 Å rms vs. 1.9 Å rms, overall), and reduces to 0.8 Å rms for the P2 to P2′ residues that have been observed to contribute most to inhibitor binding.

Table 4. *RMS differences between the U-75875 model and crystal structure*

	Atom position[a]		Contact distance[b]	
Residue	rms	*n*	rms	*n*
Noa	2.3	14	1.0	85
His	1.6	10	0.7	115
Cha	1.1	11	0.4	86
Val	1.5	8	1.2	68
Ile	0.8	8	0.7	85
Amp	3.1	8	2.8	61
Total	1.9	59	1.3	500 (all residues)
Total	1.3	37	0.8	354 (excluding Noa, Amp)

[a] The rms column indicates the root mean square distance, in Å, between pairs of corresponding atoms in the model and the crystal structure; *n* is the number of such pairs in each residue, over which the calculation was done.

[b] For a given atom in a residue in the crystal structure, the distance to all protein atoms within 5 Å was determined and so were the corresponding distances in the model. The contact distance difference rms calculation employed the *differences* between those distances in the model and the crystal structure; *n* is the number of such distances that were found.

Differences between the modeled inhibitor and its crystal structure

The most visible differences between the two are in the end-group placements (Fig. 4A) and in the positioning of the second hydroxyl group in the diol portion of the insert (Fig. 5A). At the P3 end of the inhibitor, the model-building procedure found several placements of the Noa group that would maximize its hydrophobic interactions with the enzyme. In the placement shown in Figure 4A, for example, the naphthyl ring lies parallel to the lower strand of the enzyme's flap and in full contact with Gly 48 and Gly 49 of that strand. There are additional minor hydrophobic contacts with Phe 53, on the upper strand of the flap, and with Pro 81 on the opposite side of the cleft. In contrast, the naphthyl ring of the Noa group in the crystal structure makes relatively little contact with the protein. It lies perpendicular both to the flap and to the plane of the naphthyl group as seen in the model. Only two of its atoms are in contact with enzyme atoms; the rest of the ring system remains exposed. It is possible to explain this difference on the basis of the potential function used in the determination of intermolecular interaction energies during the model building, since this procedure rates favorably those conformations of the inhibitor that bury hydrophobic regions on both the inhibitor and the enzyme. However, this trend may not always be reflected in the actual binding of inhibitor to protein.

At the opposite end of the binding cleft, the difference between the two structures is more difficult to explain. In the model, the Amp group is high in the cleft, close to one of the flaps and one wall of the cleft. In the crystal structure, the Amp group is quite low, close to the floor and to the opposite wall of the cleft. In both structures there are very few interactions between the Amp ring and the protein. In the modeled structure this would suggest either that such contacts were not possible due to backbone hydrogen-bonding constraints, or that there was little energetic value to be found in making the contacts. Where minimum contacts with the enzyme exist for a chemical group, predictions obtained by modeling methods need not always yield observed structural results.

The modeling predicted an asymmetric positioning of the two hydroxyl groups relative to the catalytic aspar-

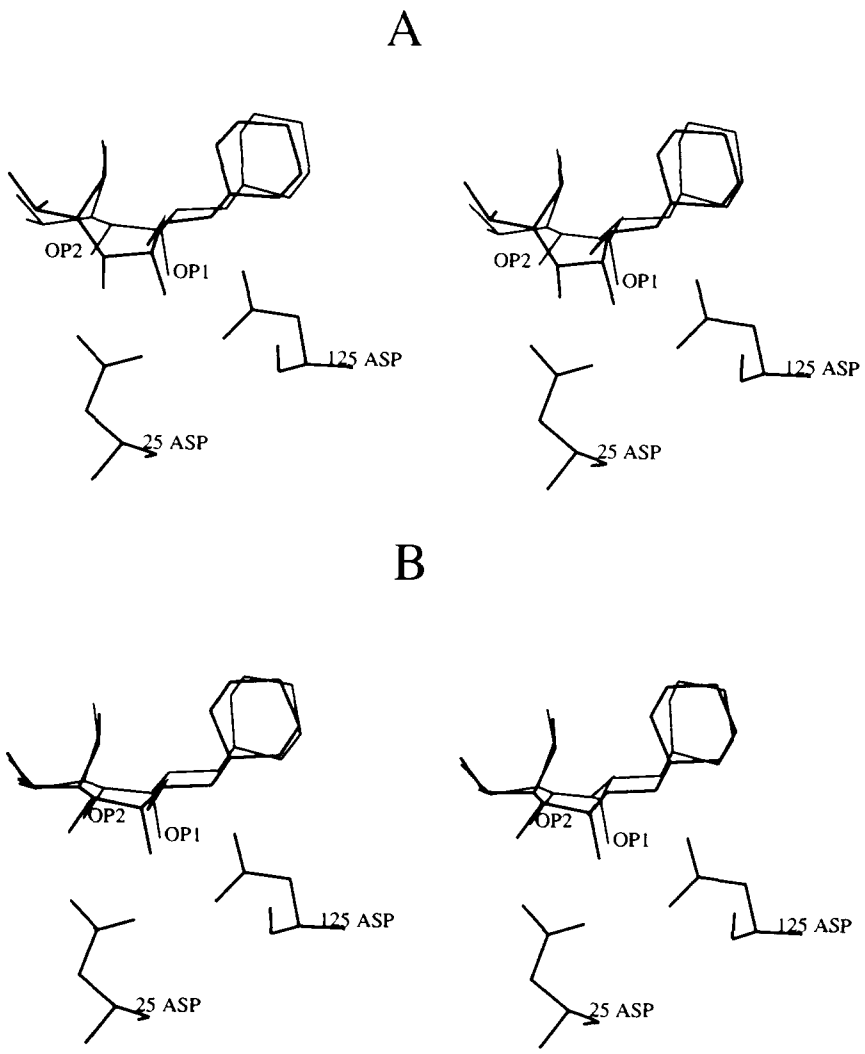

Fig. 5. Close-up stereo views of the positioning of the two hydroxyl groups of the modeled inhibitor (darker lines) and its X-ray crystal structure (thin lines) relative to the catalytic aspartates of the protease. **A**: Original model. **B**: Reconstructed model.

tates of the protease. In the model, OP1 was situated between the aspartates, similar to what has been seen in crystal structures of all monohydroxy inserts to date. The second hydroxyl, OP2, was located on the P1′ side of the aspartates. It was able to hydrogen-bond to only one of them, rather than to both, as did OP1. The crystal structure of U-75875 confirmed part of this picture. The two hydroxyl oxygens were indeed asymmetrically positioned (Kinemage 6), and OP1 was located generally as predicted. However, OP2 was almost 3 Å away from its predicted location and was located on the P1 side of the aspartates rather than the P1′ side. Figure 5A provides a close-up view of this region. Much of the relative displacement of the OP2s can be attributed to a 107° difference in the (HO)C-C(OH) dihedral angles (−57° in the model vs. +50° in the crystal structure). A difference of that magnitude places the two hydroxyl oxygens on opposite sides of Asp 25—both are able to form a hydrogen bond to Asp 25, but from different directions. Thus, it is clear that where two favorable possibilities based on geometrical constraints exist, modeling methods may occasionally choose the one that is not present in the crystal structure.

To identify the origin of this large discrepancy, we reconstructed the model of the dihydroxy-containing insert of U-75875 (see the Materials and methods section). Conversion of the CH_2-N insert of the MVT-101 crystal structure (Miller et al., 1989) to CH_2-CH_2, followed by substitution of the appropriate protons with OHs to provide the desired R,R stereochemistry, gave a dihedral at the (HO)C-C(OH) bond of 22° before minimization. OP2 remained on the same side of Asp 25 as is observed in the crystal structure of U-75875. Surprisingly, the next step of the model reconstruction, that of replacing the MVT-101 P1 and P1′ side chains with the Cha and Val of U-75875, followed by minimization of the entire P1–P1′ insert, provided a (HO)C-C(OH) dihedral of 28°, not the −57° value seen in the original model. Figure 5B illustrates the reconstructed model insert. This version matches the U-75875 crystal structure very closely. The

difference between the central dihedrals decreased from 107° (in the original model) to 22° (in the reconstructed model), and the separation of the OP2s was reduced from 3.0 Å to 0.7 Å. Importantly, the modeled OP2 hydrogen-bonded to Asp 25 from the P1 side, just as is seen in the U-75875 crystal structure.

Because of the apparent nonreproducibility of a critical portion of the model, we retraced in detail the steps taken to build the original model and identified the source of the error. OP2 had originally been placed in the S configuration, rather than R. The minimization step forced the central dihedral to rotate to bring OP2 into hydrogen-bonding distance from Asp 25. This placed OP1 and OP2 on opposite sides of Asp 25. The misassigned OP2 configuration had then been corrected by switching the H and OH positions on OP2, followed by reminimization, rather than restarting from scratch. The minimization of the corrected R-OP2 took it into a local minimum and was apparently unable to pull it back to the P1 side of Asp 25.

Having explained the discrepancy between the dihydroxy portion of the original model and that of the crystal structure, we continued with the model reconstruction by "growing" the P2–P3 and P2′–P3′ groups, following the same procedure used to construct the original model. No information from the current HIV-1 PR/U-75875 crystal structure was used. The resulting corrected model is compared with the crystal structure in Figure 4B. The P1 and P1′ residues on either side of the corrected dihydroxy match the crystal structure much more closely than did the original model. The rms difference in the Cha atoms has been reduced from 1.1 Å (see Table 4) to 0.9 Å, and the contact distance to enzyme atoms has improved slightly, from 0.4 Å rms to 0.3 Å rms. For the Val atoms at P1′, the match with the crystal structure improves considerably, from 1.5 Å rms to 0.4 Å rms, with the contact distance match changing from 1.2 Å to 0.7 Å rms. Beyond the P1 and P1′ positions, however, very little change occurs. Those residues are apparently too far away from the dihydroxy moiety to be affected by changes in its stereochemistry. The one hydrogen bond that was missed in the original model, that between the P1 N and Gly 27 O, is present in the corrected model.

Discussion

The number of structures of different complexes of HIV-1 PR with inhibitors is becoming quite large (summarized in Wlodawer et al., 1992), but nevertheless every new structure seems to bring new and unexpected findings. The inhibitor studied here, U-75875, is unique among those reported to date in having a diol insert replacing the usual scissile bond. We found only one of the hydroxyl groups to interact with both active site aspartates of the enzyme, whereas the role of oxygen OP2 may be to twist the carboxylates out of plane, in a manner not previously reported for HIV-1 PR, and not seen to that extent in any other aspartic proteases.

Among the crystal structures of native cellular aspartic proteases, the largest deviation from planarity of the two aspartates in the active site was reported for pepsin (Abad-Zapatero et al., 1990), where nonplanarity of the two groups, defined in the previous section, was 0.11 Å, with the maximum deviation of 0.22 Å for OD2 of Asp 32. Larger deviations from planarity have been observed in inhibitor complexes of aspartic proteases (Šali et al., 1989), but details have not been given. The nonplanarity for the native HIV-1 PR was 0.14 Å, for the complex with MVT-101 (Miller et al., 1989) was 0.09 Å, for U-85548e (Jaskólski et al., 1991) was 0.09 Å, and for the JG-365 (Swain et al., 1990) was 0.22 Å. The last value approaches 0.25 Å for the complex reported here, but the maximum deviation was much lower, 0.36 Å for OD1 of Asp 25, reflecting the fact that nonplanarity was mostly due to the parallel shift in the two planes, rather than a twist in one of them. In the case of U-75875, however, the nonplanarity is most probably caused by either steric hindrance between OP2 of the inhibitor and OD1 of Asp 25, or possibly by other repulsive forces. It is not clear if that effect plays a part in enhancing the binding of the inhibitor to the enzyme, or if its contribution to the binding energy is negligible.

Another unexpected phenomenon observed in this structure, not directly related to the presence of the inhibitor, is oxidation of Cys 67 in both molecules of the enzyme. This phenomenon may be caused by the irradiation of the crystal in the X-ray beam, but its exact source is not well understood at this time. It may be worthwhile, however, to reanalyze other high-resolution structures of proteins containing free cysteines to see if similar behavior went unrecognized in the past.

The U-75875 model was originally developed to provide a structural basis for synthetic modifications of the inhibitor side chains and to lead to a better understanding of the interactions of the central diol with the enzyme. It was able to explain some of the structure–activity relationship (SAR) related to the U-75875 series. This included (1) the relative insensitivity of inhibitor binding to the type of end groups chosen for the P3 and P3′ positions (since they projected slightly outside the binding cleft and could find several orientations, according to the model), and (2) a preference for nonequivalent P1 and P1′ side chains, on either side of the central diol (since the model predicted an asymmetric positioning of the diol relative to the catalytic aspartates of the enzyme).

Overall, the model appears to have matched the crystal structure of the inhibitor quite well, especially in the central regions that contribute the strongest binding interactions. None of the SAR explanations that were based on the model-built inhibitor need to be revised in light of the inhibitor's crystal structure. Although the P3 and P3′ end groups did not match the positions seen in the crys-

tal structure as closely as was hoped, it is also the case that there are numerous possible binding locations in these regions. Because of this, and the solvent exposure of these regions, it may be that the molecular mechanics techniques used in the modeling were unable to pick up the subtleties and nuances of interaction that appear in the crystal structure.

The construction of the Cha-Vam diol insert was considered to be the most important part of the model-building process, for two reasons. First, although diol inserts were known at that time to contribute significantly to inhibitor binding (Thaisrivongs et al., 1991), there were no crystal structures of such inserts to define how the hydroxy groups interacted with the enzyme. It would have been perfectly reasonable to assume a symmetric positioning of the diols with respect to the aspartates, for example. And second, the methodology employed to grow the P3–P2 and P2′–P3′ residues would rely, we felt, on an accurate positioning of the P1–P1′ insert. Our assumption was that a poorly modeled insert would lead, by extension, to a poorly modeled inhibitor overall.

This did not turn out to be the case. Although the corrected insert model improved the P1–P1′ match with the crystal structure considerably, there was relatively little effect on the remainder of the structure. It appears, therefore, that the overall quality of the model is less sensitive to the quality of the insert portion of the model than we had originally thought. By extension, it is reasonable to expect that X-ray crystal structures of dihydroxy-based inserts of varying stereochemistries, if they can be obtained, will show major differences within the P1–P1′ region, but little perturbation beyond that area. Further, it will be interesting to see if crystal structures of R,S-dihydroxy inserts exhibit the same type of flanking of the aspartates as has been inadvertently suggested by our modeling. This assumes, of course, that the stereochemical change does not eliminate HIV-1 PR inhibitory activity. In one case, for example, the change from R,R to R,S diol configuration in a pair of compounds related to U-75875 showed a 20-fold drop in binding affinity (Thaisrivongs et al., 1991).

Materials and methods

Synthesis of U-75875 and preparation of recombinant HIV-1 protease

Details regarding the synthesis of U-75875 and its diol insert were published in an earlier communication by Thaisrivongs et al. (1991). The HIV-1 PR was the recombinant enzyme expressed in *Escherichia coli*, and purified from inclusion bodies as described by Tomasselli et al. (1991). The protease was shown to be pure by the usual criteria of compositional and sequence analysis and by specific enzyme activity. Immediately following refolding of the enzyme, it was treated with a twofold molar excess (with respect to the HIV-1 PR dimer) of U-75875, which forms a tight one-to-one complex with the enzyme dimer (K_i = <1 nM) and blocks activity completely. The protease can be stored indefinitely as this complex with inhibitor and is unable to undergo autolysis.

Crystallization and data collection

Crystals of the protease complexed with the inhibitor, U-75875, were obtained by the hanging-drop vapor diffusion technique (Wlodawer & Hodgson, 1975). A solution of 20% ammonium sulfate with 0.1 M sodium acetate (v/v, pH 6.8) was used as precipitant. The protein complex was concentrated to ~10 mg/mL, after which the inhibitor was added at a molar ratio of 2:1 inhibitor to protein dimer and the mixture was incubated overnight at room temperature prior to setting up crystallizations. The 6-μL drops consisted of 50% of this protein–inhibitor complex and 50% of the precipitant. Crystals appeared at 20 °C within 24 h and grew over 5 days to their final size. The space group is $P2_12_12_1$ with unit cell parameters $a = 52.0$ Å, $b = 58.6$ Å, $c = 61.9$ Å, containing one protease dimer and one molecule of the inhibitor per asymmetric unit. X-ray data were collected using a Siemens area detector mounted on a Rigaku RU-200 rotating anode source. A single crystal (approximately $1.0 \times 0.3 \times 0.2$ mm) was used. The data extended to 1.9 Å and were complete to 2.0 Å (10,110 reflections with $I > 1.5\sigma(I)$ from 10.0 Å to 2.0 Å were included in refinement).

Structure solution and refinement

Three other HIV-1 PR inhibitor complexes previously solved in this laboratory are isomorphous with the complex described here (Miller et al., 1989; Swain et al., 1990; Jaskólski et al., 1991). However, the starting model for the homodimer enzyme was taken from the as yet unpublished, refined model of a protease–inhibitor complex containing a Phe-Sta insert (J.K.M. Rao, unpubl.). That model in turn was originally based on the structure of Miller et al. (1989), refined further at 2 Å resolution. The very first $(|F_o| - |F_c|)\alpha_c$ map using the phases calculated from the protein alone clearly showed the inhibitor density in the active site cleft. Since the chemical structure of the inhibitor was known (Thaisrivongs et al., 1991), it was possible to orient the inhibitor correctly in the first instance to fit the electron density in the difference map. Thus, after three cycles of initial refinement using X-PLOR (Brünger et al., 1987), the inhibitor was modeled into the electron density by using the modeled structure, and the ϕ, ψ, and ω angles were manually manipulated for the best fit. Each round consisted of a refinement of positional coordinates (250 cycles), simulated annealing refinement at 2,000 K with a slow-cooling pro-

tocol (timestep 0.0005 ps; Brünger et al., 1990), an overall or individual temperature factor refinement (20 cycles) and further fitting of the electron density maps using FRODO (Jones, 1985). The structure of the protease-inhibitor complex was then refined using the restrained least squares program PROFFT (Hendrickson, 1985; Finzel, 1987; Sheriff, 1987) to an R factor of 0.197. Solvent molecules were included using a program written by J.K.M. Rao (unpubl.), and later manually checked on graphics using FRODO (Jones, 1985). The final model is characterized by an R factor of 0.169, and by an rms deviation between ideal and observed bond distances of 0.02 Å (Table 1).

Coordinates have been deposited in the Protein Data Bank, file number 1HIV.

Development of the inhibitor model

The procedure used to develop models of U-75875 bound to HIV-1 protease has already been described in detail (Thaisrivongs et al., 1991). However, because portions of that procedure are relevant to the analysis presented in the current work, they will be summarized here. The starting point for the modeling was the X-ray crystal structure of the inhibitor MVT-101 (Ac-Thr-Ile-NleΨ[CH_2NH]Nle-Gln-Arg-NH_2) complexed to HIV-1 protease (Miller et al., 1989). The central CH_2-N of MVT-101 was converted first to CH_2-CH_2, and then to CH(OH)-CH(OH) in the R,R configuration, using standard molecular graphics methods. The P1 and P1′ side chains of MVT-101 were then mutated to the cyclohexyl (Cha) and isopropyl (valine) of U-75875, to provide the "insert" construct. The remainder of MVT-101 was discarded. The U-75875 insert was then subjected to energy minimization in the MVT-101 version of the HIV-1 PR crystal structure. Energy minimization was accomplished with the AMBER forcefield (Weiner et al., 1984) as implemented in MacroModel/BatchMin version 2.5 (Mohamadi et al., 1990), using the PRCG minimizer and an rms gradient of 0.1 kcal/Å as the convergence criterion. The Noa-His and Ile-Amp groups in the P3-P2 and P2′-P3′ locations, respectively, were then attached by the GROW program (Moon & Howe, 1991), which employs a conformational search procedure to select from a library of amino acid conformations those that form the strongest interactions with the protein. Conformations are scored according to a molecular mechanics-based scoring function, which includes van der Waals and electrostatic nonbonded contributions, plus desolvation and ligand internal strain penalties.

The resulting conformational models of U-75875 were then minimized in the MVT-101 version of the HIV-1 PR crystal structure. Although the models demonstrated a number of alternative binding modes, especially in the P3 and P3′ regions, only the highest scoring model was used in the comparison described herein. Since the crystal structures of the U-75875 and MVT-101 inhibitor complexes are in the same coordinate frame of reference, the comparison could be made by direct overlays.

Acknowledgments

We thank Dr. Ronald Rubin for help in data collection and Dr. Maria Miller for sharing with us her unpublished, highly refined structure of the MVT-101 complex. We are indebted to Dr. Stephen Freer for his computer program, which was used to create an inhibitor standard group for PROFFT refinement. This research was sponsored in part by the National Cancer Institute, DHHS, under contract no. NO1-CO-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. The US Government and its agents and contractors retain a nonexclusive royalty-free license in copyright covering this article.

References

Abad-Zapatero, C., Rydel, T.J., & Erickson, J. (1990). Revised 2.3 Å structure of porcine pepsin: Evidence for a flexible subdomain. *Proteins Struct. Funct. Genet. 8*, 62-81.

Ashorn, P., McQuade, T.J., Thaisrivongs, S., Tomasselli, A.G., Tarpley, W.G., & Moss, B. (1990). An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. *Proc. Natl. Acad. Sci. USA 87*, 7472-7476.

Bone, R., Vacca, J.P., Anderson, P.S., & Holloway, M.K. (1991). X-ray crystal structure of the HIV-1 protease complex with L-700,417, and inhibitor with pseudo C_2 symmetry. *J. Am. Chem. Soc. 113*, 9383-9385.

Brünger, A.T., Krukowski, A., & Erickson, J.W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A46*, 585-593.

Brünger, A.T., Kuriyan, J., & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science 235*, 458-460.

Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A., & Knigge, M. (1990). Design activity, and 2.8 Å crystal structure of a C_2 symmetric inhibitor complexed to HIV-1 protease. *Science 249*, 527-533.

Finzel, B.C. (1987). Incorporation of fast Fourier transforms to speed restrained least-squares refinement of protein structures. *J. Appl. Crystallogr. 20*, 53-55.

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). Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. *J. Biol. Chem. 265*, 14209-14219.

Gustchina, A. & Weber, I.T. (1990). Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: The role of the flap. *FEBS Lett. 269*, 269-272.

Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol. 115*, 252-270.

Jaskólski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B.H., & Wlodawer, A. (1991). Structure at 2.5 Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry 46*, 1600-1609.

Jones, A.T. (1985). Interactive computer graphics: FRODO. *Methods Enzymol. 115*, 157-171.

Kröhn, A., Redshaw, S., Ritchie, J.C., Graves, B.J., & Hatada, M.H. (1991). Novel binding mode of highly potent HIV-1 proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J. Med. Chem. 34*, 3340-3342.

Lapatto, R., Blundell, T.L., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danely, D.E., Geoghegan, K.F., Hawrylik, S.J., Lees, S.E., Scheld, K.G., & Hobart, P.M. (1989). X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature 342*, 299–302.

Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., & Wlodawer, A. (1989). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science 246*, 1149–1152.

Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., & Still, W.C. (1990). MacroModel – An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comp. Chem. 11*, 440–467.

Moon, J.B. & Howe, W.J. (1991). Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *Proteins Struct. Funct. Genet. 11*, 314–328.

Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., & Springer, J.P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature 337*, 615–620.

Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., Kröhn, A., Lambert, R.W., Merrett, J.H., Mills, J.S., Parkes, K.E.B., Redshaw, S., Ritchie, A.J., Taylor, D.L., Thomas, G.J., & Machin, P.J. (1990). Rational design of peptide-based HIV proteinase inhibitors. *Science 248*, 358–361.

Săli, A., Veerapandian, B., Cooper, J.B., Foundling, S.I., Hoover, D.J., and Blundell, T.L. (1989). High resolution x-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: The analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J. 8*, 2179–2188.

Sheriff, S. (1987). Addition of symmetry-related restraints to PROTIN and PROLSQ. *J. Appl. Crystallogr. 20*, 55–57.

Swain, A.L., Gustchina, A., & Wlodawer, A. (1992). Comparison of three inhibitor complexes of human immunodeficiency virus protease. In *Structure and Function of the Aspartic Proteinases: Genetics, Structure and Mechanisms* (Dunn, B., Ed.), pp. 433–441. Plenum Press, New York.

Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H., & Wlodawer, A. (1990). X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. USA 87*, 8805–8809.

Thaisrivongs, S., Tomasselli, A.G., Moon, J.B., Hui, J., McQuade, T.J., Turner, S.R., Strohbach, J.W., Howe, W.J., Tarpley, W.G., & Heinrikson, R.L. (1991). Inhibitors of the protease from human immunodeficiency virus: Design and modeling of a compound containing a dihydroxyethylene isostere insert with high binding affinity and effective antiviral activity. *J. Med. Chem. 34*, 2344–2356.

Tomasselli, A.G., Howe, W.J., Sawyer, T.K., Wlodawer, A., & Heinrikson, R.L. (1991). The complexities of AIDS: An assessment of the HIV protease as a therapeutic target. *Chim. Oggi 9*, 6–27.

Weiner, S.J., Kollman, P.A., Case, D.A., Chandra Singh, U., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P.A. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc. 106*, 765–784.

Wilke, M.J., Higaki, J.N., Craik, C.S., & Fletterick, R.J. (1991). Crystal structure of rat trypsin-S195C at −150 °C: analysis of low activity of recombinant and semisynthetic thiol proteases. *J. Mol. Biol. 219*, 511–523.

Wlodawer, A. & Hodgson, K.O. (1975). Crystallization and crystal data of monellin. *Proc. Natl. Acad. Sci. USA 72*, 398–399.

Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., & Kent, S.B.H. (1989). Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science 245*, 616–621.

Wlodawer, A., Swain, A.L., & Gustchina, A. (1992). Comparison of crystal structures of inhibitor complexes of the human immunodeficiency virus protease. In *Molecular Aspects of Chemotherapy* (Shugar, D., Rode, W., & Borowski, E., Eds.), pp. 173–186. PWN/Springer-Verlag, Warsaw, Berlin.