Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution structure of unbound human immunodeficiency virus 1 subtype C protease: implications of flap dynamics and drug resistance

Roxana M. Coman,[a] Arthur H. Robbins,[a] Maureen M. Goodenow,[b] Ben M. Dunn[a]* and Robert McKenna[a]*

[a]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32603, USA, and [b]Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL 32603, USA

Correspondence e-mail: bdunn@ufl.edu, rmckenna@ufl.edu

Received 8 February 2008
Accepted 30 April 2008

PDB Reference: HIV-1 subtype C protease, 2r8n, r2r8nsf.

The X-ray crystal structure of the unbound state of human immunodeficiency virus 1 (HIV-1) subtype C protease (C PR) has been determined to 1.20 Å resolution in the tetragonal space group $P4_12_12$, with one monomer per asymmetric unit and unit-cell parameters a = 46.7, c = 100.8 Å, allowing full anisotropic least-squares refinement. The refined model has a conventional R factor of 14.1% for all reflections and estimated standard deviations in bond lengths and angles for all main-chain non-H atoms of 0.014 Å and 0.030°, respectively. The structure is compared with three unbound subtype B proteases (B PRs) to identify structural changes arising from the naturally occurring polymorphisms and delineate their implications in antiretroviral drug resistance/susceptibility. The unbound C PR exhibits a larger distance between the tips of the flaps, a downward displacement of the 36–41 loop and an increased thermal stability of the 10s loop when compared with the B PR structures. The C PR structure presents the highest resolution of the unbound state of a non-subtype-B PR and adds to the understanding of flap dynamics and drug resistance.

1. Introduction

Human immunodeficiency virus type 1 (HIV-1) protease (PR) is an aspartic hydrolase that functions as an obligatory homodimer with 99 amino acids in each subunit (labeled 1–99 and 1′–99′). Its role is to cleave the gag and gag/pol polyproteins into structural and enzymatic proteins and to induce the formation of mature infectious virions. The inhibition of this enzyme yields immature HIV virions that are incapable of spreading the infection. Because of its essential role in gaining viral infectivity, HIV-1 PR has been considered an attractive target for discovering new and potent anti-HIV drugs (Spaltenstein *et al.*, 2005; Wlodawer & Erickson, 1993). Extensive structural studies have been performed in an attempt to better understand the molecular mechanisms that govern the interactions between this enzyme and substrates or inhibitors. The HIV-1 subtype B PR (B PR) structure has been determined both alone (unbound; Heaslet, Lin *et al.*, 2007; Logsdon *et al.*, 2004; Spinelli *et al.*, 1991) and complexed with different protease inhibitors (PIs; Louis *et al.*, 2007; Velazquez-Campoy, Muzammil *et al.*, 2003; Vondrasek & Wlodawer, 2002). The crystal structures show that HIV-1 PR forms a binding site that consists of subsites S4–S4′, which span about eight residues (P4–P4′) of a peptide substrate (Schechter & Berger, 1967). Many HIV-1 PR mutants have also been crystallized unbound or complexed with peptidomimetic or non-peptidomimetic inhibitors (Heaslet, Lin *et al.*, 2007; Louis *et al.*, 2007; Vondrasek & Wlodawer, 2002). Some of these mutants show

structural changes consistent with the differences observed in their enzymatic activity (Vondrasek & Wlodawer, 2002). The availability of such a diverse panel of HIV-1 B PR variants has played a major role in the process of drug development.

This wealth of information regarding interactions between HIV-1 PR and substrates or inhibitors has until recently only been available for B PR. However, HIV-1 is characterized by a large genetic diversity and is classified in subtypes, sub-subtypes and circulating and unique recombinant forms (Brodine *et al.*, 1995; Fleury *et al.*, 2003; Kantor & Katzenstein, 2004); the differences between these are at both the genetic and protein levels, and are called naturally occurring polymorphisms (NOPs). It has been argued that these polymorphisms might play roles such as increasing the catalytic activity of non-B subtype PRs (non-B PRs; Velazquez-Campoy *et al.*, 2001), resulting in development of diverse mutational pathways during antiretroviral treatment (Dumans *et al.*, 2004; Grossman *et al.*, 2004), influencing the speed of acquiring PI-related resistance mutations (Hirsch *et al.*, 2000; Vergne *et al.*, 2000), contributing to resistance and/or maintenance of viral fitness once primary resistance mutations occur (Rose *et al.*, 1996; Velazquez-Campoy, Vega *et al.*, 2003) and promoting a poorer response to therapy (Servais *et al.*, 2001).

The first crystal structure of a non-B PR, subtype F PR, was reported recently (Sanches *et al.*, 2004, 2007). In this study, the authors compared B and F PRs and concluded that two naturally occurring polymorphic substitutions in F and other non-B PRs, M36I and L89M, may lead to early development of drug resistance in patients infected with non-B HIV-1 subtypes. In addition, recently Coman and coworkers reported the crystallization and inhibition of the C PR complexed with the PIs indinavir (IDV) and nelfinavir (NFV) (Coman *et al.*, 2007).

In this work, the differences and similarities between B and C PRs and the roles of NOPs in C PR are studied, with the goal of understanding how these sequence variations affect the overall characteristics of C PR and ultimately how the response to treatment is modulated by the pre-existence of the NOPs. Combining kinetic and structural data will provide a broad understanding of the interactions that are being affected by residue changes and how different PIs or substrates interact with the same enzyme. This knowledge will provide clinicians with information for optimization of therapeutic regimens and researchers with clues to designing inhibitors that will retard the evolution of resistant PR variants and exhibit less cross-resistance with other classes of PIs.

2. Materials and methods

2.1. Protein expression and purification

The subtype C near-full-length gag/pol clone was isolated from an HIV-positive patient from India and provided by the NIH Research and Reference Reagent Resources Program (clone p94IN476.104; Rodenburg *et al.*, 2001). This clone contained nine amino-acid differences (T12S, I15V, L19I, M36I, S37A, H69K, N88D, L89M and I93L) from the B PR sequence (LAI strain) and these NOPs are located outside the active-site cavity of the PR (Fig. 1). N88D is considered a major mutation for the clinically used PI NFV. The C PR analyzed here was obtained by back-mutating the aspartic acid at position 88 to asparagine and by introducing three other amino-acid changes (Q7K, L33I, L63I) known to block the self-cleavage process.

The recombinant C PR was subcloned as previously described (Clemente *et al.*, 2006; Goodenow *et al.*, 2002). The mutations blocking the autolysis sites were introduced using a Quick Change Site-directed Mutagenesis approach (Stratagene). The enzyme was expressed using the pET23a expression vector (Novagen) and transformed into the *Escherichia coli* expression cell line BL21 (DE3) Star pLysS (Invitrogen). Protein expression, inclusion-body isolation, protein refolding and purification were carried out as described previously (Clemente *et al.*, 2006).

2.2. Crystallization

The purified C PR was concentrated to 3.5 mg ml^{-1} using a 5 kDa VivaSpin 15R Concentrator (VivaScience) in 20 m*M* sodium acetate pH 4.5 with 2 m*M* dithiothreitol. Initial crystallization trials were conducted using the hanging-drop vapor-diffusion method at room temperature (McPherson, 1982). Crystal drops were prepared by mixing 2 µl enzyme solution with 2 µl reservoir solution, equilibrated by vapor diffusion against 1 ml reservoir solution at 293 K and screened using conditions from various crystallization kits (Hampton Research). Initial screening yielded microcrystals from various concentrations of sodium chloride as precipitant and citric acid and sodium citrate as buffers. Based on these results, useful X-ray diffraction-quality crystals of C PR were obtained by mixing 2 µl enzyme solution with 2 µl of reservoir solution consisting of 30 m*M* citric acid pH 5.0 and 1 *M* sodium chloride with Triton X-100 as an additive.

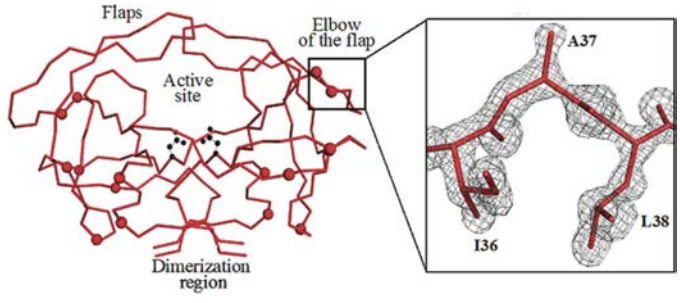


Figure 1
Structure of the unbound C PR. C^{α} tracing of C PR; the sites of naturally occurring polymorphisms (NOPs) are shown as red spheres and the catalytic aspartic residues in ball-and-stick representation. The inset shows a $2F_o - F_c$ electron-density map of residues 36–38. The map is contoured at 3σ. Figures were rendered with *PyMOL* (DeLano Scientific).

2.3. Data collection and reduction

Data were collected using a MAR CCD 225 detector at the SER-CAT beamline BM22 at the Advanced Photon Source, Argonne National Laboratory. The crystal-to-detector distance was 200 mm. The crystals were soaked in 35% glycerol solution and flash-cooled at 100 K. All diffraction data frames were collected using a 0.5° oscillation angle with an exposure time of 5 s per frame. The data set was indexed and scaled with *HKL*-2000 software (Otwinowski & Minor, 1997).

2.4. Rotation and translation search

Cross-rotation, translational searches and rigid-body refinement were performed using the *CNS* package (Brünger *et al.*, 1998). The unbound B PR (PDB code 1hhp; Spinelli *et al.*, 1991), with all solvent removed, was used as the molecular-replacement phasing template, using data between 8.0 and 4.0 Å resolution.

2.5. Refinement

Initial positional and *B*-factor refinement steps were done using the *CNS* suite (Brünger *et al.*, 1998) for data to 1.6 Å resolution. Further refinement was carried out using *SHELX* (Sheldrick, 2008) for all data to 1.2 Å resolution. 5% of the observed reflections were randomly selected and used to calculate $R_{\rm free}$ during the refinement process. Interactive manual model building was performed using the molecular-graphics program *O* (v.10.0.1 ; Jones *et al.*, 1991) with $2F_{\rm o} - F_{\rm c}$ and $F_{\rm o} - F_{\rm c}$ electron-density maps. In the later stages of *SHELX* refinement, H atoms were calculated in riding positions.

2.6. *TLSMD*

Anisotropic temperature factors from *SHELX* refinements were analyzed by the *TLSMD* method (Painter & Merritt, 2006*a*,*b*). A translation–libration–screw model was calculated for the monomer divided into segments, up to 15 for the polypeptide chain, and a least-squares residual was calculated. The residual is a measure of agreement between observed anisotropic temperature factors and those calculated from the translational and rotational displacements of the TLS model. From these calculations, a TLS model based upon seven segments was chosen.

The quality of the final refined structure was validated with *PROCHECK* (Laskowski *et al.*, 1993).

3. Results

3.1. Crystallization

Crystals of the unbound C PR appeared under the following conditions: 1 *M* NaCl, 30 m*M* citric acid pH 5.0 with Triton X-100 as an additive (Hampton Research). The initial inspection of the crystallization drop, 12 h after the equilibration against the precipitant solution at room temperature, revealed fine precipitation that accumulated over the next 48 h. On day 3 several diamond-shaped crystals appeared and increased in size over the following 5 d.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P4_12_12$
Unit-cell parameters (Å)	a = 46.7, c = 100.8
Unit-cell volume (Å^3)	219834
$V_{\rm M}$ (Å^3 Da^{-1})	2.56
Solvent fraction (%)	52
Total reflections	279351
Unique reflections	35611
Crystal mosaicity (°)	0.4
Resolution range (Å)	30.0–1.2 (1.24–1.2)
Completeness (%)	99.3 (95.7)
$R_{\rm merge}$† (%)	7.6 (27.3)
Redundancy	7.9 (4.7)
Average $I/\sigma(I)$	14.0
$I/\sigma(I) > 3$ (%)	76.0 (41.7)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl)\rangle$ is the average intensity

3.2. Diffraction data collection, processing and scaling

A total of 360 images were used in the data set. This resulted in 279 351 reflections measured to 1.2 Å resolution. The data were scaled in the Laue group 4/*mmm* with unit-cell parameters a = 46.7, c = 100.8 Å and merged and reduced to a set of 35 611 independent reflections (99.3% completeness, 95.7% in the outer resolution shell) resulting in a scaling $R_{\rm merge}$ of 0.076 (0.276 in the outer resolution shell) (Table 1). Using the unit-cell volume (2.2×10^5 Å) and the molecular weight of C PR (10 740 Da per monomer), a $V_{\rm M}$ value (Matthews, 1968) of ∼2.56 Å^3 Da^{-1} (52% solvent content) was calculated assuming eight monomers (four dimers) in the unit cell using *CNS* (Brünger *et al.*, 1998).

3.3. Molecular replacement: particle orientation and position

The PDB entry 1hhp (Spinelli *et al.*, 1991) of the unbound B PR structure was used as the molecular-replacement model to determine the C PR orientation and position in the tetragonal cell. The cross-rotation function search, using data between 8.0 and 4.0 Å resolution, provided a single (but weak) solution with the correlation coefficient of 0.078, with the next highest peak having a correlation coefficient of 0.068.

Using this orientation matrix, a translation function search in space group $P4_12_12$ with the B PR monomer gave a single peak with correlation coefficient of 0.616 and a packing value of 0.581. Translation-function searches for $P4_22_12$ and $P4_32_12$ were also performed, but no significant peaks were found.

3.4. Structure refinement and validation

The structure was initially refined with a cycle of rigid-body, individual *B*-factor and positional refinement using the *CNS* package (Brünger *et al.*, 1998). The resultant $R_{\rm work}$ was 39.4% at 2.5 Å resolution.

Initial $F_{\rm o} - F_{\rm c}$ and $2F_{\rm o} - F_{\rm c}$ electron-density maps were calculated and contoured at 3.0σ and 1.5σ, respectively. The

Table 2
Refinement statistics.

R_{final}†	14.0
R_{free}†	17.8
No. of water molecules	167
No. of glycerol molecules	2
R.m.s.d for bond lengths (Å)	0.014
R.m.s.d for angles (°)	0.030
Ramachandran statistics (%)	
Most favored regions	96.2
Allowed regions	3.8
Average *B* factors (Å^2)	
Main chain	13.5
Side chain	19.1
Waters	37.0
Glycerols	25.2
PDB code	2r8n

† $R_{final} = \sum(|F_o| - |Fc|)/\sum|F_o| \times 100$. R_{free} is identical to R_{final} for 5% of data that were omitted from refinement.

maps were of good quality and the main-chain electron density was continuous. The map also showed that the flap regions of the unbound B PR structure used for phasing did not fit within the electron-density map and rebuilding of the backbone and side chains of residues 47–54 was required. After several more cycles of refinement, mutating the side chains of B PR to C PR and adding water and two glycerol molecules, the C PR structure converged with an R_{work} of 21.0% and an R_{free} of 21.7% using data to 1.6 Å resolution.

The C PR structure was then further refined using the programs *SHELX* and *SHELXPRO* (Sheldrick, 2008) to 1.2 Å resolution. The water and two glycerol molecules placed in the model during the *CNS* refinement steps were not removed prior to input into *SHELX*. The first cycle of isotropic refinement resulted in an R_{work} of 23.8% and an R_{free} of 26.4%. After a further five cycles of refinement gradually improving the C PR model, including building dual side-chain conformers, the protein atoms were refined anisotropically and, with the removal of 50 poorly refined water molecules, the R_{work} and R_{free} were improved to 14.9% and 18.3%, respectively. The additional anisotropic refinement of all non-H atoms resulted in an R_{work} of 14.0% and an R_{free} of 17.8%. The final step of refinement in *SHELX* was performed using 100% of the data, including the 5% of the data reserved for R_{free} calculation, and yielded a final R factor of 14.1% (Table 2).

The quality of the refined structure of C PR was verified with the *PROCHECK* program (Laskowski *et al.*, 1993). 96.2% of the dihedral angles were located in the most favored regions, with all others in the additional allowed regions.

3.5. *TLSMD*

The *TLSMD* analysis of C PR indicated that even with a TLS model with up to 15 individual segments the least-squares residual continued to decrease. Therefore, as a compromise, the TLS model chosen consisted of seven segments, keeping the number of segments small while accepting a modestly good residual. In this model, amino-acid residues 42–53 of the flap region were covered by one TLS segment. These residues had their maximum translational displacement nearly normal to the crystallographic twofold axis, similar to the results of an unbound B PR with a similar flap conformation (Heaslet, Lin *et al.*, 2007). The rotational components of the screw model in this flap region were dominated by a single vector approximately along the β-strands of the flap with a rotational value of 23.2°. Mean isotropic atomic displacements calculated from the TLS model for the flap segment were 0.82 Å for the translational motion and ranged from approximately 0.25 to 0.65 Å for the rotational motion.

3.6. Structure analysis

The 1.2 Å resolution structure of the unbound form of C PR showed excellent electron density for all protein atoms, glycerol and water molecules. The protein crystallized as a homodimer, with one monomer in the crystallographic asymmetric unit (labeled resides 1–99), with the unbound substrate-binding site located between a crystallographic twofold dimer interface, covered by two extended polypeptide arms (residue 43–57), known as the flaps, one from each monomer (Fig. 1).

Some residual diffuse density, exhibiting a 'C-shape' appearance, was located between the open flaps of the $F_o - F_c$ electron-density maps and was interpreted as a string of poorly defined water molecules.

In this study the C PR structure was compared with three unbound B PRs: PDB entries 1hhp (Spinelli *et al.*, 1991), 2pc0 (Heaslet, Rosenfeld *et al.*, 2007) and 1rpi (Logsdon *et al.*, 2004) determined to 2.7, 1.4 and 1.8 Å resolution, respectively. The first two B PRs do not harbor any drug-resistance mutations, while 1rpi is a multi-drug-resistant variant harboring nine substitutions known to confer drug resistance to PIs (Fig. 2).

The average *B* factors for the C PR structure for the main-chain and the side-chain atoms were 13.5 and 19.1 Å^2, respectively. The solvent for the final model included 167 water and two glycerol molecules, with average *B* factors of 37.0 and 25.2 Å^2, respectively (Table 2). The distribution of the main-chain atom *B* factors for the unbound C PR showed maximal values for the C- and N-termini in the flap and elbow regions (34–51) at the tip of the 60s loop (65–70) and in the

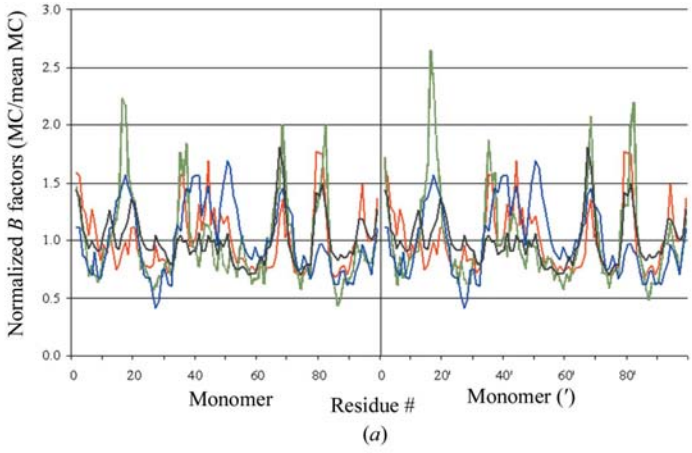
```
            10        20        30        40        50
            |         |         |         |         |
2R8N PQITLWKRPLVSIKVGGQIKEALLDTGADDTVIEEIALPGRWKPKMIGGI
1HHP PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
2PC0 PQITLWKRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPKMIGGI
1RPI PQITLWQRPIVTIKIGGQLKEALLNTGADDTVLEEVNLPGRWKPKLIGGI

            60        70        80        90
            |         |         |         |
2R8N GGFIKVRQYDQIIIEICGKKAIGTVLVGPTPVDIIGRNMLTQLGCTLNF
1HHP GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
2PC0 GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
1RPI GGFVKVRQYDQVPIEICGHKVIGTVLVGPTPANVIGRNLMTQIGCTLNF
```

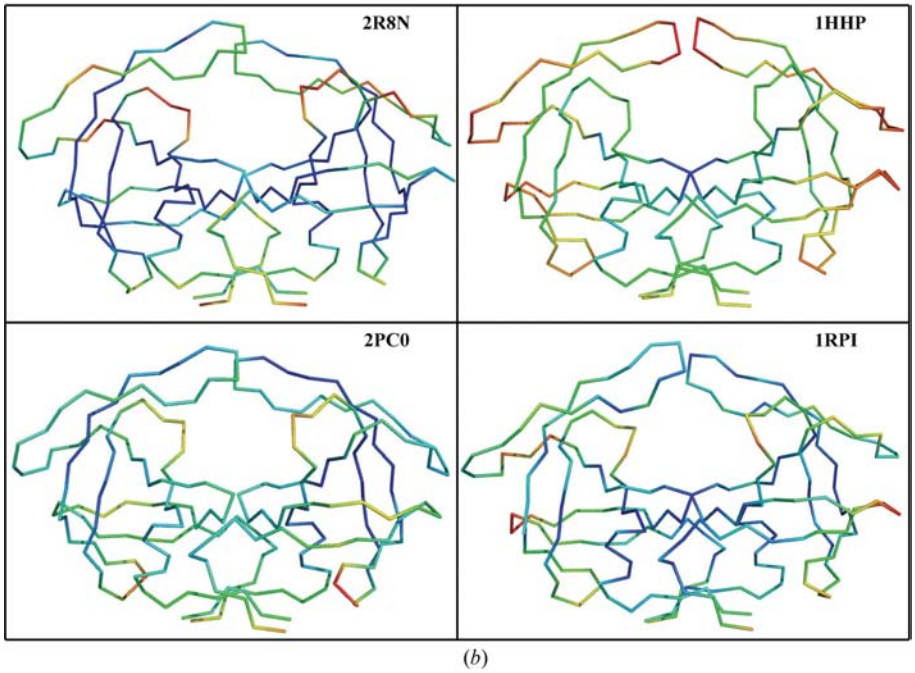
Figure 2
Sequence alignment of C PR (PDB code 2r8n) and three B PRs (PDB codes 1hhp, 2pc0 and 1rpi; Spinelli *et al.*, 1991; Heaslet, Rosenfeld *et al.*, 2007; Logsdon *et al.*, 2004). Amino-acid differences are highlighted in green.

active site (78–84) (Fig. 3*a*). Comparison of the B PR structures showed similar *B*-factor profiles, but there were also significant differences in that 1hhp had decreased *B* factors for the active-site residues 78–84, and 2pc0 and 1rpi had low *B* factors in the flap and elbow regions, residues 34–51. The *B* factors for the flap regions in C PR show intermediate values when compared with the B PR structures.

Also of interest was the significant decrease in *B* factors for C PR in the 12–20 β-sheet region, where three of the NOPs are located. This thermal stability was also observed for 2pc0 and partially for 1hhp, but was not seen in the 1rpi structure.



(*a*)



(*b*)

Figure 3
Thermal parameters for the C PR (PDB code 2r8n) and 1hhp (Spinelli *et al.*, 1991), 2pc0 (Heaslet, Rosenfeld *et al.*, 2007) and 1rpi (Logsdon *et al.*, 2004) B PRs. (*a*) Plot of the normalized main-chain atom mean *B* factors. Normalized *B* factors were obtained by dividing the mean *B* factor for the main-chain atoms of each residue by the average *B* factors for the all main-chain PR atoms. The residues in the two subunits are labeled 1–99 and 1′–99′. The color code is C PR, red; B PR 1hhp, blue; B PR 2pc0, black; B PR 1rpi, green. (*b*) C^{α} rainbow histogram tracing of the *B* factor for the main-chain atoms for the C and B PRs. The color range covers, in equal-sized *B*-factor increments, from the lowest (dark blue) to the highest (red) *B* actors. This figure was rendered with *PyMOL* (DeLano Scientific).

Because of the high resolution and quality of the C PR structure, analysis of the anisotropy (defined as the ratio between the minimum and maximum eigenvalues of the matrix of anisotropic displacement parameters; Merritt, 1999) of the structure was performed. This was of interest as this can provide insight into the direction of the motion of the flap regions known to have different conformations in different PR structures. This analysis indicated that the flaps have a general lateral motion relative to the active-site cleft (Fig. 4) and this motion is accompanied, to a lesser extent, by analogous movements in the elbow of the flap (data not shown). These observations were similar to those observed from the *TLSMD* calculations.

High-resolution structural information also allows better interpretation of the structural disorder, including amino-acid side chains that exhibit alternate conformations (Esposito *et al.*, 2000). In the C PR structure, several residues exhibited side chains with alternate conformations (Glu21, Glu34, Glu35, Pro44, Arg57, Lys69 and Val82). All of these residues are located on the outer loops of the enzyme, in contact with the solvent.

A least-squares superimposition of the unbound form of C PR and the three B PR structures was performed (Fig. 5*a*). The 1hhp, known as the 'semi-open' form, and the 2pc0 B PR structures were both crystallized in space group $P4_12_12$ with unit-cell parameters similar to those of the C PR structure, while the 1rpi B PR crystallized in space group $P4_1$. The r.m.s. deviations (r.m.s.d.) for C^{α} atoms between 1hhp, 2pc0 and 1rpi B PR structures were 1.09, 0.28 and 0.65 Å compared with C PR, respectively (Fig. 5*b*), whereas the catalytic triplet residues 25–27 located in the active site showed relatively very low r.m.s.d. values of 0.14, 0.21 and 0.35 Å, respectively, which is consistent with the highly conserved core structure. However, an interesting observation was that the main-chain atoms of the active-site residue Val82, which is in close proximity to the catalytic triad, showed an r.m.s.d. of 0.8–1.40 Å. The highest r.m.s.d difference observed was between the flaps of the PRs (as detailed below).

The relatively high r.m.s.d. between 1hhp B PR and C PR emphasizes several differences of more than 1.0 Å deviation. These included residues 35–42 (the elbow of the flaps), 49–53 (the

tip of the flaps), 63–70 (the 60s loop) and 80–81. The highest r.m.s.d.s were located in the outer loops, especially in the loops harboring the NOPs such as the elbow of the flap and the 60s loop. The most striking difference between the C and 1hhp PRs is a significant conformational change in the flap orientation, with the closest distance between the tips of the flaps of 1hhp B PR being 4.4 Å, while for C PR this value is 12.2 Å (Fig. 6*a*). This would imply a displacement of more than 8 Å of the flaps of the C PR relative to the 1hhp B PR structure. Interestingly, the PR C structure showed a similar distance between the flaps as that of the 2pc0 and 1rpi B PRs, which have an opening of the flaps of 12.2 Å (Fig. 6*b*) and 12.5 Å (Fig. 6*c*).

M36I is one of the NOPs occurring in C PR that is considered to be a secondary drug-resistance mutation in B PR (Patick *et al.*, 1998). Analyzing the PRs, it was observed that the side chain of the smaller Ile residue has approximately twofold less van der Waals interactions than observed in 1hhp and 2pc0 B PRs harboring the Met residue (Fig. 2). It should also be noted that there was a downward shift in the C PR 36–41 loop with an average displacement of 1.6 and 1.3 Å when compared with the 1hhp and 1rpi PRs (Fig. 7*a*). No displacement was noticed when compared with 2pc0.

There are three other NOPs that occur with high frequency in C PR: H69K, L89M and L93I (Fig. 2). H69K is located close to the base of the PR, within a loop that exhibited an r.m.s.d. of 1.7, 0.63 and 1.0 Å when C PR is compared with 1hhp, 2pc0 and 1rpi B PRs, respectively. The L89M polymorphism occurs in more than 95% of the subtype C strains (Stanford University HIV Drug Resistance Database; http://hivdb.stanford.edu/); this residue is located in the core of the PR and participates in an extensive network of hydrophobic interactions. There is a twofold increase in the number of interactions in C PR when compared with B PRs. The I93L polymorphism is located in close spatial vicinity of position 69 and in both B and C PRs makes numerous interactions with the surrounding residues (Fig. 7*b*).

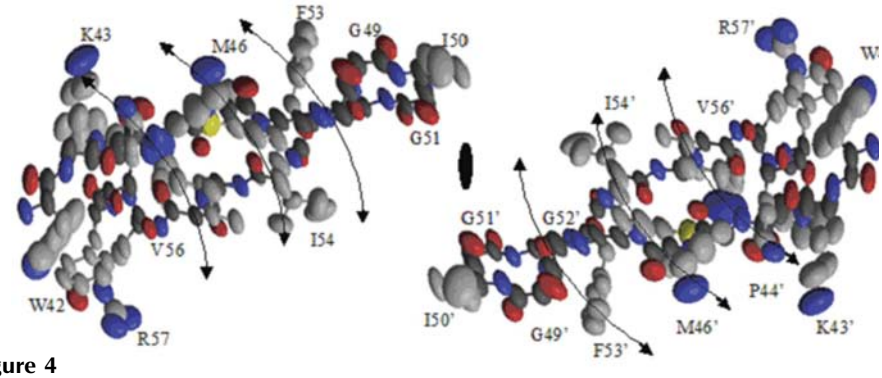


Figure 4
Thermal ellipsoid diagram for the unbound C PR. The diagram represents the anisotropy for residues 42–59 within the flaps, including the side chains. View looking down the C PR homodimer active site. C, O, N and S atoms are colored gray, red, blue and yellow, respectively. The arrows indicate the resultant direction of the flap motion. This figure was created using *RASTEP* (Merritt, 1999; Merritt & Bacon, 1997).

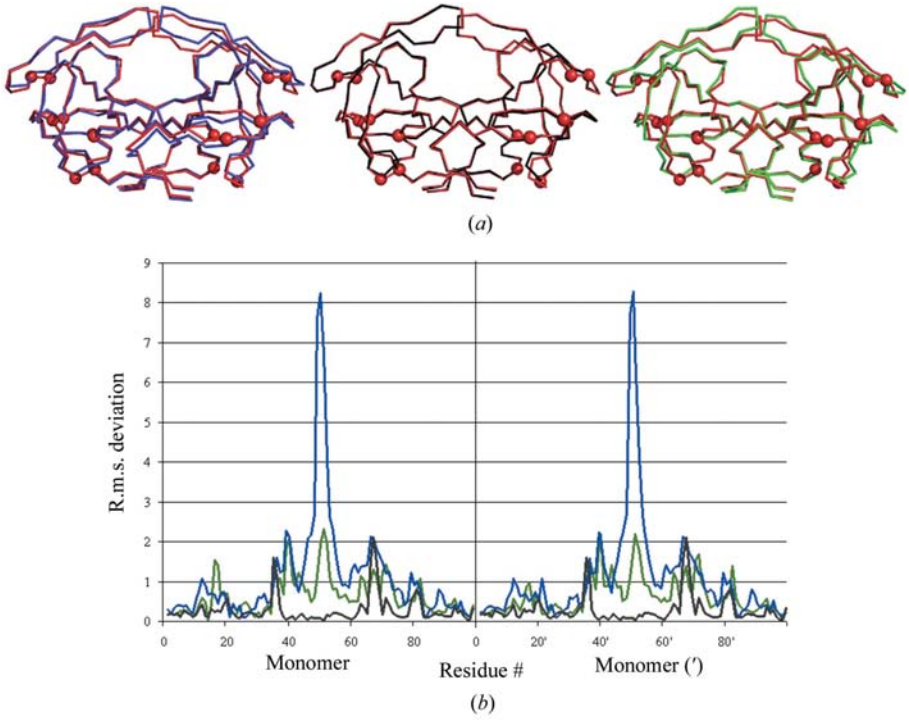


Figure 5
Superimposition of the C PR with the 1hhp (Spinelli *et al.*, 1991), 2pc0 and 1rpi (Logsdon *et al.*, 2004) B PRs. (*a*) C^{α} tracing of C PR (red) superimposed on 1hhp (blue/left), 2pc0 (black/middle) and 1rpi (green/right) B PRs. The NOPs in C PR are represented as red spheres. (*b*) The r.m.s.d. (Å) per residue plotted for the C^{α} atoms of 1hhp (blue), 2pc0 (black) and 1rpi (green) B PRs compared with the C PR. The residues in the two subunits are labeled 1–99 and 1′–99′. This figure was rendered with *PyMOL* (DeLano Scientific).

4. Discussion

Crystals of the unbound C PR have been grown and its structure determined to 1.2 Å resolution, representing the first structure of the unbound C PR and the highest resolution solved structure of a non-B PR reported to date.

This structure is of interest because many *in vivo* and *in vitro* studies (Clemente *et al.*, 2006; Gonzalez *et al.*, 2006; Kantor & Katzenstein, 2003; Peeters, 2001; Sanches *et al.*, 2007; Tanuri *et al.*, 1999; Velazquez-Campoy, Vega *et al.*, 2003) have advanced the hypothesis that the NOPs within the PR

play a role in modulating antiretroviral drug susceptibility with the possibility of faster development of drug resistance during therapy. In this study, the structure of the unbound form of C PR is compared with three unbound B PRs in an attempt to understand the structural effects arising from the NOPs and their implications in antiretroviral drug resistance/susceptibility.

One of the PR regions believed to be involved in modulating the affinity of the PR for inhibitors is the flap domain. Understanding the factors underlining the PR flap mobility has profound implications in elucidating the detailed mechanism of substrate/inhibitor binding of this enzyme and in the design of new therapeutic agents such as allosteric inhibitors intended to interfere with the flap opening and thereby with enzymatic function. The mechanisms and the factors involved in coordinating and modulating the motion of the flaps have been the focus of study for many researchers. Several studies have shown that the flaps open upward and laterally (Ishima *et al.*, 1999; Nicholson *et al.*, 1995; Toth & Borics, 2006; Wlodawer & Erickson, 1993), while others argued that the tip of the flaps curl inside, making hydrophobic contacts with several residues located in the active site (Heaslet, Lin *et al.*, 2007; Scott & Schiffer, 2000). It is generally agreed that the large motion of the tip of the flap is accompanied by changes in the hinge as well as the elbow of the flap (Clemente *et al.*, 2004; Perryman *et al.*, 2006). Several NMR and molecular-dynamics studies investigated the conversion between closed, semi-open and fully open forms of HIV-1 PR flaps. These conformations appear to be in dynamic equilibrium, with the semi-open form being the most prevalent (Freedberg *et al.*, 2002; Hamelberg & McCammon, 2005; Hornak *et al.*, 2006*b*; Nicholson *et al.*, 1995). In the C PR structure the magnitude of the atomic motion in the flaps does not appear to be significantly higher than the core of the enzyme. This is also the case for the 1rpi, a multi-drug-resistant B PR, but is significantly not so for the 1hhp B PR. These data, correlated with relatively low *B* factors for the flaps (Figs. 3*a* and *b*), argue for a limitation of the flap movements, probably owing to crystal contacts as has been proposed by Hornak *et al.* (2006*a*). Among the crystal contacts involved in holding the flap open are hydrogen bonding of the carbonyl of Gly49 to the side-chain amino group of Arg41′ and hydrophobic interaction of Ile50 with Pro81′. The 2pc0 and 1rpi B PR structures contain about 100 and 130 water molecules, respectively, in the active-site cavity. The refined 1hhp B PR structure has no assigned water molecules, which is most likely to be a consequence of the poor resolution. Martin *et al.* (2005) proposed that these water molecules form a scaffold in the active-site cavity, preventing the PR from collapsing in the absence of a ligand, as observed in the C PR structure (Table 2).

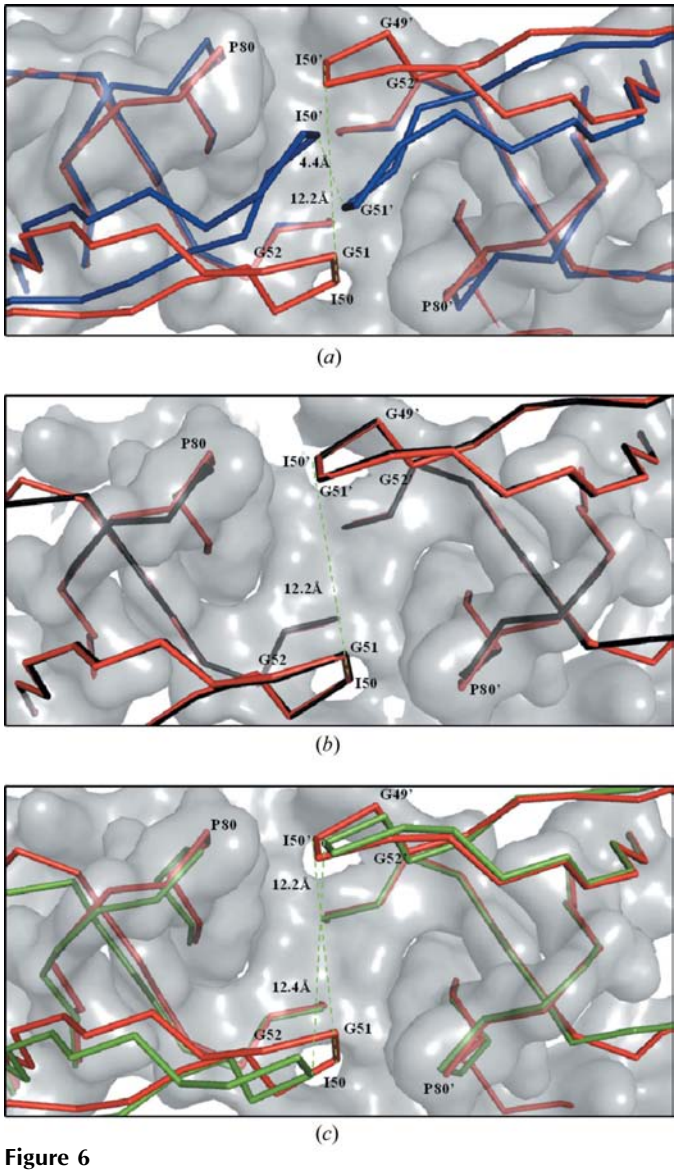


Figure 6
Comparison between the flaps of the C PR (PDB code 2r8n) and 11hhp (Spinelli *et al.*, 1991), 2pc0 (Heaslet, Rosenfeld *et al.*, 2007) and 1rpi (Logsdon *et al.*, 2004) B PRs. C^{α} tracing of the flap regions of (*a*) the unbound C PR (red) superimposed on 1hhp (blue), (*b*) the unbound C PR (red) superimposed on 2pc0 (black) and (*c*) the unbound C PR (red) superimposed on 1rpi (green). The gray surface represents the active site. This figure was rendered with *PyMOL* (DeLano Scientific).

Among the B PR structures used for comparison in this study, the 2pc0 structure crystallized in space group $P4_12_12$ and has widely open flaps, similar to the C PR structure. This finding argues for crystal contacts having a prominent role in propping open the flaps when unbound HIV PRs crystallize in space groups $P4_1$ or $P4_12_12$. However, it could be that the unbound PR prefers the open conformation in solution and the prevalence of this form induces the enzyme to crystallize in space groups $P4_1$ or $P4_12_12$; consequently the crystal contacts are formed owing to the open form of the PR and are not solely the cause of the flaps staying open.

Other regions of interest that could further elucidate the changes in the flaps during binding/release of the substrate/inhibitor are the hinge and the elbow of the flaps. When superimposing the unbound C PR with the three B PR structures, several interesting differences are observed. Position 36 occupies a region in the PR that is highly mobile during flap opening and closing in the course of ligand binding. It has been argued that the M36I mutation may promote long-range

structural changes in the active site or changes in the flexibility of the PR, which may lead to either the closed or the open conformation of the PR being dominant (Clemente *et al.*, 2004). As observed in the C PR structure and several other previous studies (Clemente *et al.*, 2004; Martin *et al.*, 2005), Met36 makes extensive interactions with residues located in the 10s and 60s loops. In C PR the bulkier Met is exchanged for a smaller Ile and consequently there is a decrease in the van der Waals interactions between the 10s and 60s loops. This effect is augmented by the I15V polymorphisms in which a smaller Val replaces the Ile residue. The overall effect is a decreased number of interactions between these two loops in C PR, allowing an increased stability and decreased flexibility of the elbow of the flap. Previous studies of B and F PRs have also argued that this polymorphic change causes a collapse of the elbow of the flap, resulting in displacement of the main-chain of this loop toward the loop 76–83, stabilizing the catalytic S1/S1′ pockets (Sanches *et al.*, 2007). A similar effect has been noted in the C PR structure, where the catalytic residues Pro81 and Val82 are shifted towards the interior of the active site.

Analyzing these data and comparing the unbound C PR and unbound B PRs structures, it appears that the flap and the elbow are a functional unit and that the changes or the motion in the elbow of the flap are transmitted to the tip of the flap and *vice versa*. Consequently, the differences observed between C and B PRs at the elbow could arise from a different orientation at the tip of the flap and not from NOPs.

The amino-acid residue at position 89 is located in the hydrophobic core of the PR and when mutated to a Met makes extensive hydrophobic contacts with neighboring residues, to a greater extent than in B PR, which harbors a Leu at this position. Variation in the number of hydrophobic residues appears to be important for both maintaining the structural stability of the enzyme and allowing conformational changes. It has been hypothesized that the hydrophobic core residues slide by each other, exchanging one hydrophobic van der Waals contact for another, with little energy penalty, while maintaining many structurally important hydrogen bonds (Foulkes-Murzycki *et al.*, 2007). Such hydrophobic sliding may represent a general mechanism by which proteins undergo conformational changes. Consequently, mutation of these residues in the PR would alter the packing of the hydrophobic core, affecting the conformational flexibility of the enzyme. It has been proposed that these residues impact on the dynamic balance between processing substrates and binding inhibitors and thus any change in this region could contribute to drug resistance/susceptibility (Foulkes-Murzycki *et al.*, 2007). The increased number of van der Waals interactions in the presence of L89M polymorphism might increase the stability of the C PR and affect the dynamic properties of the PR and potentially affect its ability to bind inhibitors and substrates. Furthermore, a previous study hypothesized that Met89 was assumed to mimic the role of the L90M mutation by displacing Asp25 and thus constraining the S1/S1′ pockets (Sanches *et al.*, 2007). In the C PR structure it was observed that L89M can also displace the 60s loop laterally and downwards and could also participate in formation of a more stable network of van der Waals interactions.

The C PR harbors three signature residues, Ser12, Val15 and Ile19, which are located in a β-sheet that forms what is called the 10s loop. The influence of the NOPs in this region has not been widely studied, but there are two interesting observations from this study. Firstly, as mentioned above, the polymorphic change from a larger Ile15 to a smaller Val15 in C PR further reduces the number of interactions between the 10s loop and the elbow of the flap, in this way changing the dynamics of the elbow of the flap. Secondly, analysis of the *B* factors (Fig. 3) showed that there is a significant difference between the main-chain *B* factors of the 10–22 residues between B and C PRs. In the C PR the 10s loop had *B* factors just below the average value, while the 1rpi PR exhibited an ∼2.5-fold increase in the *B* factors in this region. This finding leads to the conclusion that the 10s loop is more ordered and probably less flexible in C PR. A similar effect, but of a lesser magnitude, happens for the 60s loop. All of these data taken together, the increased hydrophobic contacts owing to L89M polymorphisms and decreased stability of the 10s and 60s loops, could suggest that in C PR there is an increased stability at the base of the PR. The large number of van der Waals interactions forms a scaffold on which the flaps could swing open more easily with fewer energetic requirements. Also, this arrangement could change the size of the active site to an extent where, upon addition of major drug-resistance

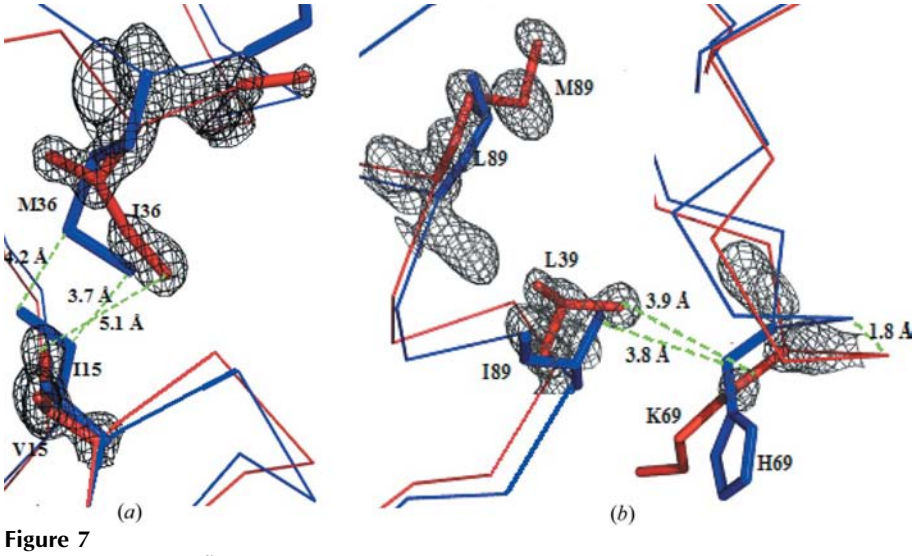


Figure 7
The NOPs in C PR. C^{α} tracing of (*a*) the flap and elbow regions of the C PR (red) superimposed on 1hhp B PR (blue) and (*b*) the 60s loop and residues 89 and 93 of the C PR (red) superimposed on 1hhp B PR (blue). The amino-acid residues are represented as sticks. This figure was rendered with *PyMOL* (DeLano Scientific).

mutations, the inhibitor binding is hindered, while at the same time maintaining a reasonable affinity for the more flexible substrate.

The role of NOPs might be that they stabilize the core of the PR while maintaining the flexibility of the flaps, promoting the open flexible conformation of C PR. Since inhibitors are rigid and are designed to bind the closed conformation, they would preferentially bind to enzymes that carry mutations that favor the closed conformation (Clemente *et al.*, 2004). Consequently, this open conformation of the C PR would be less favorable for inhibitor binding. These results correlate with our recently published analysis on the contribution of NOPs on altering the biochemical and structural properties of several drug-resistant variants of subtype C PR (Coman *et al.*, 2007) as well as with other previous structural and kinetic studies (Sanches *et al.*, 2007) that showed that the NOPs in F PR might amplify the effect of drug-resistance mutations.

This structural study revealed several differences between B and C PRs. Even though crystallography offers a static exploration of a structure, it still allows several inferences to be made about the dynamics of the flaps. These results could add to the general effort in explaining if and how the NOPs contribute to the mechanism through which C PR could gain resistance to PIs.

These data and subsequent studies with other PIs will greatly aid in our efforts to understand the influence of NOPs in modulating the enzyme sensitivity and resistance to current drug-therapy regimens and hopefully provide new insight into designing novel inhibitors that are less likely to promote the development of PR drug-resistance mutations.

The authors thank Dr Mavis Agbandje-McKenna, Dr Lakshmanan Govindasamy and John Domsic for useful discussion and technical help with data collection and analysis. The authors also thank Dr Alexander Wlodawer, Dr Zbigniew Dauter and the SER-CAT beamline staff at the Advanced Photon Source X-ray facility for their assistance in data collection, and the Center of Structural Biology at the University of Florida for its support of the in-house X-ray facility where preliminary diffraction data were obtained.

References

Brodine, S. K., Mascola, J. R., Weiss, P. J., Ito, S. I., Porter, K. R., Artenstein, A. W., Garland, F. C., McCutchan, F. E. & Burke, D. S. (1995). *Lancet*, **346**, 1198–1199.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Clemente, J. C., Coman, R. M., Thiaville, M. M., Janka, L. K., Jeung, J. A., Nukoolkarn, S., Govindasamy, L., Agbandje-McKenna, M., McKenna, R., Leelamanit, W., Goodenow, M. M. & Dunn, B. M. (2006). *Biochemistry*, **45**, 5468–5477.

Clemente, J. C., Moose, R. E., Hemrajani, R., Whitford, L. R., Govindasamy, L., Reutzel, R., McKenna, R., Agbandje-McKenna, M., Goodenow, M. M. & Dunn, B. M. (2004). *Biochemistry*, **43**, 12141–12151.

Coman, R. M., Robbins, A., Goodenow, M. M., McKenna, R. & Dunn, B. M. (2007). *Acta Cryst.* F**63**, 320–323.

Dumans, A. T., Soares, M. A., Machado, E. S., Hue, S., Brindeiro, R. M., Pillay, D. & Tanuri, A. (2004). *J. Infect. Dis.* **189**, 1232–1238.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.

Fleury, H., Recordon-Pinson, P., Caumont, A., Faure, M., Roques, P., Plantier, J. C., Couturier, E., Dormont, D., Masquelier, B. & Simon, F. (2003). *AIDS Res. Hum. Retroviruses*, **19**, 41–47.

Foulkes-Murzycki, J. E., Scott, W. R. & Schiffer, C. A. (2007). *Structure*, **15**, 225–233.

Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y. X., Kustanovich, I., Louis, J. M. & Torchia, D. A. (2002). *Protein Sci.* **11**, 221–232.

Gonzalez, L. M., Aguiar, R. S., Afonso, A., Brindeiro, P. A., Arruda, M. B., Soares, M. A., Brindeiro, R. M. & Tanuri, A. (2006). *J. Gen. Virol.* **87**, 1303–1309.

Goodenow, M. M., Bloom, G., Rose, S. L., Pomeroy, S. M., O'Brien, P. O., Perez, E. E., Sleasman, J. W. & Dunn, B. M. (2002). *Virology*, **292**, 137–149.

Grossman, Z., Paxinos, E. E., Averbuch, D., Maayan, S., Parkin, N. T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C. J. & Schapiro, J. M. (2004). *Antimicrob. Agents Chemother.* **48**, 2159–2165.

Hamelberg, D. & McCammon, J. A. (2005). *J. Am. Chem. Soc.* **127**, 13778–13779.

Heaslet, H., Lin, Y. C., Tam, K., Torbett, B. E., Elder, J. H. & Stout, C. D. (2007). *Retrovirology*, **4**, 1.

Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E. & Stout, C. D. (2007). *Acta Cryst.* D**63**, 866–875.

Hirsch, M. S., Brun-Vezinet, F., D'Aquila, R. T., Hammer, S. M., Johnson, V. A., Kuritzkes, D. R., Loveday, C., Mellors, J. W., Clotet, B., Conway, B., Demeter, L. M., Vella, S., Jacobsen, D. M. & Richman, D. D. (2000). *J. Am. Med. Assoc.* **283**, 2417–2426.

Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. (2006*a*). *J. Am. Chem. Soc.* **128**, 2812–2813.

Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. (2006*b*). *Proc. Natl Acad. Sci. USA*, **103**, 915–920.

Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M. & Torchia, D. A. (1999). *Structure*, **7**, 1047–1055.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kantor, R. & Katzenstein, D. (2003). *AIDS Rev.* **5**, 25–35.

Kantor, R. & Katzenstein, D. (2004). *J. Clin. Virol.* **29**, 152–159.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Logsdon, B. C., Vickrey, J. F., Martin, P., Proteasa, G., Koepke, J. I., Terlecky, S. R., Wawrzak, Z., Winters, M. A., Merigan, T. C. & Kovari, L. C. (2004). *J. Virol.* **78**, 3123–3132.

Louis, J. M., Ishima, R., Torchia, D. A. & Weber, I. T. (2007). *Adv. Pharmacol.* **55**, 261–298.

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: Wiley and Son.

Martin, P., Vickrey, J. F., Proteasa, G., Jimenez, Y. L., Wawrzak, Z., Winters, M. A., Merigan, T. C. & Kovari, L. C. (2005). *Structure*, **13**, 1887–1895.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1997–2004.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **227**, 505–524.

Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Lam, P. Y., Jadhav, P. K., Hodge, N., Domaille, P. & Chang, C.-H. (1995). *Nature Struct. Biol.* **2**, 274–280.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.

Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.

Patick, A. K., Duran, M., Cao, Y., Shugarts, D., Keller, M. R., Mazabel, E., Knowles, M., Chapman, S., Kuritzkes, D. R. & Markowitz, M. (1998). *Antimicrob. Agents Chemother.* **42**, 2637–2644.

Peeters, M. (2001). *Transfus. Clin. Biol.* **8**, 222–225.
Perryman, A. L., Lin, J. H. & McCammon, J. A. (2006). *Biopolymers*, **82**, 272–284.
Rodenburg, C. M., Li, Y., Trask, S. A., Chen, Y., Decker, J., Robertson, D. L., Kalish, M. L., Shaw, G. M., Allen, S., Hahn, B. H. & Gao, F. (2001). *AIDS Res. Hum. Retroviruses*, **17**, 161–168.
Rose, R. E., Gong, Y. F., Greytok, J. A., Bechtold, C. M., Terry, B. J., Robinson, B. S., Alam, M., Colonno, R. J. & Lin, P. F. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 1648–1653.
Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A. & Polikarpov, I. (2007). *J. Mol. Biol.* **369**, 1029–1040.
Sanches, M., Martins, N. H., Calazans, A., de Moraes Brindeiro, R., Tanuri, A., Antunes, O. A. C. & Polikarpov, I. (2004). *Acta Cryst.* D**60**, 1625–1627.
Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
Scott, W. R. & Schiffer, C. A. (2000). *Structure*, **8**, 1259–1265.
Servais, J., Lambert, C., Fontaine, E., Plesseria, J. M., Robert, I., Arendt, V., Staub, T., Schneider, F., Hemmer, R., Burtonboy, G. & Schmit, J. C. (2001). *Antimicrob. Agents Chemother.* **45**, 893–900.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Spaltenstein, A., Kazmierski, W. M., Miller, J. F. & Samano, V. (2005). *Curr. Top. Med. Chem.* **5**, 1589–1607.
Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. (1991). *Biochimie*, **73**, 1391–1396.
Tanuri, A., Vicente, A. C., Otsuki, K., Ramos, C. A., Ferreira, O. C. Jr, Schechter, M., Janini, L. M., Pieniazek, D. & Rayfield, M. A. (1999). *Antimicrob. Agents Chemother.* **43**, 253–258.
Toth, G. & Borics, A. (2006). *J. Mol. Graph. Model.* **24**, 465–474.
Velazquez-Campoy, A., Muzammil, S., Ohtaka, H., Schon, A., Vega, S. & Freire, E. (2003). *Curr. Drug Targets Infect. Disord.* **3**, 311–328.
Velazquez-Campoy, A., Todd, M. J., Vega, S. & Freire, E. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6062–6067.
Velazquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H. W. & Freire, E. (2003). *AIDS Rev.* **5**, 165–171.
Vergne, L., Peeters, M., Mpoudi-Ngole, E., Bourgeois, A., Liegeois, F., Toure-Kane, C., Mboup, S., Mulanga-Kabeya, C., Saman, E., Jourdan, J., Reynes, J. & Delaporte, E. (2000). *J. Clin. Microbiol.* **38**, 3919–3925.
Vondrasek, J. & Wlodawer, A. (2002). *Proteins*, **49**, 429–431.
Wlodawer, A. & Erickson, J. W. (1993). *Annu. Rev. Biochem.* **62**, 543–585.