Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Molecular replacement with pseudosymmetry and model dissimilarity: a case study

Mariusz Jaskólski,[a] Mi Li,[b,c] Gary Laco,[d]‡ Alla Gustchina[b] and Alexander Wlodawer[b]*

[a]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University and Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [b]Protein Structure Section Section, Macromolecular Crystallography Laboratory, NCI at Frederick, Frederick, MD 21702, USA, [c]Basic Research Program, SAIC-Frederick, Frederick, MD 21702, USA, and [d]National Cancer Institute, Bethesda, MD 20892, USA

‡ Present address: Lake Erie College of Osteopathic Medicine-Bradenton, Bradenton, FL 34211, USA.

Correspondence e-mail: wlodawer@ncifcrf.gov

Received 31 October 2005
Accepted 5 December 2005

PDB Reference: HTLV-1 protease, 2b7f, r2b7fsf.

Crystals of human T-cell leukemia virus protease (HTLV-1 PR) have been very difficult to prepare and only native data extending to 2.6 Å resolution could be collected. Initial attempts to solve the structure with a variety of low-sequence-identity models utilizing proteases from other retroviruses and using a number of molecular-replacement programs were unsuccessful. The structure was finally solved using *Phaser*, revealing extensive pseudosymmetry and significant deviations from the starting models, features that were likely to be responsible for the initial failures. The steps taken to solve this structure and some of its intriguing crystallographic aspects are discussed.

1. Introduction

Although many macromolecular structures can now be solved in a rapid and routine manner, largely as a consequence of improvements in the methods of data collection and processing (Kabsch, 1993; Leslie *et al.*, 2002; Otwinowski & Minor, 1997; Pflugrath, 1999), automated phasing of diffraction data (Brunzelle *et al.*, 2003; de La Fortelle & Bricogne, 1997; Levitt, 2001; Ness *et al.*, 2004; Sheldrick, 1997; Terwilliger & Berendzen, 1999) and model building (Perrakis *et al.*, 1999; Terwilliger, 2003), as well as fast structure refinement (Brünger *et al.*, 1998; Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997; Sheldrick & Schneider, 1997; Tronrud *et al.*, 1987), occasional failures are still not uncommon. This is particularly true if the properties of the protein are such that crystals are difficult to obtain in a reproducible manner or cannot be derivatized. For larger structures, molecular replacement becomes the sole possible approach if only native data are available. This technique, which has been implemented in a number of computer programs, can be quite powerful if good-quality data are on hand and if an appropriate search model is available. However, not all available programs perform equally well, especially if the crystals exhibit properties such as pseudosymmetry, if the starting models are incomplete or if they differ from the target structure in a significant way.

Crystal structures of retroviral proteases have been studied very extensively since the first structures of the proteases encoded by Rous sarcoma virus (RSV PR) and human immunodeficiency virus type 1 (HIV-1 PR) were reported (Miller, Jaskólski *et al.*, 1989; Navia *et al.*, 1989; Wlodawer *et al.*, 1989). These structures were solved independently by multiple isomorphous replacement, but structures of many variants of these enzymes, crystallized in many different crystal forms, were subsequently solved by molecular replacement (Vondrasek & Wlodawer, 2002; Wlodawer &

Table 1
Data-collection statistics.

Values in parentheses relate to the highest resolution shell (∼5% of data).

Space group	*C*2
Unit-cell parameters (Å, °)	a = 134.32, b = 77.79, c = 80.38, β = 99.3
Temperature (K)	100
Molecules per ASU	6 (3 dimers)
Resolution (Å)	50–2.6
Total reflections	161514
Unique reflections	24645
Redundancy	6.55 (3.53)
Completeness (%)	98.1 (85.7)
$\langle I\rangle/\langle\sigma(I)\rangle$	21.7 (2.8)
$R_{\rm merge}$†	0.089 (0.304)
Wilson B factor (Å^2)	53.5

† $R_{\rm merge} = \sum_h \sum_i |I_{hi} - \langle I_h\rangle| / \sum_h \sum_i |I_{hi}|$, where I_{hi} is the observed intensity of the ith measurement of reflection h and $\langle I\rangle$ is the average intensity of that reflection obtained from multiple observations.

Erickson, 1993). Retroviral proteases are homodimers of two identical subunits, each 99–125 residues long. In pairwise comparisons, the root-mean-square devations between the C$^\alpha$ atoms of enzymes encoded by closely related viruses such as HIV-1, HIV-2 and simian immunodeficiency virus (SIV) are 1.0–1.2 Å, but for more divergent retroviruses the deviations are larger, between 1.5 and 2 Å. Although even in the latter cases the core of the molecules (especially in the area surrounding the active site) is very well conserved, significant differences are found in the numerous loop regions. Because of the relatively short length of the polypeptide chains, these loops represent a significant fraction of the fold and solving a novel structure of a retroviral protease by molecular replacement is not always easy. Although molecular replacement has successfully been used in the case of HIV-2 PR (Mulichak *et al.*, 1993), SIV PR (Rose *et al.*, 1993) and equine infectious anemia virus (EIAV) PR (Gustchina *et al.*, 1996), it failed in the case of the feline immunodeficiency virus (FIV) enzyme (Wlodawer *et al.*, 1995). In the latter case, it was necessary to solve the structure *de novo* using heavy-atom derivatives. Here, we describe the successful use of molecular replacement to solve the crystal structure of an inhibitor complex of human T-cell leukemia virus type 1 (HTLV-1) protease. This project was unusually difficult since most molecular-replacement programs failed to provide the correct solution and since, in view of the lack of a close structural probe, the process was very model-dependent. Using new approaches, we also managed retrospectively to re-solve the structure of FIV PR by molecular replacement.

2. Materials and methods

2.1. Crystallization of a complex of HTLV-1 PR with inhibitor

The expression, purification and crystallization of HTLV-1 PR have been described in detail elsewhere (Li *et al.*, 2005). The enzyme was complexed with the inhibitor Ac-Ala-Pro-Gln-Val-Sta-Val-Met-His-Pro (Louis *et al.*, 1999). The complex of HTLV-1 PR (monomer MW 12 554.6 Da) with the inhibitor (MW 1091.3 Da) was prepared by mixing the protein solution and the inhibitor (dissolved in 100% DMSO) in a molar ratio of 1:10 (protein monomer:inhibitor). Crystals of HTLV-1 PR were grown by the vapor-diffusion method in hanging drops made up of 4 µl protein solution and 4 µl well solution consisting of 17% PEG 8000, 16% PEG 300 and 10 m*M* DTT in 0.1 *M* acetate buffer pH 5.2. The crystals were frozen in liquid nitrogen with a cryosolution containing 17% PEG 8000, 16% PEG 300, 10 m*M* DTT, 0.1 *M* KH_2PO_4 and 10% ethylene glycol in 0.1 *M* acetate buffer pH 5.2 prior to data collection.

2.2. X-ray data collection and analysis

Low-temperature X-ray diffraction data extending to 2.6 Å resolution were collected at the SER-CAT insertion-device beamline 22-ID (Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA) on a MAR 225 CCD detector (MAR Research). Data were processed and scaled with *HKL*2000 (Otwinowski & Minor, 1997). Initial indexing of the diffraction pattern indicated a monoclinic *C* lattice (distortion indicator 0.12%), but a trigonal lattice was also a potential possibility (3.91%). However, since no satisfactory scaling of the data could be achieved with trigonal symmetry, the final integration and scaling were carried out in space group *C*2. The statistics of data processing are shown in Table 1.

2.3. Programs used in structure solution and refinement

The molecular-replacement programs used in this study included *AMoRe* (Navaza, 1994), *EPMR* (Kissinger *et al.*, 1999), *MOLREP* (Vagin & Teplyakov, 2000) and *Phaser* (Storoni *et al.*, 2004). The structure was fitted and rebuilt with *O* (Jones & Kjeldgaard, 1997) and refined with *REFMAC*5 (Murshudov *et al.*, 1997) and *CNS* (Brünger *et al.*, 1998). The figures were prepared with *BobScript* (Esnouf, 1999) and *PyMOL* (DeLano Scientific).

3. Results

3.1. Solution of the structure

The crystals of HTLV-1 PR belong to the monoclinic space group *C*2, with unit-cell parameters a = 134.32, b = 77.79, c = 80.38 Å, β = 99.3° (Table 1). Considering the dimeric nature of the protein, a plausible solvent content (53%, $V_{\rm M}$ = 2.6 Å^3 Da^{-1}; Matthews, 1968) could be calculated for six monomers in the asymmetric unit. However, the relatively large volume of the unit cell was compatible with a wide range of asymmetric unit compositions from three ($V_{\rm M}$ = 5.3 Å^3 Da^{-1}) to ten (1.6 Å^3 Da^{-1}) monomers. Odd numbers of copies could not be excluded owing to the presence of a crystallographic dyad in space group *C*2.

Numerous attempts to solve the structure using nearly all of the molecular-replacement programs available to us and using various models of the dimer constructed from the structures of retroviral proteases deposited in the Protein Data Bank (Berman *et al.*, 2000) proved to be futile. A successful solution was only obtained with the program *Phaser* (Storoni *et al.*, 2004) run in a single-model mode. *Phaser* could solve the

Table 2
Models used for molecular-replacement calculations and the results of different runs.

Model	Model resolution (Å)	Virus	Sequence indentity (%)	Program	MR data resolution (Å)	*B* factor (Å^2)	CC	*R* factor	*Z* score	LL gain	Solution
1kzk	1.09	HIV-1	27.5	*Phaser*	20–2.6	From PDB	N/A	N/A	10.16	189.5	Yes
1nh0	1.03	HIV-1	27.5	*Phaser*	20–2.6	From PDB	N/A	N/A	11.16	176.9	Yes
1ody	2.00	HIV-1	27.5	*Phaser*	20–2.6	From PDB	N/A	N/A	14.54	270.0	Yes
1bai	2.40	RSV	21.5	*Phaser*	20–2.6	From PDB	N/A	N/A	7.96	112.6	Yes
2fiv	2.00	FIV	24.1	*Phaser*	20–2.6	From PDB	N/A	N/A	7.80	107.2	No
2fmb	1.80	EIAV	27.5	*Phaser*	20–2.6	From PDB	N/A	N/A	10.43	141.6	Yes
1kzk	1.09	HIV-1	27.5	*AMoRe*	15–2.6	From PDB	0.366	0.514	N/A	N/A	No
1nh0	1.03	HIV-1	27.5	*AMoRe*	15–2.6	From PDB	0.361	0.519	N/A	N/A	No
1ody	2.00	HIV-1	27.5	*AMoRe*	15–2.6	From PDB	0.452	0.501	N/A	N/A	Yes
1kzk	1.09	HIV-1	27.5	*AMoRe*	15–2.6	30	0.406	0.525	N/A	N/A	No
1nh0	1.03	HIV-1	27.5	*AMoRe*	15–2.6	30	0.408	0.527	N/A	N/A	No
1ody	2.00	HIV-1	27.5	*AMoRe*	15–2.6	30	0.422	0.529	N/A	N/A	No
1kzk	1.09	HIV-1	27.5	*AMoRe*	15–2.6	25	0.400	0.525	N/A	N/A	No
1nh0	1.03	HIV-1	27.5	*AMoRe*	15–2.6	25	0.400	0.524	N/A	N/A	No
1ody	2.00	HIV-1	27.5	*AMoRe*	15–2.6	25	0.405	0.523	N/A	N/A	No
1kzk/main	1.09	HIV-1	27.5	*AMoRe*	15–2.6	30	0.390	0.563	N/A	N/A	No
1nh0/main	1.03	HIV-1	27.5	*AMoRe*	15–2.6	30	0.382	0.563	N/A	N/A	No
1ody/main	2.00	HIV-1	27.5	*AMoRe*	15–2.6	30	0.382	0.564	N/A	N/A	No
1ody	2.00	HIV-1	27.5	*AMoRe*	15–3.0	From PDB	0.330	0.517	N/A	N/A	No
1ody	2.00	HIV-1	27.5	*AMoRe*	15–3.5	From PDB	0.240	0.520	N/A	N/A	No

structure using various retroviral proteases as the molecular probe, but the quality of the solutions reflected the level of sequence identity estimated in pairwise comparisons using *CLUSTALW* (Thompson *et al.*, 1994). The best unambiguous solution (*Z* score 12–11) was obtained with HIV-1 PR (sequence identity of 27.5% calculated for the 116-residue long construct of HTLV-1 PR). We initially utilized the protease model from the highest resolution (1.03 Å) structure (PDB code 1nh0) of an inhibitor complex (Brynda *et al.*, 2004), but, retrospectively, other HIV-1 PR coordinate sets also gave similar or even much better solutions (Table 2). Consistently, the *Phaser* solutions corresponded to three protease dimers (*AB*, *CD* and *EF*) arranged into a nearly equilateral triangle around a non-crystallographic threefold axis along *c** (Fig. 1). Packing considerations precluded finding more than three dimers in the asymmetric unit.

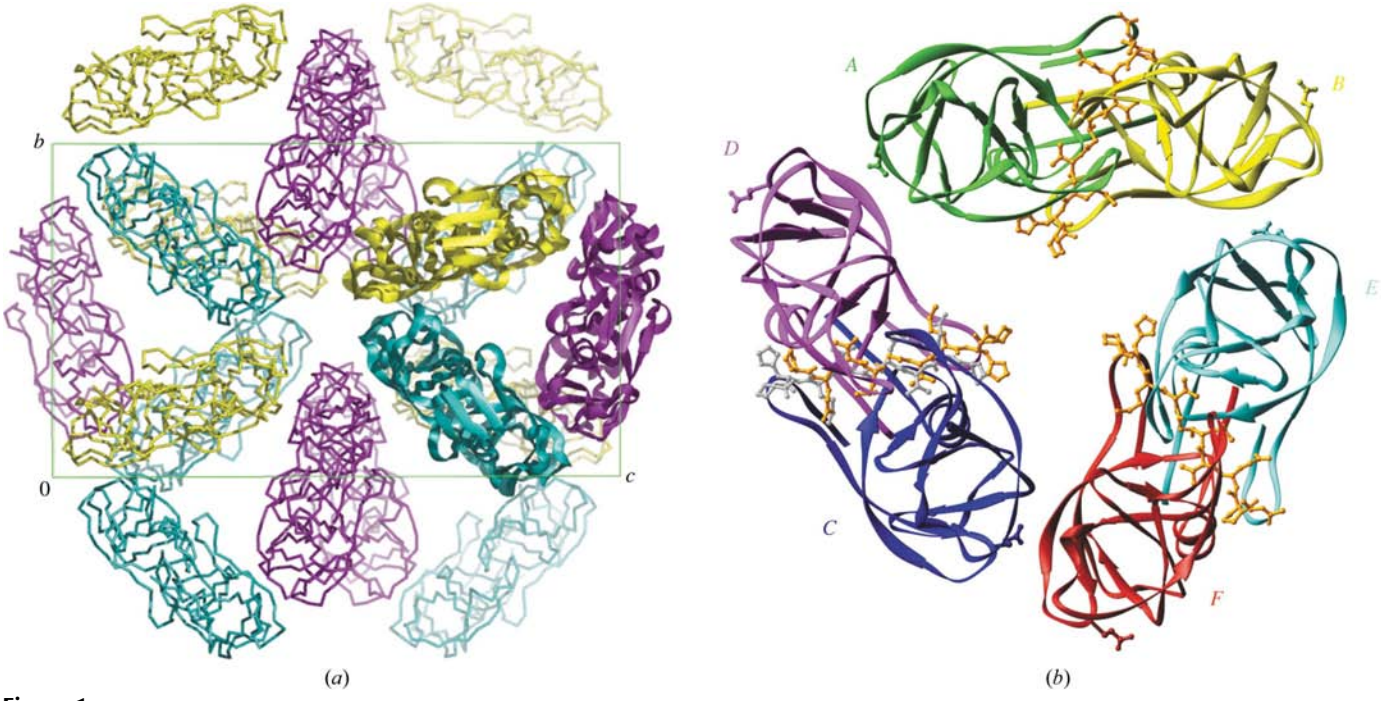

Figure 1
The structure of HTLV-1 PR. (*a*) Crystal packing viewed along [001]. The trimer of dimers of HTLV-1 PR forming the asymmetric unit of the monoclinic *C*2 cell is shown with molecules *AB* colored yellow, *CD* magenta and *EF* blue. The unique trimer is in ribbon representation, whereas the symmetry mates are shown as traces. (*b*) The unique trimer shown in ribbon representation with each monomer colored separately. The full atomic models for the inhbitor molecules are in ball-and-stick representation. The side chains of Gln20 residues involved in hydrogen-bonded interactions with the neighboring molecules are also shown in ball-and-stick representation.

Table 3
Refinement statistics.

Reflections (refinement/test)	23030/1143
R†	0.198
R_{free}‡	0.278
R.m.s.d. bond lengths§ (Å)	0.022
R.m.s.d. bond angles§ (°)	2.18
Ramachandran φ/ψ statistics (excluding inhibitor) (%)	
Most favored	89.0
Additional allowed	9.9
Generously allowed	1.1
Average B factor (Å^2)	
Protein	35.8
Inhibitor/ions	45.3
Solvent	38.3
No. of protein atoms	5298
No. of inhibitor atoms	284
No. of solvent molecules	173
No. of heteroatoms	10

† $R = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. ‡ R_{free} is defined in Brünger (1992). § Ideal values for stereochemical parameters were as defined by Engh & Huber (1991).

The phases generated by *Phaser* were improved by map averaging according to the non-crystallographic threefold symmetry using *DM* (Cowtan, 1994) as implemented in *CCP*4*i* (Collaborative Computational Project, Number 4, 1994). The *DM* run, which also included multi-resolution histogram matching and solvent flattening (45% solvent), produced excellent results, as illustrated by the map correlation coefficients (0.820–0.995) and map R_{free} (0.122). An F_o map calculated with the improved phases was very clear in the well conserved regions of the protease, including the side chains of the two Asp32 residues that create the active site (Fig. 2), although it was rather sparse or uninterpretable in the surface-loop areas, where the superimposed models of other retroviral proteases also disagree. A remarkable feature of this initial map was the presence of strong electron density defining nearly the whole inhibitor molecule (nine amino acids, including the central statine residue) in all three crystallographically independent dimers. The inhibitors were very clear in dimers *AB* and *EF* (labeled *I* and *K*, respectively), whereas in dimer *CD* the inhibitor map was generally poorer and only the main chain was visible. However, at a later stage, difference Fourier syntheses clearly showed inhibitor side-chain locations consistent with two orientations of the ligand in the binding site of dimer *CD*, modeled with labels *J* and *J*′.

The NCS-averaged F_o map served to build the initial model, which contained a number of gaps in the loop regions, including the flap arms. Further model building continued in the program *O* (Jones & Kjeldgaard, 1997) using $2mF_o - DF_c$ and $mF_o - DF_c$ maps generated after each cycle of refinement, which consisted of simulated-annealing refinement with *CNS* (Brünger *et al.*, 1998) combined with conventional refinement in *REFMAC*5 (Murshudov *et al.*, 1997), using maximum-likelihood target functions in both programs. The initial R factor was 0.386 and fell after each round of structure rebuilding/refinement. The improvement of the model was monitored by R_{free} (Brünger, 1992) calculated for 1143 randomly selected reflections which were excluded from refinement. The parameters of the multiple copies of the molecules were refined with medium NCS restraints, which were relaxed to loose towards the end of the refinement. Three separate sets of NCS restraints were introduced. One set was applied to protein subunits *A*, *C* and *E*, another to protein subunits *B*, *D* and *F* and a third to the inhibitor molecules *I*, *K*, *J* and *J*′ (the latter two representing the twofold-disordered inhibitor in dimer *CD*). As seen in Fig. 1(*b*), molecules *A* and *E* interact with the C-terminal parts of the inhibitors *I* and *K*, respectively, while *B* and *F* interact with the N-terminal parts of these molecules. The good quality of the final electron-density maps justified the interpretation of a limited number of water molecules, including those with a characteristic tetrahedral coordination by the flaps and the central part of the inhibitor in each dimer, corresponding to Wat301 in HIV-1 PR (Miller, Schneider *et al.*, 1989). The final round of refinement, which included bulk-solvent correction, resulted in a model with good stereochemistry (r.m.s. deviation from ideal bond lengths 0.022 Å), characterized by an R factor of 0.198 and an R_{free} of 0.278 (Table 3). The final structure has been deposited in the PDB with accession code 2b7f.

The initial electron-density maps were quite highly biased to the original model, causing some problems in revising the model. However, the presence of six molecules in the asymmetric unit was a bonus since a likely tracing for one of them could be propagated to the other molecules and tested during the refinement procedure. This approach was particulary useful for the flap loops in molecules *C* and *D*, which were not

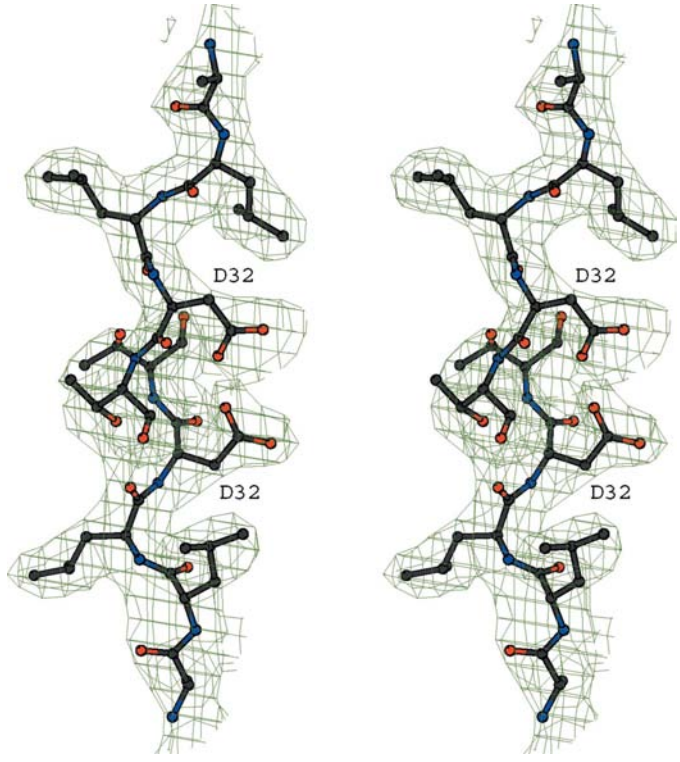

Figure 2
Stereoview of an F_o electron-density map (contour level 1σ) resulting from molecular replacement, averaging and solvent flattening, covering the region of the active site of the enzyme (ball-and-stick model) in dimer *AB*.

visible in the initial maps and are still of comparatively poor quality even in the final map owing to the twofold disorder of the inhibitor. However, in most other areas the final maps were very good.

3.2. Analysis of crystal packing

The asymmetric unit of the HTLV-1 PR crystal contains three dimers (labeled *AB*, *CD* and *EF*) related by a non-crystallographic threefold axis that is aligned with the *c** direction. A packing diagram is shown in Fig. 1(*a*). While the diffraction data could be easily indexed in a monoclinic *C*2 cell (Table 1), some indication of quasi-trigonal packing was obvious from the metrics of the lattice, since the *C*-centered (001) plane has an almost perfectly trigonal pattern of nodes (the nodes 000, 010 and $\frac{11}{22}0$ form an equilateral triangle). This pseudosymmetry became particularly clear in self-rotation superpositions of the Patterson map.

The molecular dyads of the three dimers forming the molecular-replacement solution are aligned within 10° with the local pseudo-threefold axis. Both these local symmetry elements can be seen in self-superpositions of the Patterson map, which show strong peaks for $\kappa = 120°$ and 180° at $\omega = 0°$ (Fig. 3). In this arrangement, the flap regions of the three dimers are all situated on one 'face' of the triangle, while the four-stranded interface β-sheets (composed of all the terminal peptides in each dimer) are situated on the other face. The inhibitor molecules *I* and *K*, bound by the *AB* and *EF* dimers of the triangle, respectively, have unambiguous and consistent polarity, with their C-termini pointing towards the pseudo-threefold axis (Fig. 1*b*). The inhibitor molecule in dimer *CD* shows evident twofold disorder in the active site (*J* and *J*′). The two orientations have been modeled with half occupancy, with molecule *J* oriented consistently with the pattern defined by inhibitors *I* and *K*.

The *C*-lattice translation propagates the *AB*–*CD*–*EF* trimer into a flat molecular (001) layer of trigonal pattern (Fig. 1*a*). This packing results from the trigonal metric symmetry of the *ab* plane and from the approximate alignment of the *AB*–*CD*–*EF* triangle with those pseudo-trigonal directions (the angle between the long axis of dimer *CD* and the *b* axis is about 8°). It should be noted that because of the pseudo-threefold symmetry, the *ab* plane contains three twofold axes, namely the crystallographic dyad along [010] and two non-crystallographic dyads at 120°. The main lattice contacts are within the layers so that the structure can be considered to be composed of tightly knitted trigonal layers stacked in the *c** direction. The molecular layer has a thickness of about 40 Å and extends between $z = 0$ and $z = 1/2$. The crystallographic twofold axis generates consecutive layers in the *c** direction, which alternately face each other with the flap faces or the dimeric β-sheet faces.

In general, the inter-layer contacts of molecule *AB* are with *E*′*F*′, whereas dimer *CD* forms contacts only with its symmetry-related replicas. Several direct hydrogen-bonding interactions involve residues that belong to the flaps of two adjacent dimers. The carboxylate of Asp65 of molecule *A* makes a pair of hydrogen bonds with the main-chain amide of residue 64 and Thr63 $O^{\gamma 1}$ of molecule *F*′, and a reciprocal interaction is made by *F* and *A*′. The side-chain amide of Gln64 of molecule *B* interacts with Asn53 of molecule *C*′. In contrast, Asp65 of molecule *E* is quite distant from any symmetry-related molecules, whereas Asp65 of molecule *D* forms a hydrogen bond with the amide N atom of residue Thr54 of *D*′. The side-chain NH_2 group of Gln64 of molecule *D* is within hydrogen-bonding distance of the main-chain

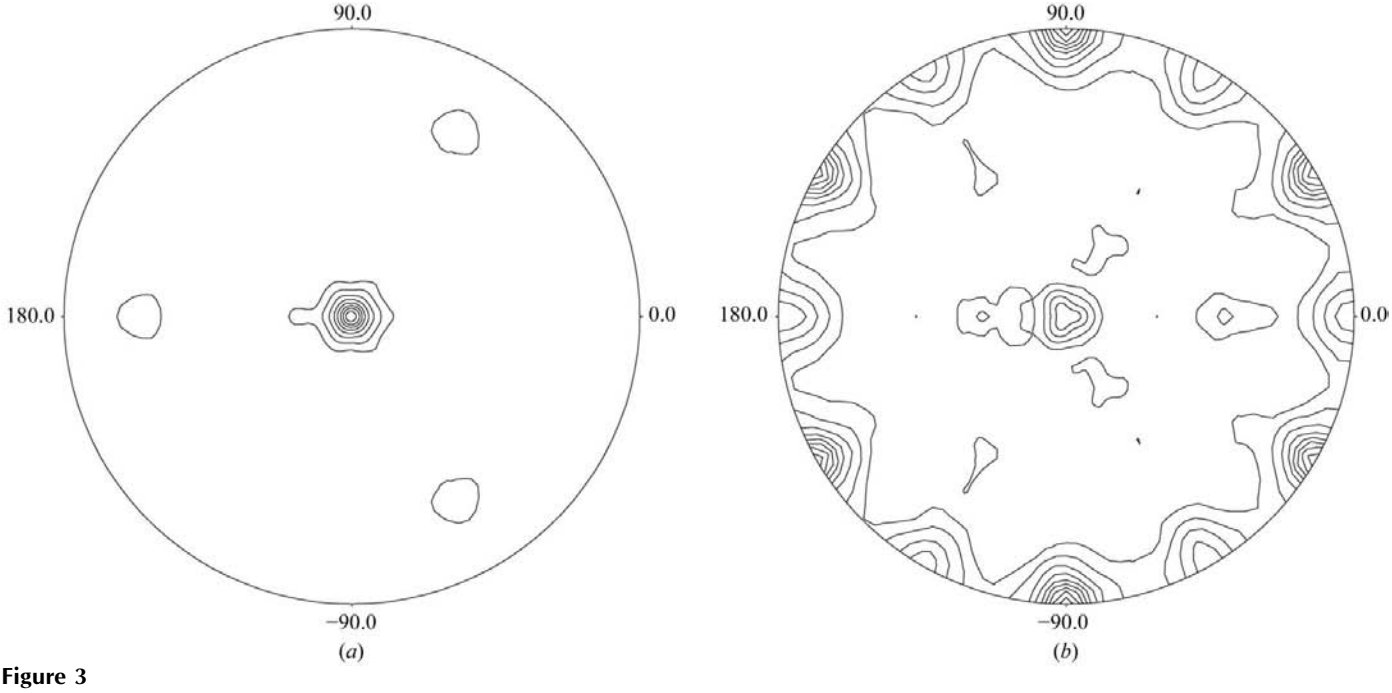

Figure 3
Stereographic projections of Patterson self-rotation function corresponding to (*a*) $\kappa = 120°$ and (*b*) $\kappa = 180°$.

amide of its symmetry-related counterpart. Although these interactions of the flap arms are not very extensive, their symmetric character for molecule *D* may be related to the twofold disorder of inhibitor *J*/*J*′ and to the generally poorer quality of the electron-density map in the vicinity of the active site of dimer *CD*.

Direct interactions between the three dimers in the pseudotrigonal plane include hydrogen bonds between the side chains of Gln20 belonging to molecules *A*, *C* and *E* (Fig. 1*b*) and the main-chain N and O atoms of Thr26 of molecules *D*, *F* and *B*, respectively. Similar interactions are also found between Gln20 of molecules *B*, *D* and *F* and Thr26 of *C*′, *E*′ and *A*′, respectively, which are crystallographic copies of *C*, *E* and *A*. Residues 14, 72–73 and 86–88 are involved in creating hydrophobic contacts within the in-plane trimers of dimers.

3.3. Comparison of the individual molecules

Since the crystals of HTLV-1 PR contain six individual polypeptide chains forming three distinct dimers, it is possible to observe the combined effects of crystal packing and refinement errors in the differences between them. Arbitrarily taking molecule *A* as a reference and superposing it on the other molecules using the program *ALIGN* (Cohen, 1997) yields r.m.s. deviations between all 116 C^{α} atoms of 0.45, 0.37, 0.51, 0.28 and 0.46 Å for molecules *B*, *C*, *D*, *E* and *F*, respectively. A slight asymmetry of the dimers, related to the polarity of the inhibitors, is reflected in the lower deviation for molecule *E*. Some side chains, for example Lys14, show two distinct sets of orientations, one in molecules *A*, *C* and *E* and another in *B*, *D* and *F*. This distinction is most likely to be the result of different interactions required by crystal packing. An interesting case is provided by the conformation of the side chain of Trp98, which is virtually identical in molecules *A*, *C*, *D* and *E* but is rotated by ∼45° in molecule *B* and ∼90° in *F*. Since this residue directly interacts with the opposing flap, its variability may be related to a combination of the directionality of the inhibitor and differences in crystal contacts between dimers *AB* and *EF* and perhaps also to the disorder of the inhibitor in *CD*, where multiple conformations of Trp98 may exist but could not be modeled owing to poor electron density. However, these comparisons show that the differences between the individual molecules are not large and the non-crystallographic threefold symmetry is generally obeyed.

3.4. A new approach to solving the structure of FIV PR

The structure of FIV PR reported ten years ago (Wlodawer *et al.*, 1995) was solved using a combination of molecular replacement with multiple isomorphous replacement, since maps calculated with MR phases alone did not contain enough information to allow proper tracing of the complete chain of the molecule. In view of the present success of molecular replacement with HTLV-1 PR and to test whether more advanced approaches would be more successful in solving that structure without resorting to MIR phasing, we carried out the MR procedure with *Phaser*. The diffraction data used for this purpose corresponded to the native data set measured for a complex of FIV PR with the inhibitor TL-3 that was used to refine the original structure (Wlodawer *et al.*, 1995). It consisted of 9018 reflections extending to a resolution of 1.9 Å. The intial model was chain *A* of the HIV-1 PR structure used to solve the present HTLV-1 PR structure, *i.e.* the atomic resolution model with PDB code 1nh0 (Brynda *et al.*, 2004; not available at the time the original FIV PR structure was solved). The polypeptide chain of FIV PR consists of 116 residues, at variance with the 99 residues found in HIV-1 PR, with the differences concentrated at the two termini and two insertions located between residues 67 and 70 as well as between 78 and 81 (HIV-1 PR numbering). Unlike HTLV-1 PR, the $P3_121$ crystals of FIV PR contain only a single polypeptide chain in the asymmetric unit and the dimer is generated by the operation of crystallographic symmetry, leading to a twofold-disordered inhibitor.

A run of *Phaser* in a default mode resulted in a single solution with a *Z* score of 10.14 and an LL gain of 61.97. The solution was refined with *CNS* using the rigid-body refinment option, yielding $R = 51\%$. However, the resulting map was of good quality, although the density in the loops expected to be present only in FIV PR was not contiguous. In the next step, the HIV-1 PR sequence was replaced with amino acids corresponding to the FIV PR sequence in those regions where the two sequences could be aligned, still leaving the gaps as defined above. The resulting model was further refined with *CNS*, including a step of simulated annealing (Brünger *et al.*, 1990). The resulting $2F_o - F_c$ map showed unambigous density for the missing regions of FIV PR and for the inhibitor, which was not part of the MR model. We are confident that the quality of this map would have been sufficient to complete the refinement without the need for any extra phasing information.

4. Discussion

The crystals of HTLV-1 PR discussed here presented us with various problems and difficulties and the question might be raised whether the effort required to solve this structure was justified in the first place or whether it would not have been more prudent to simply abandon these problematic crystals and search for conditions that would yield a different form. However, the multitude of setbacks encountered while working with HTLV-1 PR precluded such an approach. The original construct of the protein that contained all 125 residues never crystallized and only crystallization of a truncated 116-amino-acid construct was possible. However, even this protein was prone to aggregation and crystallization experiments had to be set up within a few hours of the final purification step or the protein would precipitate. We were also not able to freeze stock solutions of HTLV-1 PR in order to preserve it for future experimentation. Thus, the growth of single crystals was irreproducible and erratic and the crystal used for data collection was quite unique in its single character and its ability to diffract better than most others. In view of those crystal-related problems, we were unable to resort to heavy-atom

derivatives for phasing and were left with a single native data set and with molecular replacement as the only practical approach to solving the structure.

Our initial attempts to solve the structure using the coordinates of all known retroviral proteases and a variety of molecular-replacement programs were generally unsuccessful. Although the presence of a non-crystallographic threefold axis could be readily detected, no satisfactory solutions to the translation function were obtained. Since the extent of deviation of the final coordinates of HTLV-1 PR from the starting models was correctly predicted to exceed 1.5 Å, our initial efforts to solve the structure only used data extending to 3.0–3.5 Å. This is the usual practice recommended for programs such as *EPMR* or *AMoRe*, but it was clearly not successful in this case. The only successful run of the latter program utilized all data to 2.6 Å, whereas calculations utilizing 3.0 and 3.5 Å data had failed (Table 2). This result is counterintuitive and illustrates the fact that the success of molecular replacement is often critically dependent on the correct choice of parameters. We also should note that the successful run of *AMoRe* was performed only after the structure had been solved with *Phaser* and the proper solution could be easily verified. On the other hand, all runs with *Phaser* using different HIV-1 PR models, as well as most of the runs that utilized the coordinates of other retroviral proteases, were successful, although the quality of the statistical criteria varied (Table 2).

We do not have an explanation why the solutions utilizing the medium-resolution HIV-1 PR coordinate set 1ody were significantly better than when atomic resolution models were used in the molecular-replacement calculations with *Phaser* and, indeed, why 1ody was the only successful model with *AMoRe*. The r.m.s. deviation of these coordinates from the final HTLV-1 PR model is marginally higher than for the atomic resolution models 1kzk and 1nh0 and the use of an identical overall temperature factor prevented correct solution. We were not able to run *Phaser* in a multi-model mode because of packing clashes between the different components of the set detected by the program. The observations summarized in Table 2 do, however, support the conclusion that when many possible models are available, all should be investigated as potential starting points, without limiting the search to only the highest-resolution 'best' structures. We can also recommend for programs such as *AMoRe* the inclusion of higher resolution reflections if no solutions are apparent with data truncated by a high-resolution limit.

We thank Dr Randy Read for helpful hints on running *Phaser* and Dr J. K. Mohana Rao for his help in early stages of molecular replacement. We acknowledge the use of beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) located at the Advanced Photon Source, Argonne National Laboratory. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. W-31-109-Eng-38. The project was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and in part with Federal funds from the National Cancer Institute, National Institutes of Health under contract No. NO1-CO-24000. The research of MJ was supported by a Faculty Scholar fellowship from the National Cancer Institute and by a subsidy from the Foundation for Polish Science. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
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.
Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.
Brunzelle, J. S., Shafaee, P., Yang, X., Weigand, S., Ren, Z. & Anderson, W. F. (2003). *Acta Cryst.* D**59**, 1138–1144.
Brynda, J., Rezacova, P., Fabry, M., Horejsi, M., Stouracova, R., Soucek, M., Hradilek, M., Konvalinka, J. & Sedlacek, J. (2004). *Acta Cryst.* D**60**, 1943–1948.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Gustchina, A., Kervinen, J., Powell, D. J., Zdanov, A., Kay, J. & Wlodawer, A. (1996). *Protein Sci.* **5**, 1453–1465.
Jones, T. A. & Kjeldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leslie, A. G., Powell, H. R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. & Nave, C. (2002). *Acta Cryst.* D**58**, 1924–1928.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
Li, M., Laco, G. S., Jaskólski, M., Rozycki, J., Alexandratos, J., Wlodawer, A. & Gustchina, A. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18332–18337.
Louis, J. M., Oroszlan, S. & Tozser, J. (1999). *J. Biol. Chem.* **274**, 6660–6666.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J. & Wlodawer, A. (1989). *Nature (London)*, **337**, 576–579.
Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H. & Wlodawer, A. (1989). *Science*, **246**, 1149–1152.
Mulichak, A. M., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Curry, K. A., Tomich, C. S., Thaisrivongs, S., Sawyer, T. K. & Watenpaugh, K. D. (1993). *J. Biol. Chem.* **268**, 13103–13109.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. & Springer, J. P. (1989). *Nature (London)*, **337**, 615–620.
Ness, S. R., de Graaff, R. A., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pflugrath, J. P. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rose, R. B., Rose, J. R., Salto, R., Craik, C. S. & Stroud, R. M. (1993). *Biochemistry*, **32**, 12498–12507.
Sheldrick, G. M. (1997). *Methods Enzymol.* **276**, 628–641.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Vondrasek, J. & Wlodawer, A. (2002). *Proteins*, **49**, 429–431.
Wlodawer, A. & Erickson, J. W. (1993). *Annu. Rev. Biochem.* **62**, 543–585.
Wlodawer, A., Gustchina, A., Reshetnikova, L., Lubkowski, J., Zdanov, A., Hui, K. Y., Angleton, E. L., Farmerie, W. G., Goodenow, M. M., Bhatt, D., Zhang, L. & Dunn, B. M. (1995). *Nature Struct. Biol.* **2**, 480–488.
Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989). *Science*, **245**, 616–621.