

Structural Analysis of Human Immunodeficiency Virus Type 1 CRF01_AE Protease in Complex with the Substrate p1-p6[∇]

Rajintha M. Bandaranayake,¹ Moses Prabu-Jeyabalan,¹ Junko Kakizawa,²
Wataru Sugiura,² and Celia A. Schiffer^{1*}

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605,¹ and Laboratory of Therapeutic Research and Clinical Science, AIDS Research Center, National Institute of Infectious Diseases, 4-7-1 Gakuen Musashimurayama, Tokyo 208-0011, Japan²

Received 3 January 2008/Accepted 16 April 2008

The effect of amino acid variability between human immunodeficiency virus type 1 (HIV-1) clades on structure and the emergence of resistance mutations in HIV-1 protease has become an area of significant interest in recent years. We determined the first crystal structure of the HIV-1 CRF01_AE protease in complex with the p1-p6 substrate to a resolution of 2.8 Å. Hydrogen bonding between the flap hinge and the protease core regions shows significant structural rearrangements in CRF01_AE protease compared to the clade B protease structure.

Based on its genomic diversity, the human immunodeficiency virus type 1 (HIV-1) has been classified into three groups, M (major), N (nonmajor), and O (other/outlier) (16). Group M has been further defined into nine clades (clades A to D, F to H, and J and K) and a number of subclades and circulating recombinant forms (CRFs). HIV-1 protease is one of the major proteins targeted for anti-HIV drug development. The *pol* gene, which codes for protease, differs by 10 to 15% between clades (7), and sequence diversity within HIV-1 clades has been an important area of study in recent years due to its possible role in altering resistance pathways within the protease (1, 10). In particular, the HIV-1 CRF01_AE protease acquires nelfinavir resistance via an alternative mutational pathway (1), making the detailed study of non-B proteases strongly warranted.

Structural studies of clade B protease have led to the successful development of a number of protease inhibitors (PIs). However, the majority of HIV-1 infection cases in the world result from non-clade B variants, and there is limited evidence that non-clade B variants respond differently to currently available PIs (3, 23). Although a large number of clade B protease structures have been solved over the years, to date, very little structural information is available for non-B HIV proteases. The first non-clade B protease structures for clade F were published recently by Sanches et al. (18), and the crystallization of clade C PI complexes has been reported by Coman et al. (4). We present here the crystal structure of an inactive HIV-1 CRF01_AE protease variant (D25N) in complex with a decameric peptide corresponding to the p1-p6 cleavage site within the Gag and Gag-Pro-Pol polyproteins. CRF01_AE was one of the first CRFs to be identified and is now the predominant HIV-1 variant in Southeast Asia (12). The protease was de-

rived from a Japanese patient isolate and has 10 amino acid substitutions (R14K, K20R, E35D, M36I, R41K, P63L, V64I, H69K, L89M, and I93L) compared to that of clade B (Fig. 1A and B).

Crystallization and structure determination. The CRF01_AE protease was expressed and purified as previously described (14). The protein was concentrated to 1.8 mg ml⁻¹ using a 10-kDa molecular size limit Amicon Ultra-15 centrifugal filter device. The decameric p1-p6 peptide (Arg-Pro-Gly-Asn-Phe-Leu-Gln-Ser-Arg-Pro; Quality Controlled Biochemicals, Inc., Hopkinton, MA) was solubilized in dimethyl sulfoxide and equilibrated with the protein with a fivefold molar excess for 1 h on ice. Crystals were grown over a reservoir solution consisting of 126 mM phosphate buffer at pH 6.2 and 63 mM sodium citrate and ammonium sulfate in the range of 18 to 33% (20). A 2:1 volume ratio of reservoir solution and substrate-protein solution were combined to set up hanging drops with a final volume of 6 μ l. The crystals were grown at ambient temperature.

Crystallographic data were collected under cryogenic conditions using an R-AXIS IV image plate mounted on a Rigaku rotating anode X-ray generator. The data were reduced and scaled using the programs DENZO and SCALEPACK, respectively (13). Structure determination and refinement were carried out using programs within the CCP4 software suite as previously described (15). Model building was carried out, followed by real space refinement with the COOT molecular graphics software (5). Refinement of the initial models was done without the p1-p6 substrate, and the peptide was built into the $F_o - F_c$ density within the active site as the refinement progressed. A truncated p1-p6 peptide lacking ArgP5 and ProP4 was modeled into the active site, as the $2F_o - F_c$ and $F_o - F_c$ maps indicated weak and discontinuous electron density at the N terminus of the peptide. The ArgP4' of the p1-p6 peptide was modeled in as alanine, since the electron density was not well defined to model in the arginine side chain. The stereochemical parameters of the final model were checked using PROCHECK (11). The CRF01_AE protease in complex

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605. Phone: (508) 856-8008. Fax: (508) 856-6464. E-mail: Celia.Schiffer@umassmed.edu.

[∇] Published ahead of print on 23 April 2008.

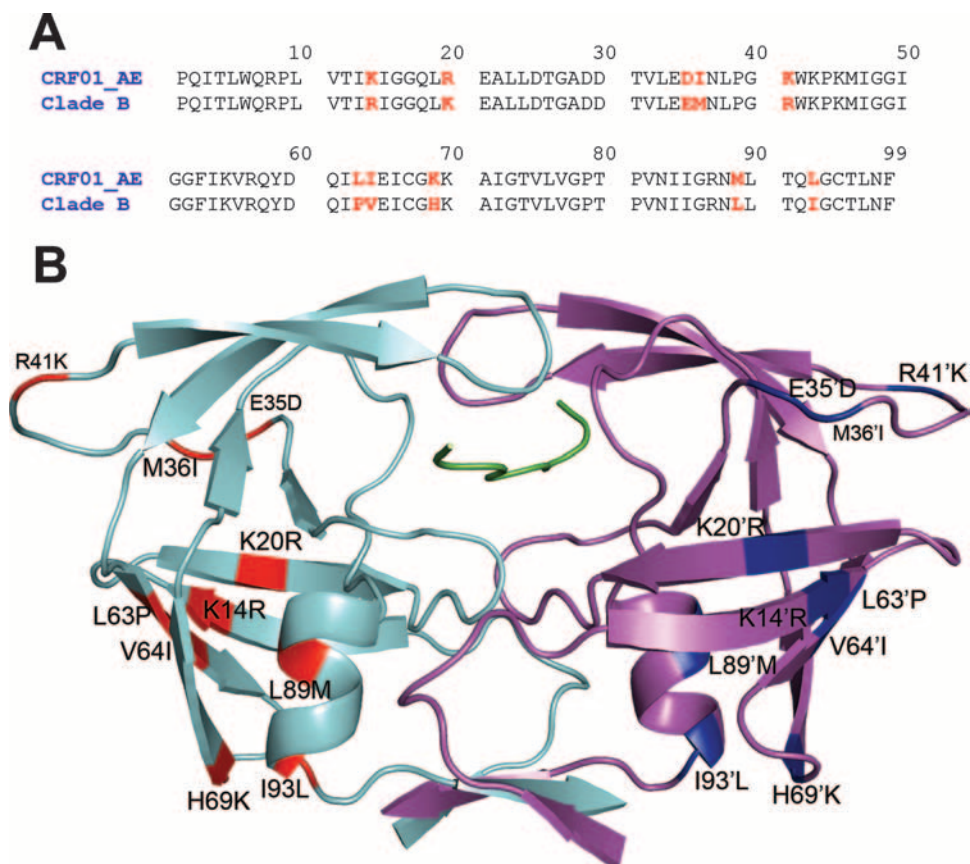


FIG. 1. (A) Amino acid sequence alignment of the CRF01_AE protease with the clade B protease. Positions where sequences differ are indicated in red. (B) CRF01_AE protease in complex with p1-p6 (green). Amino acid changes in monomer A (cyan) are indicated in red, and changes in monomer B (magenta) are indicated in blue.

TABLE 1. Crystallographic data and statistics for CRF01_AE in complex with substrate p1-p6

Parameter	Value
Resolution (Å)	2.8
Temperature (°C)	Cryogenic
Space group	P6 ₁
Cell dimensions	
a = b (Å)	62.1
c (Å)	82.1
Z ^a	6
R _{merge} (%)	9.5
Completeness (%)	99.4
Total no. of reflections	45,495
No. of unique reflections	4,523
I/σ ¹	8.4
RMSD in ^b	
Bond lengths (Å)	0.008
Bond angles (Å)	1.2
R _{factor} (%)	19.9
R _{free} (%)	25.8

^a Number of molecules in the unit cell.

^b RMSD, root-mean-square deviation.

with p1-p6 was determined to a resolution of 2.8 Å (PDB code 3D3T) (Table 1).

Protease structure comparison. The clade B D25N protease in complex with p1-p6 (PDB code 1KJF) was used for structural comparisons. The terminal regions (residues 1 to 9 and 86 to 99) from both monomers were used to superimpose the clade B structure onto the CRF01_AE complex. The superimposition was performed in a way that preserved the orientation of the substrate peptide between the two structures. A double-difference plot was generated to visualize structural differences between the two complexes. Distances between all the C α atoms within the dimer were calculated for each complex, and then the difference of the difference between the two dimers was plotted as a contour plot, as previously described (15). The presence of significant contour peaks within the plot indicates regions that differ between the two structures.

Based on the C α superimposition, the CRF01_AE and clade B structures displayed a high level of structural similarity to each other, with a root-mean-square deviation of 0.37 Å (Fig. 2A). However, peaks within the double-difference plot show that the CRF01_AE complex has significant structural rearrangements at the flap hinge region (residues 33 to 39) and near the protease core region (residues 16 to 22) (Fig. 2B). These structural differences are present in both monomers of the complex. Closer examination of Ile36 shows that its shorter

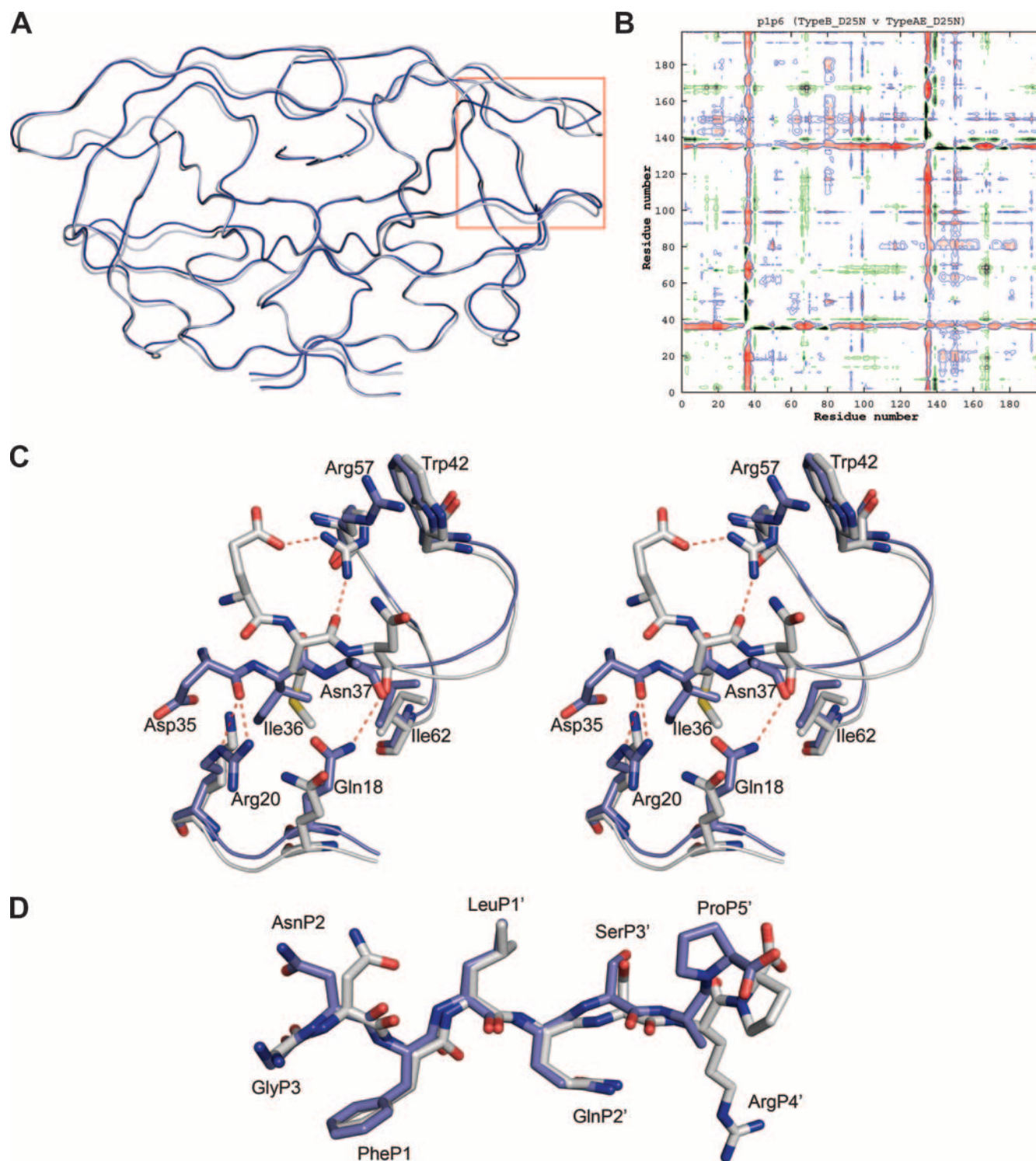


FIG. 2. (A) Ribbon diagram superposition of the CRF01_AE (blue) and the clade B (gray) p1-p6 structures. The flap hinge region is indicated by the red box. (B) Double-difference plot comparing the CRF01_AE and clade B p1-p6 structures. Contours in the plot represent the degrees of distance between the residues of the structures being compared. Black indicates a distance of <0.6 Å; red indicates -0.59 Å and -0.3 Å; blue indicates 0.3 Å and 0.59 Å; and yellow indicates >0.6 Å. (C) Stereoview of the rearrangement of the flap hinge region of the CRF01_AE structure (blue) compared to that of the clade B structure (gray). The Asn37 side chain in the CRF01_AE structure is disordered and has been modeled as alanine. (D) p1-p6 substrate conformation (blue, CRF01_AE; gray, clade B).

TABLE 2. Substrate-protease hydrogen bonds

Substrate atom	Protease atom	Distance (Å) between ^a :	
		CRF01_AE protease	Clade B protease
GlyP3 N	Gly48 O	3.5	3.0
GlyP3 O	Asp29 N	3.0	2.8
AsnP2 OD-1	Asp29 N	2.8	
AsnP2 OD-1	Asp30 N	2.8	
AsnP2 ND-2	Asp30 OD2	3.1	
PheP1 N	Gly27 O	2.6	2.9
PheP1 O	Asn25 ND2	2.8	2.7
GlnP2' N	Gly27' O	2.9	3.0
GlnP2' OE-1	Asp29' N	3.1	2.9
GlnP2' OE-1	Asp30' N	2.7	2.7
GlnP2' NE-2	Asp30' OD2	2.6	3.0
GlnP2' O	Asp29' O	2.8	3.1
Gln P2' OE-1	Asp30' OD2	3.4	3.3
SerP3' OG	Asp29' OD2	2.8	3.4
SerP3' OG	Arg8 NH1	3.1	
SerP3' N	Gly48' O	3.1	3.0
SerP3' O	Gly48' N	3.2	2.8
ArgP4' NE	Asp30' OD1	NA ^b	3.2
ArgP4' NE	Asp30' OD2	NA ^b	3.0
ArgP4' NH-2	Gln58' OE1	NA ^b	3.4

^a Distances highlighted in bold are hydrogen bonds that are observed only in the CRF01_AE structure.

^b Distances are not available (NA) as the ArgP4' side chain is disordered.

side chain is stabilized through van der Waals interactions with the side chains of Asn18, Leu38, and Arg20, allowing the flap hinge region to pack closer to the core region when compared to the longer Met36 side chain in the clade B structure (Fig. 2C). The collapse of the flap hinge toward the core is further enhanced by the formation of a hydrogen bond between the carbonyl oxygen of Asp35 and the NE or NH-2 of the Arg20 side chain. This interaction causes the Asp35 side chain to flip inward toward the core region. In comparison, the longer Glu35 side chain in clade B is flipped outward into the solvent, allowing its OE-2 oxygen to form a hydrogen bond with the side chain NH-1 of Arg57. The positioning of the Arg57 side chain also allows the NH-2 hydrogen to form a hydrogen bond with the carbonyl oxygen of Met36. The Arg57 side chain of the CRF01_AE structure is not involved in making any interactions with the flap hinge and packs against Trp42. These observations suggest that the flap hinge region of the CRF01_AE protease is likely to have reduced flexibility as a result of its tighter packing against the protease core region than against clade B.

Substrate conformation. The p1-p6 peptide is bound within the active site in an extended conformation with the Phe-Leu cleavage site at positions P1 and P1', oriented between the "catalytic" Asn25 residues. AsnP2 in the CRF01_AE structure adopts a conformation different from that of the clade B structure (Fig. 2D). OG of SerP3' also adopts an orientation that is different from that seen in the clade B structure. ProP5' is rotated by 180°, which causes the C terminus of the peptide to kink toward the P4' position, whereas in the clade B complex, the p1-p6 peptide adopts an extended conformation at the C terminus. Despite changes in peptide conformation, the protease-substrate hydrogen bonding patterns show a high degree of similarity between the two structures, with 13 substrate-protease hydrogen bonds conserved between the CRF01_AE

and the clade B structures (Table 2). However, the CRF01_AE structure makes four additional substrate-protease hydrogen bonds that are not seen in the clade B structure. The AsnP2 side chain conformation allows OD-1 to form a hydrogen bond with either Asp29 N or Asp30 N, while ND-2 forms a hydrogen bond with Asp30 OD-2. Furthermore, the AsnP2 side chain conformation allows it to make significant van der Waals interactions with Asp29 and Asp30. The hydrogen bond formed between SerP3' OG and Arg8 NH₁ is a result of SerP3' OG adopting an orientation that is different from that seen in the clade B structure. Thus, compared to the clade B structure, the p1-p6 substrate appears to form better interactions with the CRF01_AE active site.

Conclusions. The structure described in this study is the first CRF01_AE protease structure, as well as the first non-B HIV-1 protease-substrate complex structure, to be reported to date. The R20, D35, I36, K69, M89, and L93 seen in the structure have been implicated as resistance-associated mutations in clade B protease (2, 8, 9). While no significant structural changes were observed at K69, M89, and L93, the R20, D35, and I36 substitutions in the CRF01_AE protease resulted in significant structural rearrangements of the flap hinge and core regions compared to that in the clade B structure. We have observed a similar structural rearrangement in a CRF01_AE protease structure in an inhibitor complex (unpublished data), which might be an indication that the interactions observed are unique to the CRF01_AE protease.

Movement of the flaps is essential for substrate binding, and the flap hinge and core regions play key roles in flap dynamics (17, 19, 21). The close packing observed between these regions in the CRF01_AE protease structure is likely to restrict flexibility and thereby affect flap dynamics. The protease molecule itself undergoes large conformational changes in order to facilitate substrate binding and product release following substrate cleavage (6). Thus, reduced flexibility resulting from the packing of the flap hinge and core regions may have an effect on protease activity as well. Previously reported enzyme kinetics data for a CRF01_AE variant indicate that the active site specificity and catalytic efficiency are slightly lower for the CRF01_AE protease than for the clade B protease (3). Furthermore, polymorphisms occurring within these regions are thought to affect binding affinities for PIs (22). Therefore, the structural changes observed may influence how the CRF01_AE protease interacts with PIs and may thereby alter levels of resistance to currently available inhibitors compared to that of clade B protease.

We thank Madhavi Nalam and Balaji Bhyravbatla for assistance with structural refinement.

This work was supported by National Institutes of Health grant 2R01-GM064347-06.

REFERENCES

- Ariyoshi, K., M. Matsuda, H. Miura, S. Tateishi, K. Yamada, and W. Sugiura. 2003. Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. *J. Acquir. Immune Defic. Syndr.* 33:336-342.
- Becker-Pergola, G., P. Kataaha, L. Johnston-Dow, S. Fung, J. B. Jackson, and S. H. Eshleman. 2000. Analysis of HIV type 1 protease and reverse transcriptase in antiretroviral drug-naïve Ugandan adults. *AIDS Res. Hum. Retrovir.* 16:807-813.
- Clemente, J. C., R. M. Coman, M. M. Thiaville, L. K. Janka, J. A. Jeung, S. Nukoolkarn, L. Govindasamy, M. Agbandje-McKenna, R. McKenna, W. Leelamanit, M. M. Goodenow, and B. M. Dunn. 2006. Analysis of HIV-1

- CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry* **45**:5468–5477.
4. **Coman, R. M., A. Robbins, M. M. Goodenow, R. McKenna, and B. M. Dunn.** 2007. Expression, purification and preliminary X-ray crystallographic studies of the human immunodeficiency virus 1 subtype C protease. *Acta Crystallogr. F* **63**:320–323.
 5. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**:2126–2132.
 6. **Foulkes-Murzycki, J. E., W. R. Scott, and C. A. Schiffer.** 2007. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* **15**:225–233.
 7. **Gonzales, M. J., R. N. Machezano, and R. W. Shafer.** 2001. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J. Infect. Dis.* **184**:998–1006.
 8. **Julg, B., and F. D. Goebel.** 2005. HIV genetic diversity: any implications for drug resistance? *Infection* **33**:299–301.
 9. **Kantor, R., and D. Katzenstein.** 2003. Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS Rev.* **5**:25–35.
 10. **Kantor, R., D. A. Katzenstein, B. Efron, A. P. Carvalho, B. Wynhoven, P. Cane, J. Clarke, S. Sirivichayakul, M. A. Soares, J. Snoeck, C. Pillay, H. Rudich, R. Rodrigues, A. Holguin, K. Ariyoshi, M. B. Bouzas, P. Cahn, W. Sugiura, V. Soriano, L. F. Brigido, Z. Grossman, L. Morris, A. M. Vandamme, A. Tanuri, P. Phanuphak, J. N. Weber, D. Pillay, P. R. Harrigan, R. Camacho, J. M. Schapiro, and R. W. Shafer.** 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Med.* **2**:e112.
 11. **Laskowski, R. A., M. W. Mac Arthur, D. S. Moss, and J. M. Thornton.** 1993. PROCHECK. A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**:283–291.
 12. **McCutchan, F. E.** 2006. Global epidemiology of HIV. *J. Med. Virol.* **78**(Suppl. 1):S7–S12.
 13. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**:307–326.
 14. **Prabu-Jeyabalan, M., E. Nalivaika, N. M. King, and C. A. Schiffer.** 2004. Structural basis for coevolution of the human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J. Virol.* **78**:12446–12454.
 15. **Prabu-Jeyabalan, M., E. A. Nalivaika, K. Romano, and C. A. Schiffer.** 2006. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J. Virol.* **80**:3607–3616.
 16. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber.** 2000. HIV-1 nomenclature proposal. *Science* **288**:55–56.
 17. **Rose, R. B., C. S. Craik, and R. M. Stroud.** 1998. Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. *Biochemistry* **37**:2607–2621.
 18. **Sanches, M., S. Krauchenco, N. H. Martins, A. Gustchina, A. Wlodawer, and I. Polikarpov.** 2007. Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development. *J. Mol. Biol.* **369**:1029–1040.
 19. **Scott, W. R., and C. A. Schiffer.** 2000. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **8**:1259–1265.
 20. **Silva, A. M., R. E. Cachau, H. L. Sham, and J. W. Erickson.** 1996. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* **255**:321–346.
 21. **Todd, M. J., and E. Freire.** 1999. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36**:147–156.
 22. **Velazquez-Campoy, A., M. J. Todd, S. Vega, and E. Freire.** 2001. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc. Natl. Acad. Sci. USA* **98**:6062–6067.
 23. **Velazquez-Campoy, A., S. Vega, and E. Freire.** 2002. Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* **41**:8613–8619.