Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, purification and preliminary X-ray crystallographic studies of the human immunodeficiency virus 1 subtype C protease

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

[a]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32610, USA, and [b]Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL 32610, USA

Correspondence e-mail: bdunn@ufl.edu

Received 20 December 2006
Accepted 12 March 2007

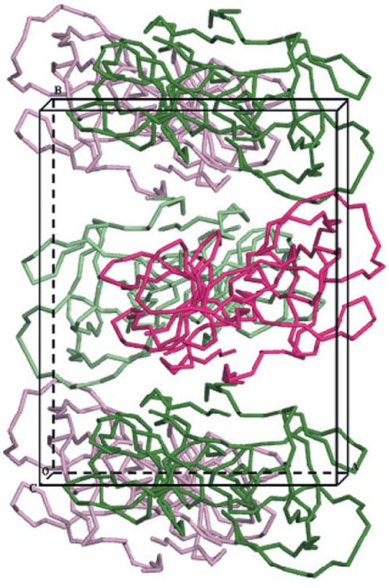
Crystals of the human immunodeficiency virus 1 (HIV-1) subtype C protease (PR) complexed with the clinically used inhibitors indinavir (IDV) and nelfinavir (NFV) have been grown in the monoclinic space group $P2_1$, with mean unit-cell parameters a = 46.7 (±0.1), b = 59.8 (±0.3), c = 87.0 (±0.4) Å, β = 95.2 (±0.5)°. The crystals of both complexes have been shown to diffract X-rays to 2.3 Å resolution. The diffraction data for the subtype C PR complexes with IDV and NFV were subsequently processed and reduced, with overall R_{sym} values of 8.4 and 11.4%, respectively. Based on the unit-cell volumes, molecular-replacement results and packing considerations, there are two protease homodimers per crystallographic asymmetric unit in each of the complexes. The data were initially phased using a model based on the crystal structure of HIV-1 subtype B PR; the structures have been determined and further refinement and analysis are in progress. These structures and subsequent studies with other inhibitors will greatly aid in correlating the amino-acid variation between the different HIV PRs and understanding their differential sensitivity and resistance to current drug therapy.

1. Introduction

The HIV/AIDS epidemic continues to spread unchecked since its first documentation in 1981 (Gottlieb *et al.*, 1983; Groopman & Gottlieb, 1983; Karpas, 2004). 25 years later, nearly 45 million individuals are living with HIV/AIDS worldwide, according to the figures released by the Joint United Nations Program on HIV/AIDS and the World Health Organization (CDC, 2006; UNAIDS/WHO, 2004). The most devastated developing regions are in sub-Saharan Africa, with more than 25 million adults and children living with HIV/AIDS (CDC, 2006; Katzenstein, 2006; UNAIDS/WHO, 2004).

HIV is characterized by significant genetic diversity among distinct types, groups and clades (Fleury *et al.*, 2003; Kantor & Katzenstein, 2004) and this variability has implications in prevention, diagnostic tests, therapy response and vaccine development (Kantor & Katzenstein, 2004; Peeters & Sharp, 2000; Romano *et al.*, 2000). HIV-1 subtype C is one of the nine HIV-1 subtypes and it is noteworthy that subtype C causes 42% of all HIV infections worldwide and is now the dominant subtype in sub-Saharan Africa (McCormack *et al.*, 2002; Osmanov *et al.*, 2002). Recent studies (Walker *et al.*, 2005) have demonstrated that the growth rate of subtype C infections in southern Africa shows no evidence of declining. In addition, the C epidemic has also spread to South and Central China, India and Brazil (Soares *et al.*, 2003; Yu *et al.*, 1998). This global epidemiological profile, the invasive behavior of subtype C infection and the overall picture in which the majority of HIV infection appears to be due to subtype C provide de facto evidence for transmissibility. Thus, it is more important than ever to improve and tailor prevention and treatment and ultimately to develop vaccine strategies effective against the most prevalent HIV-1 subtypes (Katzenstein, 2006).

Highly active antiretroviral therapy, HAART, is a combination of three or more drugs from two different drug classes and has significantly improved the prognosis of HIV-infected individuals (Carpenter *et al.*, 1997; Collier, 1996). Many of the therapeutic strategies used to treat HIV-infected patients are designed based on the knowledge of subtype B viruses. Numerous studies (Abecasis *et al.*, 2005; Cane *et al.*, 2001; Caride *et al.*, 2001; Grossman *et al.*, 2004;



Kantor *et al.*, 2005) have shown that globally non-B subtype viruses seem to be susceptible to currently used protease inhibitors (PIs), but caution is advised to be taken when designing therapeutic regimens against these strains as differences in drug-resistance evolution and pathways among various subtypes are continuing to emerge. Novel combinations of mutations seem to emerge as a consequence of the extraordinary variability of HIV subtypes and the wider use of antiretroviral regimens.

One of the targets of HAART is the HIV protease (PR), an enzyme that is essential to the viral life cycle. The HIV-1 PR is a homodimer that consists of two identical 99-residue subunits (Wlodawer *et al.*, 1989), with the active site located directly above the dimer interface and containing two catalytic aspartic residues. Extensive research has been dedicated to designing resistance-evading drugs for PR, which is critical for the maturation of viral structural (gag) and enzymatic (pol) proteins. Crystal structures of HIV-1 PR were first reported in 1989 (Miller *et al.*, 1989; Wlodawer *et al.*, 1989) and their availability has played a major role in the process of drug development (Clemente *et al.*, 2006; Mahalingam *et al.*, 2004; Vondrasek & Wlodawer, 2002).

Many crystal structures of the PR and its complexes with inhibitors are currently available, but there is no structural information on non-subtype-B PRs. The first crystallization of a non-subtype-B PR was reported recently by Sanches and coworkers, but at the time of submission of this paper the structure had not been published (Sanches *et al.*, 2004). Kinetic (Velazquez-Campoy *et al.*, 2001), thermodynamic (Todd *et al.*, 2000; Velazquez-Campoy *et al.*, 2000) and modeling (Batista *et al.*, 2006) studies show that non-subtype-B PRs exhibit reduced affinity for protease inhibitors (PIs). If these affinity differences alone are not sufficient to cause drug resistance, they can intensify the effects of mutations and eventually lead to drug resistance. Furthermore, new combinations of mutations have been shown to arise in subtype C PR upon treatment with HAART regimens containing at least one PI (Doualla-Bell *et al.*, 2006).

Solving the crystal structure of HIV-1 subtype C PR will contribute to a better understanding of enzyme–inhibitor interactions. The structure will provide insight into how the enzyme interacts with clinically used PIs for each individual subsite at the atomic level, especially for those residues that are presumably critical for substrate or inhibitor binding. Thus, the structural information obtained from the crystallographic analysis can not only verify the data observed in kinetic and functional studies, but can also provide structural reference for inhibitor design. Furthermore, it will reveal instructive information about the effect of baseline polymorphisms on the overall structure of the PR and will help in identifying the long-range interactions through which these non-active-site mutations can affect the binding of substrates/inhibitors in the active site.

2. Materials and methods

2.1. Expression and purification

The HIV-1 subtype C clone was isolated from an HIV-positive patient from India and was provided by the NIH Research and Reference Reagent Resources Program (Rodenburg *et al.*, 2001). This clone contained nine amino-acid differences (T12S, I15V, L19I, M36I, S37A, H69K, N88D, L89M and I93L) from the subtype B PR sequence (LAI strain; Fig. 1*a*) and these polymorphisms are located in the outside regions of the PR (Fig. 1*b*). One of these amino-acid differences, N88D, is considered to be a major mutation for resistance to the clinically used PR inhibitor nelfinavir. The subtype C PR analyzed here was obtained by back-mutating the aspartic acid at the position 88 to the wild-type residue asparagine and by introducing three changes known to block the self-cleavage process (Q7K, L33I, L63I).

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

2.2. Crystallization

HIV-1 subtype C PR was concentrated to 3.2 mg ml^{-1} using a 5 kDa VivaSpin 15R Concentrator (VivaScience) and incubated with indinavir (IDV) and nelfinavir (NFV) (obtained from the NIH Research and Reference Reagent Resources Program), two of the clinically used PIs, in a threefold molar excess for 60 min at 277 K in 20 m*M* sodium acetate pH 4.5, 2 m*M* DDT prior to crystallization. Initial crystallization trials were conducted using the hanging-drop vapor-diffusion method (McPherson, 1982), screening various conditions from Crystal Screen 1 (Hampton Research). Crystal drops were prepared by mixing 2 µl enzyme solution with 2 µl reservoir solution and were equilibrated by vapor diffusion against 1 ml reservoir solution at 293 K. Based on the results of the crystallization screening, useful X-ray diffraction-quality crystals of HIV-1 subtype C PR were obtained by mixing 2 µl enzyme solution with 2 µl reservoir solution consisting of 30 m*M* citric acid pH 5.0 and 1 *M* sodium chloride.

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

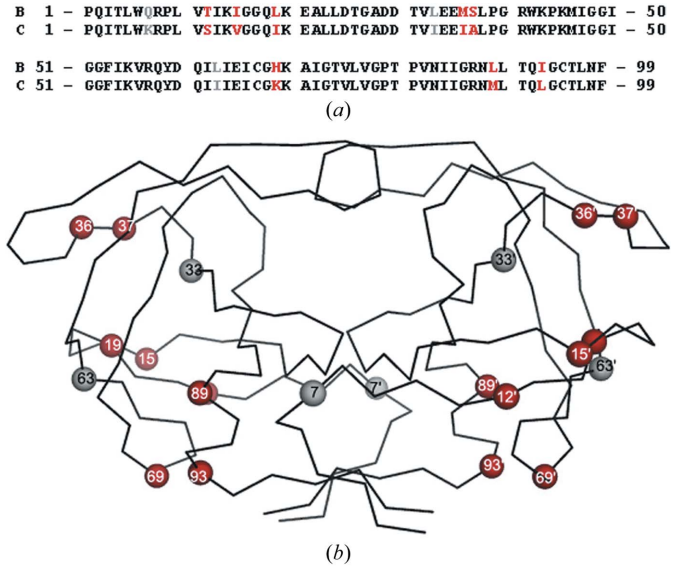


Figure 1
(*a*) Pairwise sequence alignment of HIV-1 subtype B PR LAI strain (designated B) and subtype C PR (designated C). (*b*) Cartoon representation of HIV-1 subtype B and C PRs. The naturally occurring polymorphisms in subtype C PR, represented as red spheres, are superimposed onto the crystal structure of HIV-1 subtype B PR (black ribbon). The mutations that block the self-cleavage sites are shown in gray. Amino-acid positions are as numbered.

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 sets were indexed and scaled with the *HKL*-2000 software (Otwinowski & Minor, 1997).

2.4. Sequence alignment and model building

A pairwise sequence alignment of HIV-1 subtype C and subtype B PRs was performed using *ClustalW* (Expasy Proteomics Server; Fig. 1*a*). Based on the previously determined three-dimensional crystal structure of the HIV-1 subtype B PR (PDB code 2bpx) and the *ClustalW* sequence alignment, a model of subtype C PR was built using the *PyMOL* molecular-graphics system (Fig. 1*b*).

2.5. Rotation and translation search

Cross-rotation and translational searches were performed using the program *AMoRe* (Navaza, 2001) as implemented in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) and rigid-body refinement using the *CNS* package (Brünger *et al.*, 1998). The molecular template used was a polyalanine HIV-1 subtype B homodimer (PDB code 1w5y), with the inhibitor removed from the active site. X-ray data in the resolution range 8.0 and 4.0 Å were used for these searches.

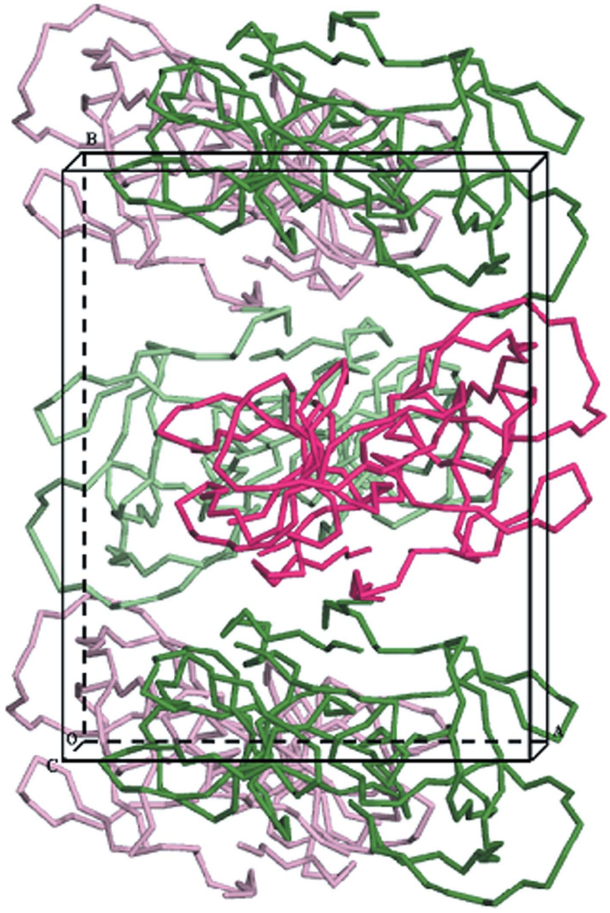


Figure 2
HIV-1 subtype C PR ribbon packing diagram in the $P2_1$ crystal lattice viewed down the *c* axis. There are four subtype C PR homodimers in the unit cell (shown in red and light red and in green and light green). The box depicts the unit cell. The model shown is based on data from subtype C PR complexed with IDV and is the same for the NFV complex.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

	Subtype C protease with indinavir (IDV)	Subtype C protease with nelfinavir (NFV)
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 46.7, b = 59.5, c = 87.4, β = 95.6	a = 46.7, b = 60.1, c = 86.7, β = 94.7
Unit-cell volume (Å^3)	241915	242491
V_M (Å^3 Da^{-1})	2.9	2.9
Solvent content (%)	55	55
Total reflections	65956	61466
Unique reflections	21311	21028
Crystal mosaicity (°)	0.9	0.9
Resolution range (Å)	20.0–2.3 (2.38–2.3)	20.0–2.3 (2.38–2.3)
Completeness (%)	96.7 (96.4)	97.3 (93.7)
R_{sym}† (%)	8.4 (33.2)	11.3 (37.3)
Redundancy	3.1 (3.0)	2.9 (2.6)
Average $I/\sigma(I)$	8.4	5.6
$I/\sigma(I) > 3$ (%)	71.1 (36.8)	59.7 (21.0)

† $R_{sym} = \sum |I - \langle I \rangle| / \sum I$, where I is the intensity of an individual reflection and $\langle I \rangle$ is the average intensity.

3. Results and discussion

3.1. Crystallization

Crystals suitable for X-ray diffraction studies appeared approximately 24 h after equilibration against the precipitant solution. The crystals grew as thin rectangular plates stacked on top of each other, with approximate dimensions of 0.4 × 0.3 × 0.1 mm. Variations on the crystallization conditions are currently being explored, changing the pH and precipitant concentration of the reservoir solution, in an attempt to improve both the crystal size and the diffraction resolution. In addition, crystallization of HIV-1 subtype C PR complexed with several other clinically used inhibitors is ongoing.

3.2. Data collection, processing and scaling

A total of 180° of data were collected (360 images) from single crystals of subtype C PR complexed with IDV and NVF and a total of 65 956 and 61 466 reflections were measured, respectively. Complete data sets were collected to 2.3 Å resolution.

The data for the subtype C PR complexed with IDV were initially processed in the Laue crystal system *P*2. The reduced data set resulted in an R_{sym} of 0.084 (0.332 in the outer resolution shell). A similar data set for the subtype C PR complexed with NFV resulted in an R_{sym} of 0.113 (0.373 in the outer resolution shell). Table 1 gives a full summary of the data-collection statistics.

On inspection of the intensities of the *h*00, 0*k*0 and 00*l* reflections, the existence of a twofold screw axis along the *b* axis could be inferred, therefore implying that the space group was $P2_1$ for both crystals. From the unit-cell volume and the molecular weight of subtype C PR complexed with IDV and with NFV, V_M values (Matthews, 1968) of ∼2.9 Å^3 Da^{-1} (56% solvent content) for both crystals were calculated using *CNS* v.1.1 (Brünger *et al.*, 1998), assuming the presence of two homodimers per asymmetric unit.

3.3. Molecular replacement: particle orientation and position

Cross-rotation function searches with the subtype C PR–IDV complex data in Laue group *P*2 using the 1w5y polypeptide model provided four solutions, two of which reflected the molecular twofold

rotation axis of the model. The four unique peaks were approximately twice the values of the next highest peaks in the rotation function.

A translation-function search using the subtype C PR–IDV complex data in space group $P2_1$ for the two noncrystallographic dimer rotation solutions provided peaks with correlation coefficients of 0.380 and 0.287, and R values of 0.473 and 0.518, respectively. A second translation search was required to define the relative separation between the two dimers along the y direction by constraining the position of the A dimer. This solution had a correlation coefficient and an R value of 0.675 and 0.358, respectively. The fractional translations were $T_x = 0.4238$, $T_y = 0.0000$ and $T_z = 0.0486$ for site A, and $T_x = -0.4222$, $T_y = 0.0245$ and $T_z = -0.4449$ for site B. Rigid-body refinement of the resulting two dimers, using data in the resolution range 15–3.5 Å, converged at $R_{\rm free} = 0.303$ and $R_{\rm work} = 0.321$. The solution for the subtype C PR–NFV complex data used the orientations and positions as defined from the subtype C PR–IDV complex data. Initial $F_{\rm o} - F_{\rm c}$ electron-density maps for both complexes, contoured at 3.0σ, showed clear interpretable density for the inhibitors IDV and NFV present in both dimers. Fig. 2 shows the packing arrangement for the two homodimers in the $P2_1$ space group for the subtype C PR–IDV complex.

Currently, the structures of the HIV-1 subtype C PR–IDV and PR–NFV complexes are being refined and are in the process of being interpreted. These findings will be compared with the kinetic data that have also been determined for various mutants of both subtype B and C PRs. These data and subsequent studies with other inhibitors will greatly aid in our efforts to understand the influence of baseline polymorphisms in modulating the enzyme sensitivity and resistance to current drug therapy.

The authors would like to thank NIAID for grant R37 AI028571 and the Center of Structural Biology (CSB) at the University of Florida (especially Dr Tom Mareci and Mavis Agbandje-McKenna) for support of this work. They would also like to thank Dr Lakshmanan Govindasamy and Dr Zoë Fisher for helpful discussions. Use of the SER-CAT beamline at the Advanced Photon Source X-ray facility was supported by the US Department of Energy, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. The authors would also like to thank Dr Zbigniew Dauter for his help with data collection and initial data analysis.

References

Abecasis, A. B., Deforche, K., Snoeck, J., Bacheler, L. T., McKenna, P., Carvalho, A. P., Gomes, P., Camacho, R. J. & Vandamme, A. M. (2005). *AIDS*, **19**, 1799–1806.

Batista, P. R., Wilter, A., Durham, E. H. & Pascutti, P. G. (2006). *Cell Biochem. Biophys.* **44**, 395–404.

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.

Cane, P. A., de Ruiter, A., Rice, P., Wiselka, M., Fox, R. & Pillay, D. (2001). *J. Clin. Microbiol.* **39**, 2652–2654.

Caride, E., Hertogs, K., Larder, B., Dehertogh, P., Brindeiro, R., Machado, E., de Sa, C. A., Eyer-Silva, W. A., Sion, F. S., Passioni, L. F., Menezes, J. A., Calazans, A. R. & Tanuri, A. (2001). *Virus Genes*, **23**, 193–202.

Carpenter, C. C., Fischl, M. A., Hammer, S. M., Hirsch, M. S., Jacobsen, D. M., Katzenstein, D. A., Montaner, J. S., Richman, D. D., Saag, M. S., Schooley, R. T., Thompson, M. A., Vella, S., Yeni, P. G. & Volberding, P. A. (1997). *J. Am. Med. Assoc.* **277**, 1962–1969.

CDC (2006). *Ann. Pharmacother.* **40**, 1708.

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.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Collier, A. C. (1996). *Adv. Exp. Med. Biol.* **394**, 355–372.

Doualla-Bell, F., Avalos, A., Gaolathe, T., Mine, M., Gaseitsiwe, S., Ndwapi, N., Novitsky, V. A., Brenner, B., Oliveira, M., Moisi, D., Moffat, H., Thior, I., Essex, M. & Wainberg, M. A. (2006). *Antimicrob. Agents Chemother.* **50**, 2210–2213.

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.

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.

Gottlieb, M. S., Groopman, J. E., Weinstein, W. M., Fahey, J. L. & Detels, R. (1983). *Ann. Intern. Med.* **99**, 208–220.

Groopman, J. E. & Gottlieb, M. S. (1983). *Nature (London)*, **303**, 575–576.

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.

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

Kantor, R. *et al.* (2005). *PLoS Med.* **2**, e112.

Karpas, A. (2004). *Biol. Rev. Camb. Philos. Soc.* **79**, 911–933.

Katzenstein, D. (2006). *J. Infect. Dis.* **194**, *Suppl.* 1, S45–S50.

McCormack, G. P., Glynn, J. R., Crampin, A. C., Sibande, F., Mulawa, D., Bliss, L., Broadbent, P., Abarca, K., Ponnighaus, J. M., Fine, P. E. & Clewley, J. P. (2002). *J. Virol.* **76**, 12890–12899.

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

Mahalingam, B., Wang, Y. F., Boross, P. I., Tozser, J., Louis, J. M., Harrison, R. W. & Weber, I. T. (2004). *Eur. J. Biochem.* **271**, 1516–1524.

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

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

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B. & Esparza, J. (2002). *J. Acquir. Immune Defic. Syndr.* **29**, 184–190.

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

Peeters, M. & Sharp, P. M. (2000). *AIDS*, **14**, *Suppl.* 3, S129–S140.

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.

Romano, L., Venturi, G., Ferruzzi, R., Riccio, M. L., Corsi, P., Leoncini, F., Vinattieri, A., Incandela, L., Valensin, P. E. & Zazzi, M. (2000). *AIDS*, **14**, 2204–2206.

Sanches, M., Martins, N. H., Calazans, A., Brindeiro R. de M., Tanuri, A., Antunes, O. A. & Polikarpov, I. (2004). *Acta Cryst.* D**60**, 1625–1627.

Soares, M. A., De Oliveira, T., Brindeiro, R. M., Diaz, R. S., Sabino, E. C., Brigido, L., Pires, I. L., Morgado, M. G., Dantas, M. C., Barreira, D., Teixeira, P. R., Cassol, S. & Tanuri, A. (2003). *AIDS*, **17**, 11–21.

Todd, M. J., Luque, I., Velazquez-Campoy, A. & Freire, E. (2000). *Biochemistry*, **39**, 11876–11883.

UNAIDS/WHO (2004). *AIDS Epidemic Update.* http://whqlibdoc.who.int/unaids/2004/9291733903.pdf.

Velazquez-Campoy, A., Todd, M. J. & Freire, E. (2000). *Biochemistry*, **39**, 2201–2207.

Velazquez-Campoy, A., Todd, M. J., Vega, S. & Freire, E. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6062–6067.

Vondrasek, J. & Wlodawer, A. (2002). *Proteins*, **49**, 429–431.

Walker, P. R., Pybus, O. G., Rambaut, A. & Holmes, E. C. (2005). *Infect. Genet. Evol.* **5**, 199–208.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. (1989). *Science*, **245**, 616–21.

Yu, X. F., Chen, J., Shao, Y., Beyrer, C. & Lai, S. (1998). *Lancet*, **351**, 1250.