Domain Flexibility in Retroviral Proteases: Structural Implications for Drug Resistant Mutations[†,‡]

Robert B. Rose,[§] Charles S. Craik,[§,||] and Robert M. Stroud*[,§,||]

Department of Biochemistry and Biophysics, University of California in San Francisco, San Francisco, California 94143-0448

Received July 2, 1997; Revised Manuscript Received November 14, 1997

ABSTRACT: Rigid body rotation of five domains and movements within their interfacial joints provide a rational context for understanding why HIV protease mutations that arise in drug resistant strains are often spatially removed from the drug or substrate binding sites. Domain motions associated with substrate binding in the retroviral HIV-1 and SIV proteases are identified and characterized. These motions are in addition to closure of the flaps and result from rotations of ∼6−7° at primarily hydrophobic interfaces. A crystal structure of unliganded SIV protease (incorporating the point mutation Ser 4 His to stabilize the protease against autolysis) was determined to 2.0 Å resolution in a new space group, $P3_221$. The structure is in the most "open" conformation of any retroviral protease so far examined, with six residues of the flaps disordered. Comparison of this and unliganded HIV structures, with their respective liganded structures by difference distance matrixes identifies five domains of the protease dimer that move as rigid bodies against one another: one terminal domain encompassing the N- and C-terminal β sheet of the dimer, two core domains containing the catalytic aspartic acids, and two flap domains. The two core domains rotate toward each other on substrate binding, reshaping the binding pocket. We therefore show that, for enzymes, mutations at interdomain interfaces that favor the unliganded form of the target active site will increase the off-rate of the inhibitor, allowing the substrate greater access for catalysis. This offers a mechanism of resistance to competitive inhibitors, especially when the forward enzymatic reaction rate exceeds the rate of substrate dissociation.

Drug resistant mutations in human immunodeficiency virus (HIV)[1] protease occur frequently in response to antiproteolytic antivirals in patients with HIV infection. Several of these mutations lie within the region of the protease that binds the inhibitor, and thus there is in principle a rationale for understanding their action. They may interfere directly with inhibitor binding. However, many other mutations, the majority, occur far from the active site and do not easily fit this model. We determined an apo-simian immunodeficiency virus (SIV) protease structure that is unique among HIV/SIV proteases in its degree of accessibility of the active site. At least this degree of openness is required for the enzyme to bind substrate. Through analysis of this structure, and comparison with the many liganded structures of the proteases, we identify five domains within the dimer that move as rigid bodies in accommodation to ligand binding, as an induced fit mechanism. In the context of these domain motions, almost all (all except one in our set) of the resistant mutations that are far from the active site fall on the boundaries between the moving surfaces of the domains. This therefore provides a plausible paradigm for understanding these mutations, in terms of their impact on the required domain dynamics of the protease. This model also suggests a rationale for design of inhibitors that would avoid this mechanism of interference with inhibitor binding, by targeting the open structure. Hence they would avoid selection of mutations that achieve resistance in this manner.

The first unliganded and liganded structures of HIV protease showed a significant conformational change associated with the binding of a peptidomimetic inhibitor (*1, 2*). The binding pocket of the liganded, "closed", structure forms a tunnel, preventing access by the intact gag and gag-pol polyprotein substrates of the virus without a necessary conformational change. The two catalytic aspartic acids are located at the "floor" of the binding pocket in the dimeric protease. Substrate binding is therefore a multistep process: first an "open" structure binds substrate, followed by a conformational change to the closed structure (*3*). Characterization of the nature of this conformational change has focused on two flexible β-strands, the "flaps", which form the "ceiling" of the binding pocket. In the unliganded HIV-1 protease crystal structures, the flaps are in high thermal motion and are shifted by 7 Å relative to the closed conformation (*2, 4−6*). In an unliganded structure of Rous sarcoma virus protease, nine residues at the tips of the flaps

[†] This work was supported by NIH Grant GM 39552 (RMS), NIH Grant CA-09215, 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 code: 1az5.

* Correspondence should be addressed to Dr. Robert M. Stroud, Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448. Phone: (415) 476-4224. Fax: (415) 476-1902. E-mail: stroud@msg.ucsf.edu

[§] Graduate Group in Biophysics.

[||] Department of Pharmaceutical Chemistry.

[1] Abbreviations: EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; HIV, human immunodeficiency virus; rms, root mean square; rmsd, root mean square deviation; SIV, simian immunodeficiency virus; amino acids of proteins are designated by the three-letter code; amino acids of peptide ligands are designated by the one-letter code.

S0006-2960(97)01607-3 CCC: $15.00
Published on Web 02/06/1998

are disordered (*7*). Rapid flap motions have been detected by NMR (*8*) and by fluorescence changes (*9*). Larger motions of the flaps have been modeled by molecular mechanics (*10*−*12*), including motions of residues that border the flaps, described as the "cantilever" (*10, 11*).

Conformational changes in addition to flap closure were also noted in the first HIV-1 protease structures. A "hinge motion" between monomers of the protease homodimer on binding of substrate narrows the binding pocket (*1*). We have determined the structure of an unliganded SIV protease which is more "open" than the unliganded HIV-1 protease structures. In this structure the tips of the flaps are disordered, as in the unliganded Rous sarcoma virus protease structure. By comparing this structure to liganded SIV protease structures, we have identified five rigid body domains of the protease dimer. We show that the conformational changes associated with ligand binding are best described in terms of the rigid body rotations of these domains.

The retroviral proteases are related evolutionarily to the pepsin-like monomeric aspartyl proteases, as indicated by the conserved catalytic Asp-Thr-Gly sequence and by overall structural homology (*13*−*15*). The monomeric proteases are bilobal instead of dimeric, with a pseudo-2-fold symmetry axis relating the two lobes. One catalytic aspartic acid is positioned on each lobe. Only one of the lobes has retained a flap, which also closes over a substrate when it binds (*16*). Comparisons of liganded and unliganded structures have shown rigid body rotations of the two lobes (*17*). Conformational changes of the pepsin-like aspartyl proteases have been invoked to explain kinetic observations such as multiple substrate specificity (a feature of many aspartyl proteases), the increase in k_{cat} with substrate length, biphasic substrate binding, and the rate enhancement of k_{cat} in the presence of small noncleaved peptides (*18, 19*). The movements implied by the crystal structures have reinforced these explanations of the kinetic data (*20, 17*). The rigid body motions of the retroviral proteases identified here are similar to those of the monomeric family of aspartyl proteases. This conservation of the domain structure across the entire superfamily of aspartyl proteases attests to their functional importance.

MATERIALS AND METHODS

Recombinant SIV protease, with the point mutation Ser 4 His, was purified and frozen as described previously (*21*). After the protein was thawed for crystallization, 1 N sodium hydroxide was added to pH 10 to prevent precipitation. The protein solution was transferred to a Collodion membrane (Schleicher and Schuell) and concentrated for crystallization as described previously (*21*).

Unliganded SIV protease crystals were grown by hanging drop vapor diffusion. Crystals grew as pyramids to a size of 500 × 200 × 100 μm. These are readily distinguished from the orthorhombic plates observed for liganded complexes of SIV protease (*21, 22*). Diffraction data were collected on an R-axis II image plate detector using Cu Kα X-rays generated from a Rigaku 18 kW generator. The reflection intensities were integrated and scaled using the HKL suite of programs (Otwinowski, 1993). The rotational symmetry was determined with the program XPREP (copyright 1991, Siemens Analytical X-ray Instruments). The orientation of the molecule in the new space group was determined by molecular replacement using the program X-PLOR (*23*). X-PLOR was also used to refine the structure. Manual building was carried out using the program CHAIN (*24*).

Structure Comparisons

The unliganded SIV protease structure was compared with two liganded structures: one bound to the product peptide F-L-E-K (*22*) (PDB entry 1yti) (*22*) and the other bound to a peptidomimetic inhibitor (PDB entry 1siv) (*25*). A further comparison was made between an unliganded HIV-1 protease structure (PDB entry 3hvp) (*2*) and a liganded HIV-1 protease structure bound to the product peptide Ac-S-L-N-F (PDB entry 1yth) (*22*). The structures were superimposed by least squares minimization of differences between carbon α positions (*26*).

Rigid body domains were identified using difference distance matrixes between the α-carbon atoms of the liganded and unliganded structures (*27*). The difference distance matrix as calculated by X-PLOR was displayed graphically using Mathematica (*28*). Domains composed of spatially contacting residues were identified (*26*). Residues were assigned to the same domain if the interatomic distances varied less than 0.5 Å between the liganded and unliganded structures. Secondary structural elements were assigned completely to one of the domains. Outliers that deviated by more than 0.5 Å from the difference distance of the rest of the domain were assessed independently.

Rigid body motion is described in terms of screw axes which optimally superimpose one rigid body onto another (*29*). To identify the screw axes and describe the degree of rotation and translation between domains of the protease, the core domain was compared with the flap domain of the same monomer and with the terminal domain. First the core domain of one monomer of the liganded and unliganded proteins was superimposed. The rotation and translation required to superimpose each of the other (the flap, and terminal) domains defines the rotation axis. The relative rotation and translation values were determined. The program GEM calculates the screw axis resulting from a superposition (available through http://util.ucsf.edu/). In addition, 18 liganded HIV-1 protease structures deposited in the PDB (Table 1) were superimposed to determine whether the rigid body rotation of domains was a function of the ligand bound.

RESULTS

Unliganded SIV Protease Structure

Crystallographic statistics are reported in Table 2. X-ray data were reduced in the trigonal point group 32 with unit cell dimensions of $a = 45.4$ Å, $b = 45.4$ Å, $c = 87.9$ Å ($R_{symm} = 8.6\%$). The crystal density indicated that the crystal contained one protease monomer per asymmetric unit. The protease dimer axis was therefore coincident with the crystallographic 2-fold axis. The positioning of the molecule was determined by molecular replacement. The initial search model consisted of an SIV protease monomer as determined in our laboratory (PDB entry 1sam), excluding the ligand,

Table 1: Rms Deviation of HIV-1 Protease Structures from the Brookhaven Protein Data Bank

PDB code	ligand	space group	rmsd (Å) from 7hvp[a]	ref
3hvp	none	$P4_{1212}$	1.2[b]	(*2*)
3phv	none	$P4_{1212}$	1.3[b]	(*4*)
1hhp	none	$P4_{1212}$	1.2[b]	(*6*)
1aid	UCSF8[c]	$P2_{12121}$	0.58	(*37*)
1aaq	peptidomimetic	$P6_1$	0.70	(*57*)
1hbv	SB203238	$P6_1$	0.59	(*58*)
1hef	SKF 108738	$P6_{122}$	0.55	(*59*)
1hih	CGP 53820	$P2_{12121}$	0.62	(*60*)
1hiv	U75875	$P2_{12121}$	0.54	(*61*)
1hos	SB204144	$P6_1$	0.57	(*62*)
1hps	SB206343	$P6_1$	0.60	(*63*)
1hpv	VX-478	$P6_1$	0.61	(*64*)
1htg	GR137615	$P2_{12121}$	0.64	(*65*)
1hvi	A77003	$P2_{12121}$	0.53	(*66*)
1hvr	XK263	$P6_1$	0.59	(*67*)
1sbg	SB203386	$P6_1$	0.55	(*68*)
4hvp	MVT-101	$P2_{12121}$	0.37	(*1*)
4phv	L-700,417	$P2_{12121}$	0.63	(*69*)
5hvp	acetylpepstatin	$P2_{12121}$	0.69	(*70*)
7hvp	JG-365	$P2_{12121}$	0.0	(*71*)
8hvp	U-85548E	$P2_{12121}$	0.66	(*42*)
9hvp	A-74704	$P6_1$	0.64	(*72*)

[a] The rmsd is calculated by a least-squares superposition of the carbon α atoms of the structure onto 7hvp and calculating the root-mean-square deviation of the carbon α atoms from 7hvp. [b] Excluding residues 46−55 of the flaps of both monomers which are displaced in the unliganded structure. [c] Crystallized with a point mutant of HIV-1 protease, Gln 7 Lys.

Table 2: Data Collection and Refinement Statistics

highest resolution (Å)	2.0
no. of unique reflections[a]	7300
redundancy	5-fold
completeness of 2.06−2.00 Å resolution shell (%)	94
R_{symm}[b](%)	8.6
final *R* factor (7−2 Å) (%)	20.4
final *R*-free (7−2 Å) (%)	26.6
average *B* factor (Å^2)	20.1
rms deviation of bond lengths (Å)	0.014
rms deviation of bond angles (deg)	3.0
rms deviation of dihedral angles (deg)	27.6

[a] No σ cutoff applied. [b] $R_{symm} = \sum_i\sum_j|\langle I_i\rangle - I_{ij}|/\sum_i\sum_j I_{ij}$.

solvent, and the flap residues 48−52 (*21*). After Patterson correlation refinement of the top 100 solutions, only one solution gave a correlation coefficient greater than 0.1, equal to 0.12. The translation search identified the correct enantiamorph as $P3_221$. The *R* factor after the translational search was 48%.

Extra parameters were included based on use of the free *R* factor (*30*). After rigid body refinement, the structure was manually rebuilt. This model was further optimized by positional and individual *B* factor refinement. 20 water molecules were included on the basis of positive difference density peaks. Simulated annealing reduced the *R* factor by 1% and the *R* free by 1.4%. Manual rebuilding and further refinement led to a final *R* factor of 20.8% and *R* free of 26.6% for a resolution range of 7 to 2 Å.

Difference density in the protease active site indicates only a water molecule associated with the catalytic aspartic acids, as observed in the unliganded HIV-1 protease structure (*2*). This water molecule is 2.7 Å from the carboxyl oxygen of Asp 25 nearest the dimer 2-fold axis and 3.1 Å from the other oxygen.

Superimposing this structure onto liganded SIV protease structures shows that residues 46−47 and 54−55 are displaced by as much as 1.5 Å in the unliganded structure. The *B* factors of these residues increased to greater than 70 Å^2, from an average of 20 Å^2 for the protein. The gap between the ordered parts of the flap residues is 20 Å wide, large enough to allow a protein substrate to enter the site by displacing only disordered regions. The water molecule that is normally hydrogen bonded to the amide nitrogens of Ile 50 of both flaps when the flaps are closed is not visible, as was true of the unliganded HIV-1 protease structures. The side chains of Phe 53 and Ile 54 are also disordered and could not be modeled.

Domains of SIV Protease

Domains were defined as spatially connected regions with interatomic distances varying by less than 0.5 Å. Figure 1 shows the difference distance matrix between the unliganded SIV protease structure and the structure bound to the peptide product F-L-E-K. Residues 1−99 and 100−198 along either axis correspond to the two monomers of the protease dimer. The monomers of the unliganded structure are identical, making the choice of monomer to associate with the monomers of the liganded structure equivalent. Significant movement between the two monomers is represented on the matrix by large white patches at the intersection of the residues of the two monomers. Linear sequences of amino acids that constitute rigid subdomains appear as darkened areas along the diagonal of the matrix. Vertical lines delineate the rigid subdomains of the protease: residues 1−9, 10−32, 33−62, 63−85, and 86−99, labeled from **a** to **e** and from **a′** to **e′** respectively for each monomer. The white zone in the middle of the **c** and **c′** domains results from omission of the residues at the tips of the flaps which are disordered in the unliganded structure. These ten subdomains can be combined into five rigid domains. Noncontinuous sequences belonging to the same rigid domain appear as darkened blocks at their intersection on the matrix. The "terminal" domain encompasses subdomains **a**, **e**, **a′**, and **e′**. The two "core" domains consist of subdomains **b** and **d** or **b′** and **d′**. The two "flap" domains are composed of subdomain **c** or **c′**. Although the flap domain is closely associated with subdomain **d**, it moves relative to subdomain **b** and is therefore not included as part of the core domain. The relative movement between subdomains will be discussed further below.

Figure 2 maps the five rigid domains, along with the ten subdomains, onto the protease structure. The rigidity of these domains can be seen in Figure 3 by superimposing the domains of the liganded and unliganded structures. Each of these domains superimposes with an rms deviation for the carbon α atoms of 0.4 Å. By comparison, the rms deviation of the carbon α atoms for the complete dimer is 1.5 Å, while the rms deviation for each monomer is 0.7 Å.

To evaluate the significance of these deviations, we must compare them to the expected measurement error in the atomic coordinates. The error in the atomic coordinates for each domain can be estimated from the average *B* factor of the α carbon atoms of that domain (*31*). The overall average error in the coordinates expected for atoms within each domain of the two structures is 0.2 Å for each domain of

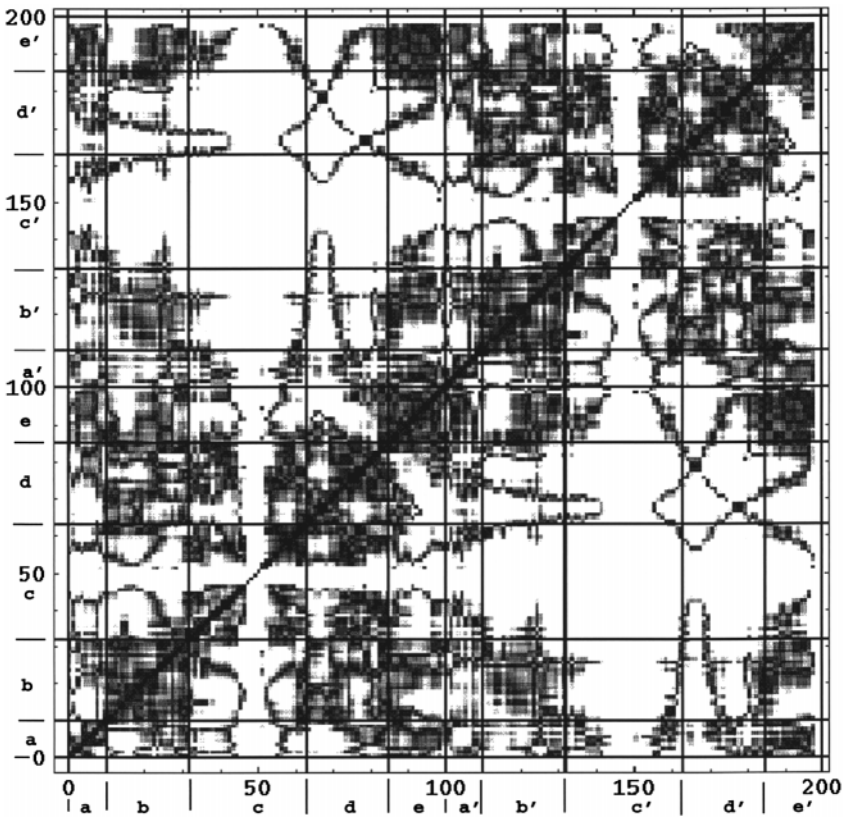


FIGURE 1: Difference distance matrix between the liganded and unliganded SIV protease structures. The matrix is shaded from black, indicating no difference distance, to white, indicating greater than 0.5 Å difference distance. Residue numbers are labeled along both axes: residues 1−99 belong to one monomer, residues 100−198 belong to the other monomer. Letters **a**−**e** and **a′**−**e′** indicate subdomains of the protease with difference distance less than 0.5 Å. (Difference distance matrix calculated by X-PLOR and plotted using Mathematica.)

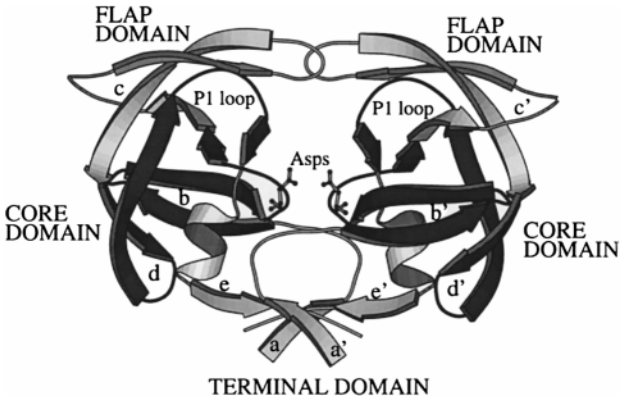


FIGURE 2: Ribbon diagram of SIV protease with the flaps in the "closed" conformation. Subdomains **a**−**e** and **a′**−**e′**, as identified in Figure 1, are labeled. The five rigid body domains are shaded: the terminal domain (residues 1−9 and 86−99 of both monomers), shaded gray; the core domains (residues 10−32 and 63−85 of a single monomer), shaded black; and the flaps (residues 33−62 of a single monomer), shaded gray (Figure drawn with Molscript (*73*)).

the unliganded structures, and 0.3, 0.3, and 0.4 Å for atoms of the terminal, core, and flap domains, respectively, for the product complex (PDB entry 1yth). If the error estimates are added in quadrature, the rms deviation expected per atom, due to random errors, is about 0.4 Å between individual atoms (*31*) This provides the rationale for use of 0.5 Å as the cutoff to define the domains by the difference distance matrix shown in Figure 1. The alignment of domains against one another is clearly much more accurate than the individual atomic deviations because there are many atoms, each with random deviations, that contribute to the determination of the molecular replacement vectors. Thus systematic domain positioning is accurate to $\sim 0.4/n^{1/2}$ where n is the number of atoms within the domain, i.e., less than 0.1 Å. The domain motions we derived correspond, after superposition of the core domains, to a difference in position of 1.1 Å at the center of mass of the terminal domain, 2.6 Å at the extremity of the domain, and 0.8 Å for the center of mass of the flap domain, 1.6 Å at the extremity. These values are significantly greater than the estimate of their error, ±0.1 Å from above. The actual rms deviations between the atoms of domains we identify of the unliganded and liganded structures after alignment, are the same as the error expected from the measurement of the atomic coordinates alone, consistent with the expectation for their belonging to the same structural domain. The protease domains are further described below.

The terminal domain consists of the four-stranded β-sheet, residues 1−3 and 95−99 of each monomer, the turn encompassing residues 4−9, and the helix, residues 86−94 of each monomer. The backbone hydrogen bonds of these residues are either solvated or satisfied within the domain. Arg 8 is the only residue in this domain in direct contact with a peptide ligand. In the absence of ligand, the side chain repositions to hydrogen bond with solvent.

The core domain consists of six primarily β-strand structures. The domain sequence, residues 10−32 and 63−85, is disrupted by the flap residues. The conserved Asp-Thr-Gly turn, including the catalytic aspartic acid, is situated at the interface of the core domains of the two monomers. The largest deviations between these domains (0.7 Å) occur in the loop 77−82 (labeled the "P1 loop" in Figure 2) which

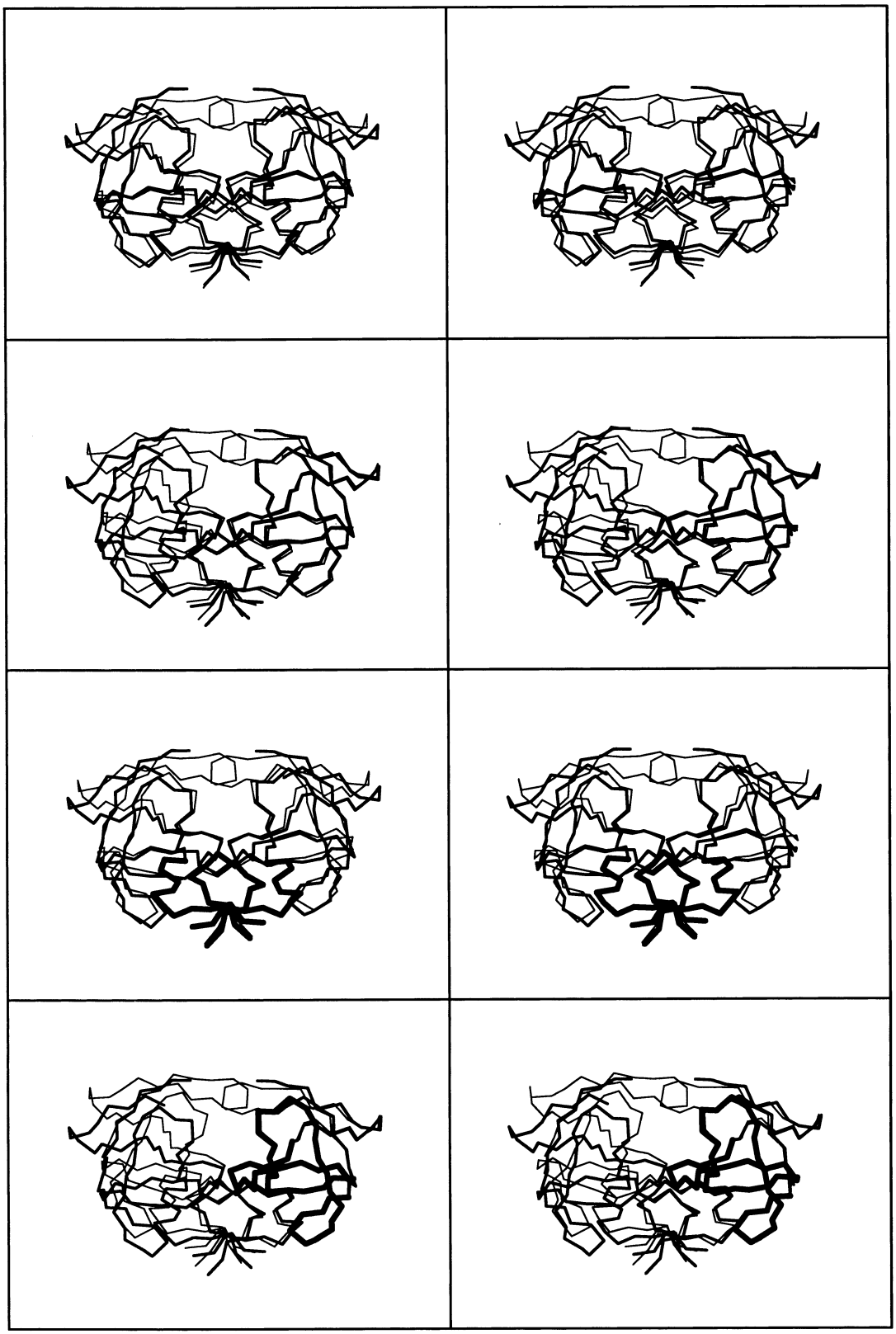


FIGURE 3: C_α traces of the unliganded (bold lines) and peptide product (F−L-E−K) liganded SIV protease structures after superposition of the C_α atoms. (a) Dimer superposition: residues 1−45 and 55−99 of both monomers of the dimer. (b) Monomer superposition: residues 1−45 and 55−99 of one monomer. (c) Terminal domain superposition: residues 1−9 and 86−99 of both monomers of the dimer. (d) Core domain superposition: residues 10−32 and 63−85 of one monomer. The flap residues 48−53 of the unliganded structure are not visible in the density (figure drawn with Molscript (*73*)).

both contacts the flap domain and forms part of the S1 subsite for the substrate.

The interface between the core and terminal domains is composed primarily of small hydrophobic residues. Figure 4a displays $2|F_o| - |F_c|$ density for part of this interface. The helix of the terminal domain packs against several β-strands of the core domain. The helical residues are completely hydrophobic except for the N-terminal residues

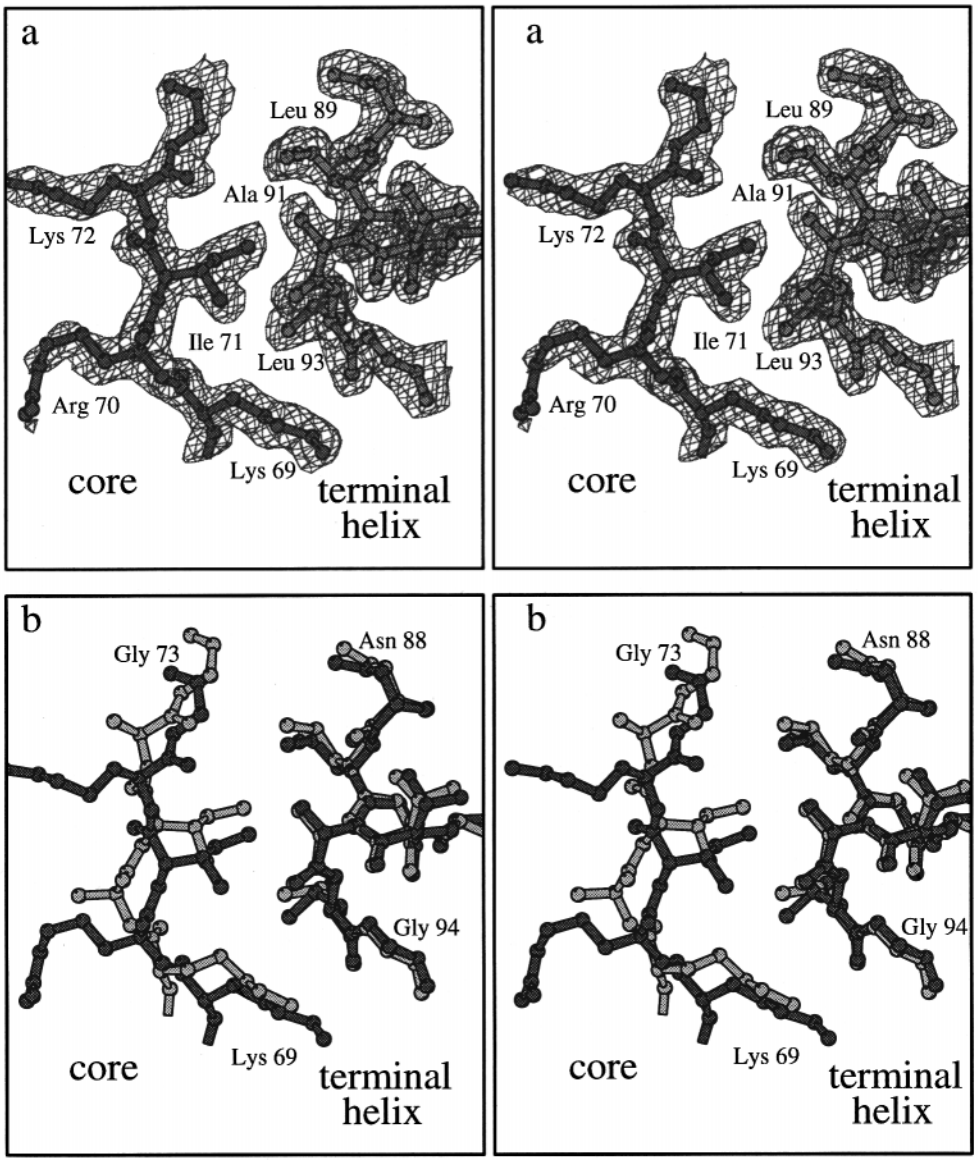


FIGURE 4: Residues at the interface between the terminal and core domains. (a) $2|F_o| - |F_c|$ density for the helix of the terminal domain (residues 89−94, colored medium gray) and a β-strand of the core domain (residues 69−73, colored black). (b) Reorientation of the residues at the interface as a result of the rigid body rotation. The terminal domains of the unliganded SIV protease, colored gray, and the protease bound to the product peptide F-L-E-K, colored black, were superimposed (figure drawn with Molscript (*73*)).

Arg 87 and Asn 88. One charge interaction is made between the carbonyl oxygen of Leu 93 and the side chain of Lys 69.

The flap domain has been assigned to residues that shift significantly when the core domains of the liganded and unliganded structures are superimposed. This includes a mostly solvent-exposed loop (residues 33−46) preceding the flap tips that are disordered in the unliganded structure (residues 48−52), plus a continuation of the β-strand of the flap (residues 57−62).

The choice of the product complex as the closed SIV protease structure in this comparison was insignificant. A similar comparison of the unliganded SIV protease in the open conformation with SIV protease bound to the peptidomimetic inhibitor SKF107457 (PDB entry 1siv) (*25*), or with the unliganded SIV protease structure in the closed conformation (PDB entry 1sip) (*32*), identified the same rigid body domains (data not shown).

Rigid Body Movement of Domains

Two axes of rotation describe the rigid body movement of the domains of SIV protease: one between the terminal and core domains; and one between the core and flap domains. These axes are drawn on the structures in Figure 5. The parametric equations for the axes are (33.06, 4.14, 3.68) + μ(−0.129, 0.159, 0.979) and (17.23, 23.54, 12.44) + μ(−0.839, 0.493, 0.231), respectively. Ligand binding results in a rotation of 7° for each core domain of SIV protease relative to the terminal domain. Translation along the screw axis is negligible. The axes are 5 Å from the carbon α atoms of the active site aspartic acids and perpendicular to the dimer 2-fold axis. The two core domains rotate together, narrowing the substrate binding pocket. The proximity of the axes to the centrally located catalytic aspartic acids reflects the greater movements at the exterior of the protein, as great as 4 Å between liganded

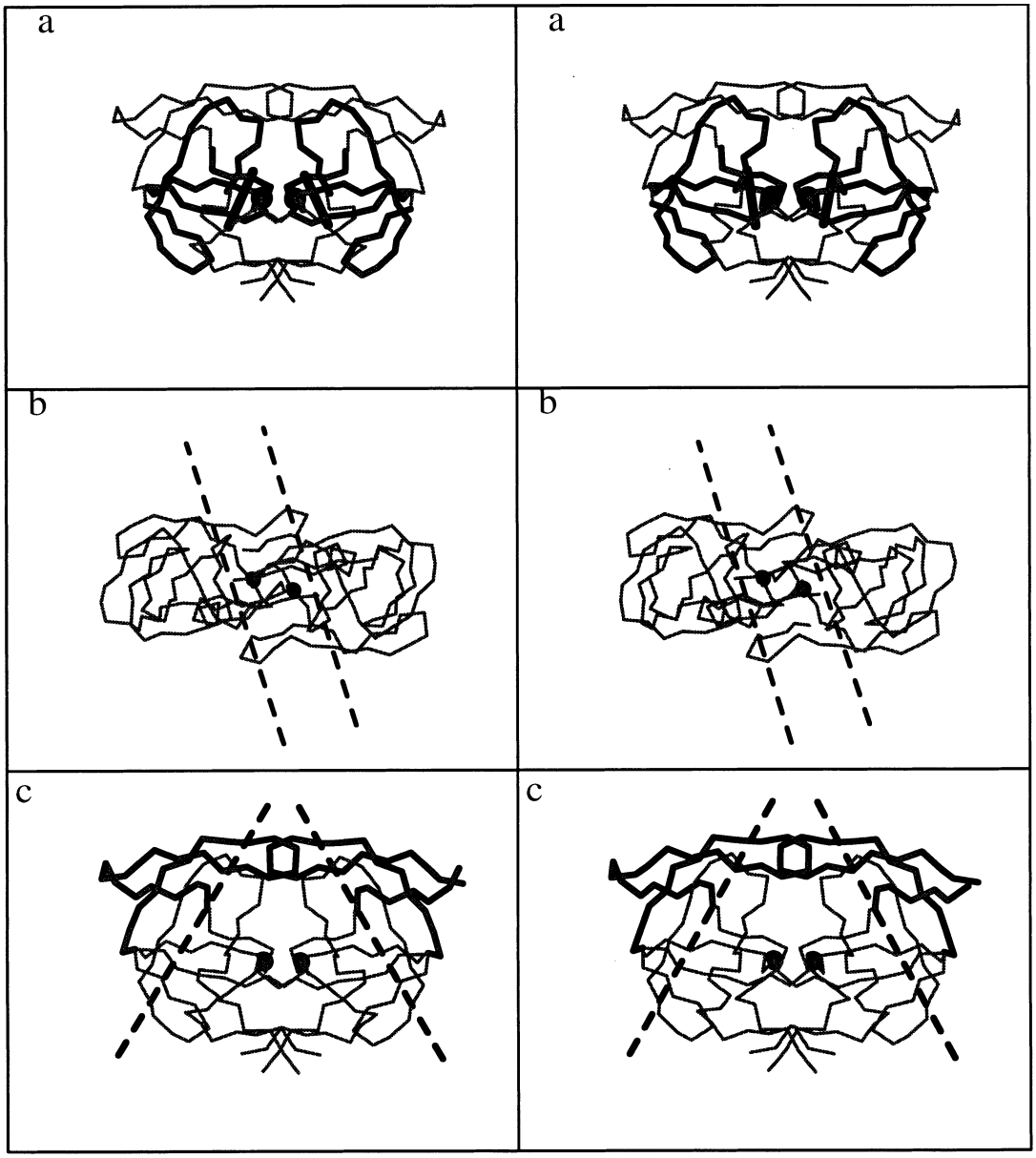


FIGURE 5: Axes determined for the rotation of the core and flap domains. a) C_α trace of the liganded SIV protease structure (1yth). The two axes about which the core domains rotate relative to the terminal domain are drawn. Dark spheres mark the catalytic aspartic acids. (b) The same as a but viewed down the dimer 2-fold axis toward the flaps. The axes of rotation are drawn with dotted lines. (c) C_α trace with the axes of rotation of the flap domains relative to the core domains is marked by dotted lines (figure drawn with Molscript (*73*)).

and unliganded structures when the terminal domains are overlapped.

Ligand binding also results in a rotation of 6° of the flap domain relative to the core domain (superimposing residues 33−45 and residues 54−62 of the flap domains). Again the translation along the screw axis is negligible. The axis intersects the protease near the interface between the flap domain and subdomain **d** of the core domain. This results in small difference distances between these residues, as noted above.

Comparison with HIV-1 Protease

The same domains describe the closing movement of the homologous HIV-1 protease (data not shown). Superimposing separate domains of the unliganded HIV-1 protease structure (PDB entry 3hvp) (*2*) on to domains of (no longer 2-fold symmetric) HIV-1 protease bound to the product peptide S−L-N−F (PDB entry 1yth) (*22*) resulted in an rms deviation for the carbon α atoms of 0.3 Å for the terminal domains, 0.5 Å for the flap domains, and 0.6 or 0.7 Å for the core domains (depending on the monomer). Excluding the flexible P1 loop, residues 77−81, the rms deviations for the core domains were even closer in structure, both within 0.5 Å of their counterpart. In comparison, the monomers, excluding the mobile flap residues (residues 46−54) superimposed with an rms deviation of 0.8 Å, and the dimers (also excluding residues 46−54) superimposed with an rms deviation of 1.2 Å. Therefore the closure necessary for substrate binding to HIV-1 protease implies motion of the same five domains as identified for SIV protease.

Ligand binding results in a rotation of 4° for the core domain relative to the terminal domain of HIV-1 protease. The axis is 2.5 Å from the carbon α atom of the aspartic acids, very similar to those found for SIV protease. The

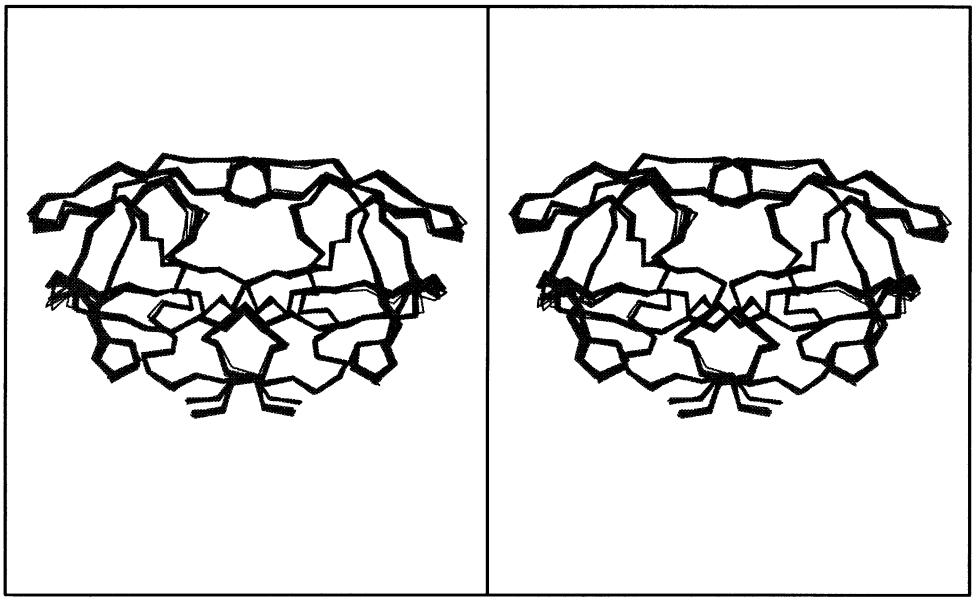


FIGURE 6: C_α traces of 17 HIV-1 protease structures after superimposition of the carbon α atoms onto 7hvp (figure drawn with Molscript (*73*)).

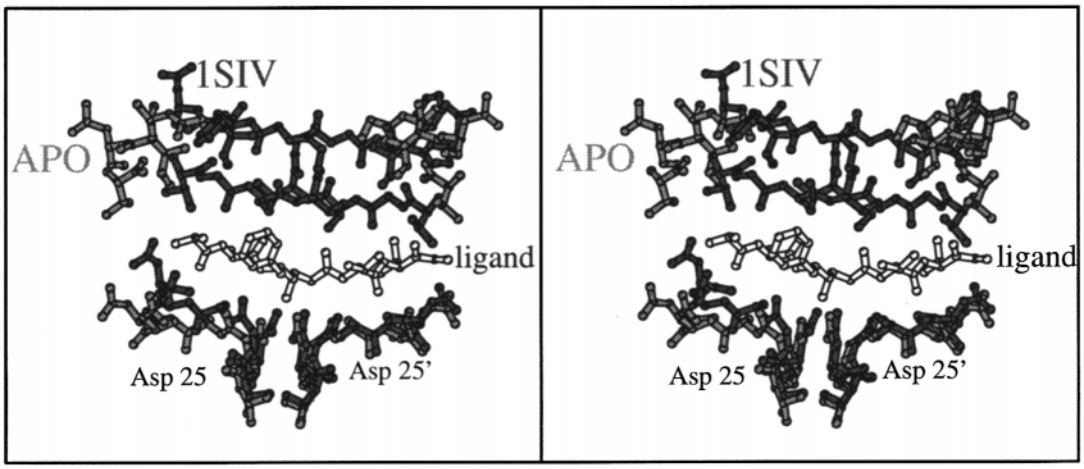


FIGURE 7: Ball-and-stick representation of the binding pocket of the unliganded (in gray) and liganded (1siv, in black) SIV protease structures with one core domain superimposed. The ligand (in white) is displayed in the active site, within hydrogen bonding distance of the superimposed monomer of the unliganded structure but too far to hydrogen bond to the unsuperimposed monomer. The α carbon atoms of the primed core domains were superimposed. The orientation of Asp 25 of the unliganded structure is seen to be rotated with respect to Asp 25 of the liganded structure (figure drawn with Molscript (*73*)).

flap domain rotates 4° relative to the core domain. These rotations are less pronounced than those of SIV protease, reflecting the less open conformation of the unliganded HIV-1 protease structure, versus the unliganded SIV structure.

To evaluate whether the orientation of the domains differ depending on the bound ligand, we superimposed 18 liganded structures deposited in the Brookhaven protein database. Table 1 lists the structures and their ligands. Figure 6 shows the carbon α traces of the structures superimposed on the dimer of 7hvp (residues 1−45 and 55−99 of both monomers). There is variation between the structures, however, the differences correlate with the thermal *B* factors in a manner consistent with plastic deformation unique to each inhibitor (*31*). The orientations of the domains are essentially identical to those of the "closed" form of the ligand-bound or substrate-bound form of the protease. Therefore any explanation for drug resistance mutations involving the conformational changes of the protease inherent in substrate binding must consider the open form of the protease or the dynamics in reaching the inhibitor-bound form.

Substrate Binding

The substrate binding pocket is dramatically reshaped upon ligand binding. Substrates bind to the protease in an extended β sheet conformation, hydrogen-bonding to the flaps and core domains (*1*). Flap residue G48, which contributes two backbone hydrogen bonds to the bound substrate, is disordered in the unliganded structure. Backbone hydrogen bonds of the P1−P4 positions of the substrate are made to one core domain, while those of the P1′−P4′ positions are made to the other core domain (nomenclature of Schecter and Berger) (*33*). In the unliganded structure, these domains are rotated apart, disrupting the β-stranded hydrogen bonds. Figure 7 shows a superposition of the unliganded and peptide-bound SIV protease structures. The monomer associated with the P1′−P4′ residues of the ligand has been used for the superposition.

Table 3: Movements of Atoms Forming the S1, S1′, and S3 Subsites between the "Open" and the "Closed" Conformation of SIV Protease

subsite	residue	$C_\alpha - C_\alpha$ distance (Å)[a]
S1′	Leu 23	1.6
S1′	Ile 81	3.1
S1′	Pro 82	2.5
S1′	Ile 84	3.9
S1	Gly 27	1.6
S3[b]	Arg 8′	0.8

[a] The carbon α atoms of the core domains (primed monomer) of the liganded and unliganded (PDB entry 1siv) SIV protease structures were superimposed, and the distances between the residues lining the S1′, S1, and S3 subsites from the unsuperimposed (unprimed) monomers were measured. [b] The Arg from the terminal domain of the primed monomer reaches into the S3 subsite.

The S1 and S3 subsites are also disrupted in the open structure. These binding pockets are formed by residues from both monomers. When one core domain of the liganded and unliganded structures is superimposed, residues Leu 23, Pro 81, Ile 82, and Ile 84 from the other monomer, forming the S1 subsite, are rotated away. Table 3 lists the shifts in position of these residues in the liganded versus unliganded structures. The side chain of Arg 8 reorients away from its position as part of the P3 binding pocket.

DISCUSSION

We report an unliganded SIV protease structure which is more "open" than the unliganded HIV-1 protease structures previously reported. By comparing this structure with SIV protease structures in the closed conformation, we identify five contiguous domains, considered as rigid domains of the protease dimer: two core domains, each containing one of the catalytic aspartic acids, a terminal domain forming the interface between the core domains, and two flap domains. The description of these domains as "rigid bodies" does not imply complete rigidity. For example, residues belonging to the P1 binding pocket, on the "P1 loop", undergo motions in addition to those of the rest of the core domain in the closed conformation. The same domain motions are also found to apply to HIV protease. Also, the tips of the flaps are more flexible than the remainder of the flap domain. The identification of rigid body domains does describe the overall movements of the protease that accompany association and dissociation of substrates, and inhibitors.

Proper identification of the rigid domains of the protease is essential in order to characterize the motions accompanying ligand binding. In addition to ordering residues at the tips of the flaps, ligand binding results in rigid body rotations of the core and flap domains. The two core domains each rotate 7° relative to the terminal domain. As a result, the catalytic aspartic acids, located close to the axes of rotation, move very little, while the surface residues move as much as 4 Å. This movement narrows the binding pocket, as observed in the first liganded HIV-1 protease structure (*1*). The two flap domains rotate 6° relative to the associated core domain, closing the flaps. The tips of the flaps, which are very mobile in the open structure, become ordered in the closed structure.

DOMAIN ROTATIONS OF MONOMERIC ASPARTYL PROTEASES

The rigid body domains of the dimeric retroviral proteases correspond closely to the rigid domains identified in the monomeric proteases (*17*). The two structurally similar N- and C-terminal lobes of pepsin-like proteases, presumed to have arisen by gene duplication from an ancestral dimeric protease (*34*), correspond to the core domains of the two retroviral protease monomers. A third "central" domain at the interface between the N- and C-terminal lobes is composed of six β-strands and two helices (*35*, *17*). This corresponds to the four β-strands and two helices of the terminal domain in the retroviral proteases. This domain structure has been conserved evolutionarily, indicating its importance for the catalytic mechanism.

Like the two core domains of the dimeric proteases, the two lobes of the monomeric aspartyl proteases rotate together in the liganded structure (*17*). The rotations described by Sali et al. ranged from 4 to 17.6°. The range of rotation angles reflects the fact that the orientation of the two lobes in the closed conformation varies with ligand and species in the monomeric aspartyl proteases. In contrast, the closed conformation of the retroviral proteases to date appears to be independent of ligand and species. Unlike the symmetrical dimeric proteases, the central β-sheet domain of the monomeric aspartyl proteases rotates with the N-terminal lobe. In both classes of protease, the catalytic aspartic acids are located near the rotation axes. The scissile bond is also positioned in both cases between two domains with the P1−P4 residues associated primarily with one domain (hydrogen bonded to the C-terminal domain in pepsin and one core domain in HIV protease), and the P1′−P4′ residues associated primarily with the other domain (the N-terminal domain in pepsin and the other core domain in HIV protease).

The interface between the core and terminal domains of the HIV and SIV proteases is formed primarily by hydrophobic side chains of the helix in the terminal domain and the β-strands of the core domain. The analogous interface in the monomeric proteases consists of a helix from the central domain and a 3_{10} helix from the C-terminal lobe. Sali et al. characterized the rearrangements at this interface that accompany the rigid body motion in terms of the "helix interface sheer mechanism" (*36*, *17*) in which large domain shifts, of several angstroms, are accommodated by small torsion angle shifts of side chains at a helix−helix interface. The side chain shifts are small enough not to disrupt the favorable packing interactions at the interface. While the interface between the core and terminal domains of the retroviral proteins are not helices packed together, a similar description applies: small shifts in the hydrophobic packing accommodate the reorientation of the domains.

Induced Fit Mechanism of HIV/SIV Proteases Involves Rotation of Five Domains

In their paper presenting an unliganded SIV protease structure in the closed conformation, Wilderspin and Sugrue proposed that the open HIV-1 protease structures were artifacts of crystal packing (*32*). We propose the opposite interpretation, namely, that their unliganded structure is closed due to crystal contacts. Their protein crystallized in space group $C222_1$, the same space group in which we previously obtained closed, liganded SIV protease structures at pH 6.5 (*21*). Their crystals were grown from 100 mM sodium acetate, pH 5.5, and 0.2 M NaCl. The open conformation of our unliganded protease grew from 100 mM sodium cacodylate, pH 6.5 and 0.3 M NaCl. The higher

Table 4: Contact Distances between Residues in the Interfaces Where Drug Resistant Mutations Arise in HIV Protease (a) between the Flap and Core Domains and (b) between the Terminal and Core Domains

source domain			target domain	
source domain	residue mutated in HIV-1 drug resistance[b]	equivalent residue in SIV protease[c]	closest contact of the mutant side chain (in Å) liganded/apo[d,g]	closest contact of the main chain at the mutant site (in Å) liganded/apo[f,h]
		(a) Between Flap and Core Domains		
CORE	Lys 20	Val 20	Ile 36 3.6/>4.0	
CORE	Val 32	Ile 32	Val 33 3.1/3.2	
			Ile 54 3.7/3.5	
			Thr 56 3.6/3.6	
CORE	Leu (Pro, Phe) 63	Glu 63	Val 62 3.5/3.7	Asn 61 3.9/>4.0
CORE	Val 75	Ile 75	Leu 38 >4.0/3.8	Lys 57 3.9/3.8
			Val 62 4.0/3.9	Glu 58 3.8/3.7
				Tyr 59 3.1/2.9
CORE	Val 77	Thr 77	Val 33 3.6/3.7	Val 33 3.2/3.1
			Leu 38 >4.0/3.9	Thr 56 3.2/3.4
			Lys 57 3.8/3.6	Lys 57 3.0/2.9
			Tyr 59 3.7/3.5	
FLAP	Leu 33	Val 33	Val 22 >4.0/3.8	Met 76 3.0/3.0
			Ile 32 3.1/3.2	Thr 77 3.2/3.1
			Met 76 4.0/>4.0	Gly 78 3.3/3.1
			Asn 83 3.9/3.8	Thr 80 >4.0/4.0
				Asn 83 3.6/3.5
FLAP	Glu 35	Gly 35	Val 77 3.5[e]	
			Gly 78 3.7	
FLAP	Met 36	Ile 36	Ile 15 3.9/4.0	Gln 18 >4.0/3.3
			Gln 18 >4.0/3.3	
			Ile 20 3.6/>4.0	
FLAP	Ile 54	Ile 54	Ile 32 3.7/3.5	
			Asp 79 >4.0/3.5	
FLAP	Lys 45	Lys 45	Met 76 >4.0/3.6	
		(b) Between Terminal and Core Domains		
TERMINAL	Arg 8	Arg 8	Val 10 3.9/4.0	Val 10 3.1/2.9
TERMINAL	Leu 89	Leu 89	Set 31 3.3/3.5	
			Ile 71 4.0/>4.0	
			Ile 75 3.7/>4.0	
			>4.0/4.0	
TERMINAL	Leu 90	Leu 90	Leu 24 4.0/>4.0	
			Asp 25 >4.0/3.8	
			Thr 26 3.9/>4.0	
TERMINAL	Arg 8′	Arg 8'	Asp 29 3.2/3.5	
	Leu 97′	Leu 97′	Thr 26 3.9/>4.0	
CORE	Leu 10	Val 10	Arg 8 3.9/4.0	Lys 7 >4.0/4.0
			Pro 9 3.6/3.7	Arg 8 3.1/2.9
CORE	Leu 24	Leu 24	Phe 3 3.8/>4.0	Phe 85 3.9/3.7
			Pro 9 3.6/3.6	Pro 90 2.9/3.0
			Phe 85 3.9/>4.0	
			Leu 90 4.0/>4.0	
			Met 95 >4.0/4.0	
CORE	Ala 71	Ile 71	Leu 89 4.0/>4.0	
			Ala 92 >4.0/3.8	
			Leu 93 4.0/3.8	

[a] Footnotes *b*−*f* apply to the distances between the flap and core domains, and footnotes *g* and *h* apply to the distances between the terminal and core domains. [b] Some of the polymorphic sequences of HIV-1 strains are listed in parentheses. [c] Distances in this table are for SIV protease instead of HIV protease in order to compare distances in the closed conformation with those of the unlganded, open conformation (for which there is no HIV structure). [d] Only the closest contact between the side chain of sites where the mutations arise, and residues in the neighboring domain (the side chain or main chain) is listed. Distances are compared between the liganded (closed) SIV protease and the unliganded (open) SIV protease structures. Residues are considered to be contacting if they are within 4.0 Å. [e] Distances for the HIV-1 protease sequence. The SIV protease sequence contains a Gly at this position, which is not contacting the flap domain. The distances are for the closed conformation of HIV-1 protease. [f] Only the closest contact between the mainchain of sites where the mutations arise and residues in the neighboring domain (the side chain or main chain) is listed. Distances are compared between the liganded (closed) SIV protease and the unliganded (open) SIV protease structures. [g] Only the closest contact between the side chain of sites where the mutations arise and residues in the neighboring domain (the side chain or main chain) is listed. Distances are compared between the liganded (closed) SIV protease and the unliganded (open) SIV protease structures. Residues are considered to be contacting if they are within 4.0 Å. [h] Only the closest contact between the mainchain of sites where the mutations arise, and residues in the neighboring domain (the side chain or main chain) is listed. Distances are compared between the liganded (closed) SIV protease and the unliganded (open) SIV protease structures.

pH may favor crystallization of the unliganded protein in the open conformation.

Although the fully closed conformation of the protease is largely ligand-independent, a range of unliganded conforma-

tions of the protease have now been reported, from the completely closed conformation of SIV protease (*32*) to the most open conformation of SIV protease reported here. The unliganded HIV-1 protease structures are intermediate between open and closed as judged by the rotation about the axis between the core and terminal domains (7° in SIV protease and 4° in HIV-1 protease) and the conformation of the flap tips, which are ordered in the open HIV-1 protease structures. Although this difference may be a function of the species, we propose that HIV-1 protease can adopt the same open conformation observed for SIV protease if not prevented by crystal packing contacts. Such a degree of opening is certainly required for substrate binding and therefore must be accessible. A conformation intermediate between the unliganded HIV-1 protease structures and the closed structure has also been reported (*37*). In this structure, HIV-1 protease is bound to a UCSF discovered, haloperidol-based inhibitor, UCSF8. The two core domains were rotated by 3.8 and 3.3° relative to the terminal domain, and the conformation of the flaps was intermediate between the unliganded and closed structures.

The fact that our unliganded protease could be crystallized in the closed conformation suggests that the free energy difference between the open and closed conformation is small because the crystallization conditions are very similar. It also suggests that the closed conformation is accessible to the protease in the absence of ligand, and that ligand binding either stabilizes this closed conformation, or more likely induces the "fit" to the ligand. Because intermediate states have also been observed, these conformations may be accessible to the unliganded protease as well. It is possible that an even more open conformation exists in solution than has been crystallized.

An induced fit mechanism, as proposed by Koshland, hypothesized that substrate binding caused a conformational change in the enzyme, properly orienting residues from the enzyme for catalysis (*38*). Small or poorly bound substrates would therefore not react because they could not make the contacts necessary to stabilize the closed, active conformation. Consistent with an induced fit mechanism, HIV protease requires up to four residues both N- and C-terminal of the scissile bond for efficient cleavage (*39*).

A similar preference for longer substrates was observed for pepsin: longer substrates increased the reaction velocity though without a corresponding increase in binding affinity (*18*). Fruton proposed that the binding energy from the substrate is primarily used to strain the scissile peptide bond toward the tetrahedral transition state (*18, 22, 40*−*42*). Modeling the substrate from products indicates that substrate strain does not contribute to catalysis (*22*). The induced fit model provides a different explanation: the binding energy of longer substrates does not increase their binding affinity because it is used to stabilize the conformational change. A good substrate increases k_{cat} because it forms the Michaelis complex properly (*43*).

The open conformation of SIV protease suggests some of the interactions that a substrate must make with the protein to effectively stabilize the closed conformation. Main chain hydrogen bonds on both sides of the scissile bond are accessible in the closed but not the open conformation. The S1, S1′, S3, and S3′ subsites are also disrupted in the open conformation. A good substrate presumably makes enough of these contacts to allow some of its binding energy to stabilize the closed conformation of the protease.

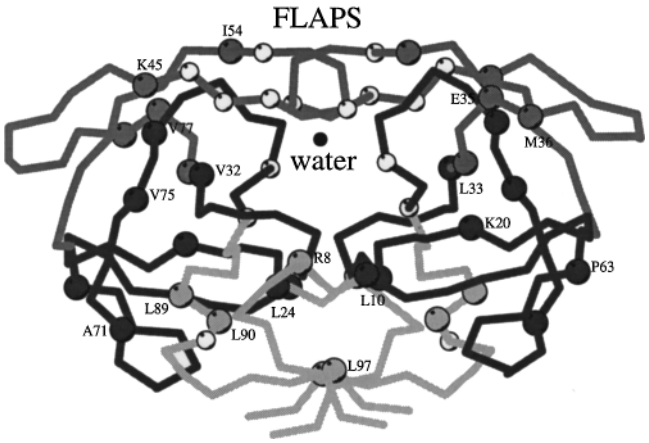


FIGURE 8: Resistance mutations mapped onto an α carbon trace of HIV-1 protease. Rigid body domains of the protease are shaded: terminal domain shaded whitest, the core domains shaded black, the flap domains shaded gray. The spheres mark residues found to mutate in cell culture or clinical trials after treatment with protease inhibitors. Four terminal domain residues (gray spheres) and 3 core domain residues (black spheres) pack at the interface of the two domains; five core domain residues and 4 flap domain residues pack at the interface of the two domains. (Only one residue of each symmetry-related dimer is labeled.) The uncolored spheres are mutations not at the interface of two domains, mostly in the active site (figure drawn with Molscript (*73*)).

Drug Resistance Mutations Related to Domain Rotations

A major difficulty in the development of anti-HIV drugs has been the rapid selection of mutations in the viral genome that result in resistance to the drug by means of resultant changes within the protein target. Mutations in at least 25 protease residues have been selected for, following treatment with individual anti-protease inhibitors in cell culture or in clinical trials (*43*−*49*). Whereas the inhibitors are effective against the wild-type protease, they compete less effectively with substrates in the mutated enzymes. Combination therapies, which would require multiple mutations to achieve resistance to multiple inhibitory drugs, may overwhelm the virus's repertoire of mutations (*50*). However the mutations selected in response to one drug are often selected for by exposure to a completely different one, suggesting that a direct and specific interaction with the drug is not the paradigm to explain drug resistance.

Figure 8 locates the sites of the 25 resistance mutations available at the time of our analysis, on the HIV-1 protease structure. The six mutations that are in direct contact with the inhibitor or substrate (Arg 8, Val 32, Ile 47, Gly 48, Val 82, Ile 84) are the easiest to rationalize. As expected, crystal structures of these binding site mutants complexed with inhibitors indicate that the mutation disrupts inhibitor binding more than substrate binding (*51*−*54*). Resistance mutations far from the active site are harder to explain. Seven sites of resistance mutations are located at the interface between the core and terminal domains as shown in Figure 9a (Arg 8, Leu 10, Leu 24, Ala 71, Leu 89, Leu 90, Leu 97 in the HIV-1 sequence), and 10 are located at the interface between the core and flap domains as revealed in Figure 9b (Lys 20, Val 32, Leu 33, Glu 35, Met 36, Lys 45, Ile 54, Pro 63, Val 75, Val 77 in the HIV-1 sequence). While Val 32 and Arg 8

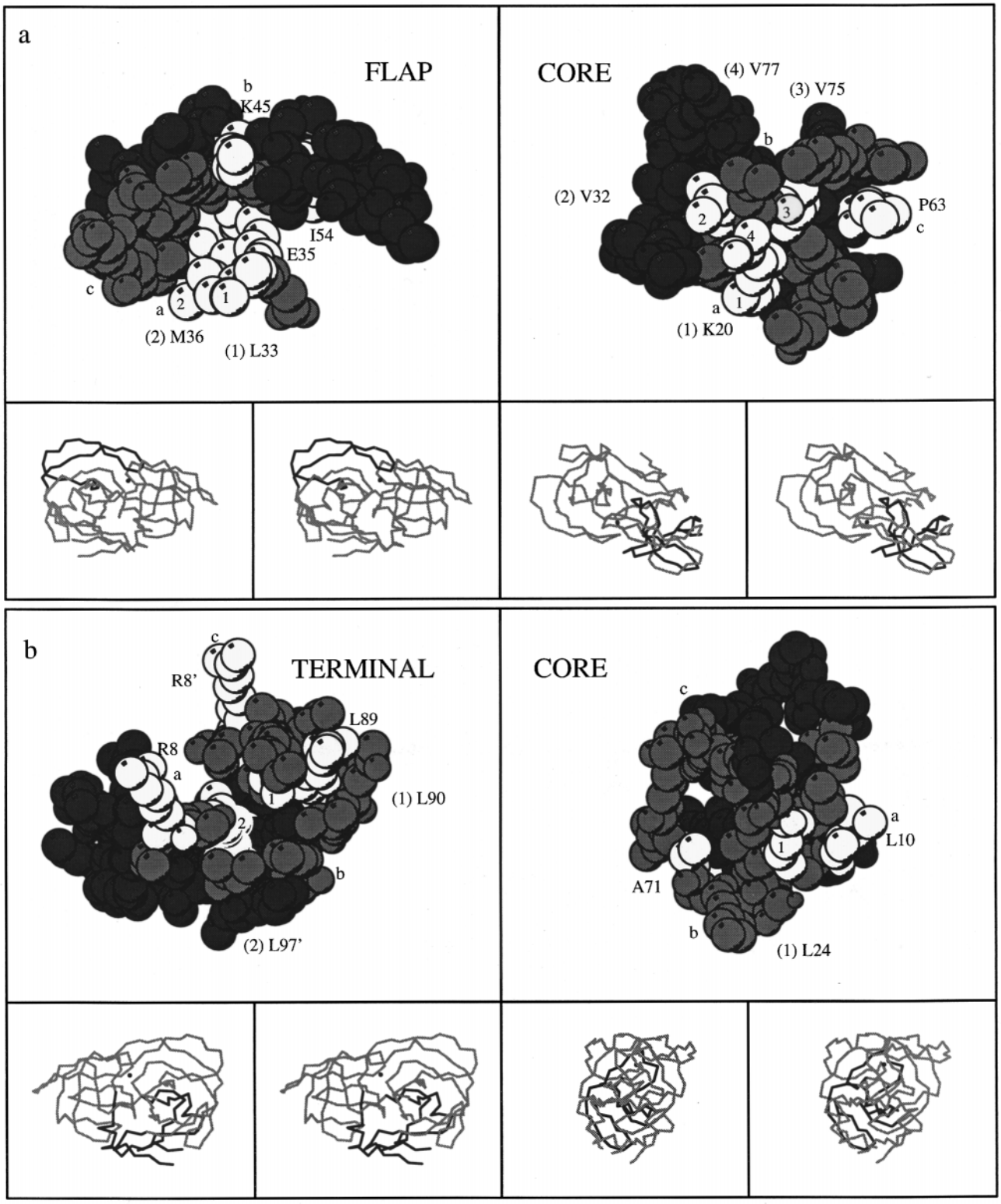
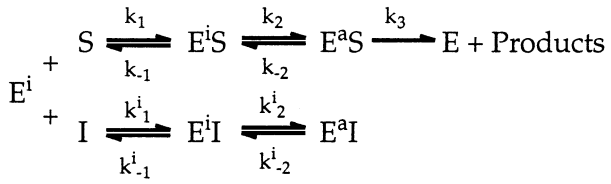


FIGURE 9: Drug resistance mutations (white van der Waals surfaces, labeled) that are in domain contact interfaces (white and gray surfaces) (18 are shown). Contact surfaces between (a) the flap and core domains, and (b) the terminal and core domains of HIV-1 protease (PDB entry 1yth). The domains are opened by −90° on the left, and +90° on the right side so that the two contacting surfaces face outward toward the viewer. The contacting surfaces pack together by superimposing the letters a, b, and c marked around the periphery of one domain with those of the opposed domain. Each of the a,b orientations is shown in the context of the complete protease dimer in the crossed-eye stereopair of the C_α trace of the protease shown beneath each of the space-filling pictures. The domain depicted in each panel is marked in bold in the C_α trace. The surface representation is composed of spheres drawn at the van der Waals radius around each atom. Gray and white spheres are in contact (within 4 Å) with the opposed surface, and black spheres depict atoms that are not contacting the opposed surface. Residues mutated in drug-resistant variants are white, with the residue type and sequence number labeled or keyed with a number (primes denote residues from the second monomer in the terminal domain). The nonmutated contacting residues are as follows: flap domain, 33, 37, 38, 56−62; core domain contacting the flap domain, 15, 16, 18, 72, 73, 74, 76, 78, 83; terminal domain, 3, 7, 9, 85−88, 92, 93, 5′, 9′, 99′; core domain contacting the terminal domain, 13, 22, 23, 25, 26, 28−31, 66, 67, 69, 70, 72, 73, 74, 82.

Scheme 1



each contact substrate, they are also located at an interdomain interface. These 19 mutations suggest that domain orientation or movement may be a factor in the development of resistance to anti-protease drugs. Four additional mutations (Met 46, Ile 50, and Phe 53 in the flap domain and Thr 91 in the terminal domain) are neither part of the substrate binding pocket nor in direct contact between domains at an interface. The first three of these however, may affect dynamics of the flap closure, and so may affect the bound versus free equilibrium, and hence K_i (*12*).

One scenario that could select against an inhibitor would be a mutation that alters the orientation of the domains in the closed conformation of the protease, thereby altering the shape of the binding pocket. Sali et al. proposed that differences in the orientation of the rigid domains of monomeric aspartyl proteases explained variations in substrate specificity (*17*). Like mutations of residues in the binding pocket, these mutations would directly impact interactions between ligand and protein. Structures of the mutated protein in the closed conformation would be predicted to show changes in the substrate binding pocket. However, unlike the monomeric aspartyl proteases, no variation in domain orientations have been observed between substrate−analogue bound protease structures, suggesting that the closed conformation is important to access for catalysis; mutations that did alter the closed conformation might be selected against due to reduced catalytic activity.

A more subtle effect would be a mutation that alters the equilibrium between the closed, and open conformations of the protease. Both substrates and inhibitors bind to the protease in a multistep process: first forming a complex with an open, inactive conformation, followed by a conformational change to the closed enzyme (*3*). Scheme 1 describes the competition between substrate and inhibitor for an open conformation of the protease. Because free enzyme would allow substrate binding, the best inhibitors would be irreversible or at least have a slow off-rate. Changes in the free energy of the conformational change from the protease's inactive form (E^i) to the active form (E^a) which favor the open conformation (i.e. decrease the association constant $K^i_2 = k_2/k_{-2}$) would increase the off-rate of the inhibitor, making unliganded protease available for substrate binding. Although such a mutation might increase the off-rate of the substrate proportionately, two factors could enhance the competitiveness of substrate over inhibitor for free E^i: if $k_2 \gg k_{-1}$, then substrate binding would immediately lead to cleavage. In addition, the on-rate for substrate might be faster than for the inhibitor ($k_1 > k^i_1$). This may particularly be true for small, rigid inhibitors, all of which to this date have been designed to match the closed protease structure or to mimic bound substrate in the closed form. Because the substrate is large and flexible, it may conform to the open conformation of the protease, resulting in a fast on-rate. The inhibitor may not develop its affinity until the induced conformation change in the enzyme occurs, and may therefore have a slow on-rate. Thus mutations that affect the rate of the closing process would be reflected in the reduced competitiveness of the inhibitor versus substrate while still implying the same closed state once bound.

Of the 25 drug resistance mutations, 19 are not in contact with the inhibitor and are often distant from the protease binding pocket (Figure 8). Many of these mutations occur multiply, i.e., the same mutation in response to different inhibitors of entirely different chemical frameworks. In this sense these drug resistant mutations are conserved (e.g. L90M, A71V), and therefore suggest that a mechanism independent of which inhibitor is used, such as the mechanism proposed here, contributes to drug resistance. Although each different protease inhibitor selects for a distinct array of mutations, the population of protease molecules containing interdomain mutations increases in response to each of the different inhibitors. And these mutated proteases often show cross-resistance to inhibitors other than those selected against, indicating that a common element of response to the different inhibitors has been affected. For example, saquinavir treatment in vivo strongly selects for the Leu 90 Met mutation, and can also select for the mutation Ala 71 Val in addition to the active site mutation of Gly 48 (*55*). The same Leu 90 Met mutation contributes to resistance to the protease inhibitor MK-639 (*44*). The same Ala 71 Val mutation also appears after treatment with ritonavir, as does the Leu 90 Met mutation (*49*). Treatment in clinical trials with MK-639 results in the mutations Met 46 Ile, Leu 10 Arg, and Leu 63 Pro as well as the active site mutations of Val 82 and Ile 84. This combination of mutations shows cross-resistance to five other structurally distinct protease inhibitors. These observations therefore support a mechanism for drug resistance that is not specific to each inhibitor structure, rather, one that affects a conserved requirement of the dynamic process of protease closure, and one that may therefore not show differences in the structures of proteases inhibited with these different drugs.

Maschera et al. propose that the resistance mutations arising in response to the inhibitor saquinavir alter the protease's equilibrium between the active and inactive form by increasing the off-rate of the inhibitor (*48*). They measured k^i_{-1} to be much less than k^i_2. Under these circumstances, the inhibitor's $k_{off} = k^i_{-1}/K^i_2$. K^i_2 would be large, favoring the closed conformation of the bound enzyme, and k_{off} would be slow. They proposed that the resistance mutations Leu 90 Met and Gly 48 Val reduced the affinity of the protease for saquinavir by decreasing K^i_2. They characterized the conformational change between E^i and E^a as flap closing, rather than the rigid body movements proposed here, though Leu 90 is far from the flaps. Leu 90 is situated on the α-helix of the terminal domain interacting with the hydrophobic surface of the core domain. Mutating these residues to Met may disrupt the packing between the two domains, thereby decreasing K^i_2. Another mutation, Arg 8 Gln, disrupts an electrostatic interaction between the terminal and core domains (*56*). This interaction is not found in the open conformation of the protease due to rearrangement of the Arg 8 side chain. In light of the domain motion, it would not be surprising if the liganded structures of these drug resistant mutant enzymes appeared essentially unchanged in their active sites. Analysis of the domain motion

is clearly necessary for an understanding of the mechanism of drug resistance by HIV protease.

These observations suggest a strategy for designing drugs against protease mutations at domain interfaces. A less rigid inhibitor, or an inhibitor that targets an open conformation of HIV protease, would compete better with the on-rate of the substrate. Such a drug would complement the existing drugs that target the closed conformation of the protease and would certainly be less likely to evoke resistance mutations by changes that are not in contact with the drug.

ACKNOWLEDGMENT

We thank Nancy Douglas for protein purification and Nick Endres for setting up crystals. We thank Tom Stout, Earl Rutenber, and Janet Finer-Moore for valuable discussions.

REFERENCES

1. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B., and Wlodawer, A. (1989) *Science 246*, 1149−1152.
2. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) *Science 245*, 616−621.
3. Furfine, E. S., D'Souza, E., Ingold, K. J., Leban, J. J., Spector, T., and Porter, D. J. T. (1992) *Biochemistry 31*, 7886−7891.
4. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., et al. (1989) *Nature 342*, 299−302.
5. Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature 337*, 615−20.
6. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H., and Poljak, R. J. (1991) *Biochimie 73*, 1391−6.
7. Jaskólski, M., Miller, M., Mohana Rao, J. K., Leis, J., and Wlodawer, A. (1990) *Biochemistry 29*, 5889−5898.
8. Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J., Daugman, J. D., Wingfield, P. T., Lam, P. Y. S., Jadhav, P. K., Hodge, C. N., Domaille, P. J., and Chang, C. (1995) *Nat. Struct. Biol. 2*, 274−280.
9. Rodriguez, E. J., Debouck, C., Deckman, I. C., Abu-Soud, H., Raushel, F. M., and Meek, T. D. (1993) *Biochemistry 32*, 3557−3563.
10. Harte, W. E., Jr., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., and Beveridge, D. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8864−8868.
11. Harte, W. E., Jr., Swaminathan, S., and Beveridge, D. L. (1992) *Proteins 13*, 175−194.
12. Collins, J. R., Burt, S. K., and Erickson, J. W. (1995) *Nat. Struct. Biol. 2*, 334−338.
13. Pearl, L. H., Taylor, W. R. (1987) *Nature 329*, 351−354.
14. Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J., and Wlodawer, A. (1989) *Nature 337*, 576−579.
15. Lin, Y., Fusek, M., Lin, X., Hartsuck, J. A., Kezdy, F. J., and Tang, J. (1992) *J. Biol. Chem. 267*, 18413−18418.
16. James, M. N. G., Sielecki, A., Salituro, F., Rich, D., and Hofmann, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6137−6141.
17. Sali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T., and Blundell, T. L. (1992) *Proteins 12*, 158−70.
18. Fruton, J. S. (1976) in *Methods in Enzymology* (Meister, A., Ed.) pp 1−35, Wiley and sons: New York.
19. Fruton, J. S. (1980) *Mol. Cell. Biochem. 32*, 105−114.
20. Abad-Zapatero, C., Rydel, T. J., and Erickson, J. (1990) *Proteins 8*, 62−81.
21. Rose, R. B., Rosé, J. R., Salto, R., Craik, C. S., and Stroud, R. M. (1993) *Biochemistry 32*, 12498−12507.
22. Rose, R. B., Craik, C. S., Douglas, N. L., and Stroud, R. M. (1996) *Biochemistry 35*, 12933−12944.
23. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
24. Sack, J. (1988) *J. Mol. Graphics 6*, 225.
25. Zhao, B., Winborne, E., Minnich, M. D., Culp, J. S., Debouck, C., and Abdel-Meguid, S. S. (1993) *Biochemistry 32*, 13054−13060.
26. Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F., and Stroud, R. M. (1990) *Proteins 8*, 315−33.
27. Richards, F. M., and Kundrot, C. E. (1988) *Proteins 3*, 71−84.
28. Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer*, 2nd ed., Addison-Wesley Publishing Company, Inc., Redwood City, CA.
29. Goldstein, H. (1981) *Classical Mechanics*, Addison-Wesley Publishing, Menlo Park, CA.
30. Brünger, A. T. (1992) *Nature 355*, 472−474.
31. Stroud, R. M., and Fauman, E. B. (1995) *Protein Sci. 4*, 2392−2404.
32. Wilderspin, A. F., and Sugrue, R. J. (1994) *J. Mol. Biol. 239*, 97−103.
33. Schecter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
34. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., and Blundell, T. L. (1978) *Nature 271*, 618−621.
35. Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. G. (1990) *J. Mol. Biol. 214*, 143−170.
36. Chothia, C., and Lesk, A. M. (1985) *Trends Biol. Sci. 10*, 116−120.
37. 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., and Salto, R. (1993) *J. Biol. Chem. 268*, 15343−15346.
38. Koshland, D. E. J. (1958) *Proc. Nat. Acad. Sci. U.S.A. 44*, 98−104.
39. Dunn, B. M., Gustchina, A., Wlodawer, A., and Kay, J. (1994) *Methods Enzymol. 241*, 254−301.
40. Pearl, L., and Blundell, T. (1984) *FEBS Lett. 174*, 96−101.
41. Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 189−215.
42. Jaskólski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., and Wlodawer, A. (1991) *Biochemistry 30*, 1600−1609.
43. Ray, J., W. J., Post, C. B., and Puvathingal, J. M. (1993) *Biochemistry 32*, 38−47.
44. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. A. (1995) *Nature 374*, 569−571.
45. Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., and Erickson, J. W. (1995) *Biochemistry 34*, 9282−9287.
46. Borman, A. M., Paulous, S., and Clavel, F. (1996) *J. Gen. Virol. 77*, 419−426.
47. Erickson, J. W., and Burt, S. K. (1996) *Annu. Rev. Pharmacol. Toxicol. 36*, 545−571.
48. Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and Furfine, E. S. (1996) *J. Biol. Chem. 271*, 33231−33235.
49. Molla, A., Korneyeva, M., Gao, Q., Vasavanonda, S., Schipper, P., Mo, H., Markowitz, M., Chernyavskiy, T., Niu, P., Lyons, N., Hsu, A., Granneman, G., Ho, D., Boucher, C., Leonard, J., Norbeck, D., and Kempf, D. (1996) *Nat. Med. 2*, 760−766.
50. Deeks, S. G., Smith, M., Holodniy, M., and Kahn, J. O. (1997) *JAMA 277*, 145−153.
51. Baldwin, E. T., Bhat, T. N., Liu, B., Pattabiraman, N., and Erickson, J. W. (1995) *Nat. Struct. Biol. 2*, 244−249.
52. Chen, Z., Li, Y., Schock, H. B., Hall, D., Chen, E., and Kuo, L. C. (1995) *J. Biol. Chem. 270*, 21433-21436.
53. Erickson, J. W. (1995) *Nature Struct. Biol. 2*, 523−529.
54. Ridky, T., and Leis, J. (1995) *J. Biol. Chem. 270*, 29621−29623.
55. Roberts, N. A. (1995) *AIDS 9* (Suppl. 2), S27−S32.

56. Ho, D. D., Toyoshima, T., Mo, H., Kempf, D. J., Norbeck, D., Chen, C., Wideburg, N. E., Burt, S. K., Erickson, J. W., and Singh, M. K. (1994) *J. Virol. 68*, 2016−2020.
57. Dreyer, G. B., Lambert, D. M., Meek, T. D., Carr, T. J., Tomaszek, T. A., Jr., Fernandez, A. V., Bartus, H., Cacciavillani, E., Hassell, A. M., Minnich, M., et al. (1992) *Biochemistry 31*, 6646−59.
58. Hoog, S. S., Zhao, B., Winborne, E., Fisher, S., Green, D. W., DesJarlais, R. L., Newlander, K. A., Callahan, J. F., Moore, M. L., Huffman, W. F., et al. (1995) *J. Med. Chem. 38*, 3246−52.
59. Murthy, K. H., Winborne, E. L., Minnich, M. D., Culp, J. S., and Debouck, C. (1992) *J. Biol. Chem. 267*, 22770−22778.
60. Priestle, J. P., Fassler, A., Rosel, J., Tintelnot-Blomley, M., Strop, P., and Grutter, M. G. (1995) *Structure 3*, 381−389.
61. Thanki, N., Rao, J. K., Foundling, S. I., Howe, W. J., Moon, J. B., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Thaisrivongs, S., and Wlodawer, A. (1992) *Protein Sci. 1*, 1061−1072.
62. Abdel-Meguid, S. S., Zhao, B., Murthy, K. H., Winborne, E., Choi, J. K., DesJarlais, R. L., Minnich, M. D., Culp, J. S., Debouck, C., Tomaszek, T. A., Jr., et al. (1993) *Biochemistry 32*, 7972−7980.
63. Thompson, S. K., Murthy, K. H., Zhao, B., Winborne, E., Green, D. W., Fisher, S. M., DesJarlais, R. L., Tomaszek, T. A., Jr., Meek, T. D., Gleason, J. G., et al. (1994) *J. Med. Chem. 37*, 3100−3107.
64. Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., et al., (1995) *J. Am. Chem. Soc. 117*, 1181−1182.
65. Jhoti, H., Singh, O. M., Weir, M. P., Cooke, R., Murray-Rust, P., and Wonacott, A. (1994) *Biochemistry 33*, 8417−8427.
66. Hosur, M. V., Bhat, T. N., Kempf, D. J., Baldwin, E. T., et al., (1994) *J. Am. Chem. Soc. 116*.
67. Lam, P. Y., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C., Weber, P. C., Jackson, D. A., Sharpe, T. R., and Erickson-Viitanen, S. (1994) *Science 263*, 380−384.
68. Abdel-Meguid, S. S., Metcalf, B. W., Carr, T. J., Demarsh, P., DesJarlais, R. L., Fisher, S., Green, D. W., Ivanoff, L., Lambert, D. M., Murthy, K. H., et al. (1994) *Biochemistry 33*, 11671−11677.
69. Bone, R., Vacca, J. P., Anderson, P. S., and Holloway, M. K. (1991) *J. Am. Chem. Soc. 113*, 9382−9384.
70. Fitzgerald, P. M., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C. T., Herber, W. K., Dixon, R. A., and Darke, P. L. (1990) *J Biol Chem 265*, 14209−14219.
71. Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B., and Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805−9.
72. 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., and et al. (1990) *Science 249*, 527−533.
73. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI9716074