STRUCTURAL BIOLOGY

ISSN 2059-7983

Comparison of a retroviral protease in monomeric and dimeric states

Stanislaw Wosicki,[a] Miroslaw Gilski,[a,b] Helena Zabranska,[c] Iva Pichova[c]* and Mariusz Jaskolski[a,b]*

[a]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland, [b]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, 61-614 Poznan, Poland, and [c]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, 166 10 Prague, Czech Republic. *Correspondence e-mail: iva.pichova@uochb.cas.cz, mariuszj@amu.edu.pl

Received 3 July 2019
Accepted 13 August 2019

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: Mason–Pfizer monkey virus; M-PMV; retroviruses; retropepsin; aspartic protease; dimerization; inhibitor; flap structure.

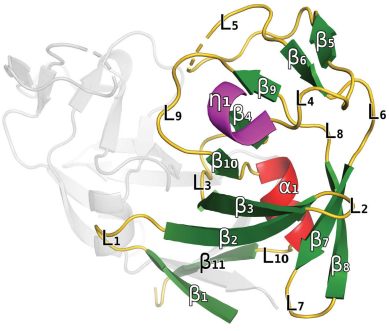
PDB references: dimeric forms of M-PMV protease: C7A/D26N/C106A mutant with inhibitor, 6s1u; D26N mutant with inhibitor, 6s1v; D26N mutant without inhibitor, 6s1w

Supporting information: this article has supporting information at journals.iucr.org/d

Retroviral proteases (RPs) are of high interest owing to their crucial role in the maturation process of retroviral particles. RPs are obligatory homodimers, with a pepsin-like active site built around two aspartates (in DTG triads) that activate a water molecule, as the nucleophile, under two flap loops. Mason–Pfizer monkey virus (M-PMV) is unique among retroviruses as its protease is also stable in the monomeric form, as confirmed by an existing crystal structure of a 13 kDa variant of the protein (M-PMV PR) and its previous biochemical characterization. In the present work, two mutants of M-PMV PR, D26N and C7A/D26N/C106A, were crystallized in complex with a peptidomimetic inhibitor and one mutant (D26N) was crystallized without the inhibitor. The crystal structures were solved at resolutions of 1.6, 1.9 and 2.0 Å, respectively. At variance with the previous study, all of the new structures have the canonical dimeric form of retroviral proteases. The protomers within a dimer differ mainly in the flap-loop region, with the most extreme case observed in the apo structure, in which one flap loop is well defined while the other flap loop is not defined by electron density. The presence of the inhibitor molecules in the complex structures was assessed using polder maps, but some details of their conformations remain ambiguous. In all of the presented structures the active site contains a water molecule buried deeply between the Asn26-Thr27-Gly28 triads of the protomers. Such a water molecule is completely unique not only in retropepsins but also in aspartic proteases in general. The C7A and C106A mutations do not influence the conformation of the protein. The Cys106 residue is properly placed at the homodimer interface area for a disulfide cross-link, but the reducing conditions of the crystallization experiment prevented S—S bond formation. An animated Interactive 3D Complement (I3DC) is available in Proteopedia at http://proteopedia.org/w/Journal:Acta_Cryst_D:S2059798319011355.

1. Introduction

Retroviral proteases function in the obligatory step of virion maturation, processing retroviral polyproteins into the final protein products, including the protease (PR) itself. The retroviral protease has the catalytic apparatus of a pepsin-like aspartic protease (Miller, Jaskólski *et al.*, 1989; thus the name retropepsin) but is assembled from two identical polypeptides (of about 100 residues), each contributing one aspartate side chain to the active site. The mechanism by which the protease is able to excise itself from the polyprotein is not very clear, particularly considering the fact that the dimerization mechanism is based on an intermolecular β-sheet woven from the four termini of the two subunits (Wlodawer *et al.*, 1989). In this context, studying systems in which the protease activation



includes several separable events or/and where the protease can be isolated in the monomeric form is of high interest. One such system is Mason–Pfizer monkey virus (M-PMV), which causes an AIDS-like syndrome in rhesus monkeys. In the virions, M-PMV PR is excised autocatalytically from the Gag-Pro polyprotein in a 17 kDa form (per monomer) that undergoes further C-terminal processing, yielding a 13 kDa PR form with comparable catalytic efficiency (Zábranský *et al.*, 1998). *In vitro*, the 13 kDa (114-residue) form can autodigest even further, producing a 12 kDa (107-residue) polypeptide, which is, however, more than ten times less active (Zábranská *et al.*, 2007). Precise structural characterization of a retroviral protease monomer is also attractive from the point of view of the design of dimerization inhibitors as new-generation drugs against retroviral infection-related diseases, including AIDS. The existing drugs targeting Human immunodeficiency virus type 1 (HIV-1) PR are all active-site inhibitors (Wlodawer & Jaskolski, 2016) that mimic the mode of substrate binding between the catalytic apparatus and a pair of extended, mobile loops called flaps. Since it is not possible to isolate HIV-1 PR monomers in crystallizable amounts, the published crystal structure of M-PMV PR in a monomeric form (Gilski *et al.*, 2011) has generated a lot of interest, especially as it was solved in an unprecedented manner with the use of the *Foldit* protein-folding game (Kleffner *et al.*, 2017) that enlisted help from hundreds of thousands of citizen scientists (Khatib *et al.*, 2011). Retroviral proteases [for example, those from HIV-1 (Wlodawer *et al.*, 1989), Human immunodeficiency virus type 2 (HIV-2 PR; Mulichak *et al.*, 1993), Simian immunodeficiency virus (SIV PR; Rose *et al.*, 1996), Feline immunodeficiency virus (FIV PR; Wlodawer *et al.*, 1995), Rous sarcoma retrovirus (RSV PR; Jaskolski *et al.*, 1990), Equine infectious anemia virus (EIAV PR; Gustchina *et al.*, 1996), Human T cell leukemia virus (HTLV-1 PR; Li *et al.*, 2005) and Xenotropic murine leukemia virus-related virus (XMRV PR; Li, DiMaio *et al.*, 2011)] form tight homodimers, with the exception of the uniquely behaved M-PMV PR (Gilski *et al.*, 2011), which also forms stable monomers.

A common template consisting of the retropepsin fold, which is based on nonviral aspartic proteases (Andreeva, 1991), defines the secondary structure as a duplication of four structural elements, as shown in Fig. 1: a hairpin (loops A1 and A2), a wide loop (B1, comprising the DTG catalytic triad, and B2), an α-helix (C1 and C2) and a second hairpin (D1, termed the 'flap', and D2). The interactions between protomers within a retroviral protease dimer consist of (i) a four-stranded antiparallel β-sheet made of the C- and N-termini of both chains, (ii) engagement of the catalytic triads in a hydrogen-bonding network, including the so-called 'fireman's grip', and (to a limited extent) (iii) hydrogen-bonding contacts between the tips of the flap loops. In the presence of a peptidomimetic inhibitor in the binding pocket, an interface water molecule (hereafter referred to as WatI) is usually hydrogen-bonded in a tetrahedral fashion by two carbonyl groups of the inhibitor acting as acceptors and two main-chain NH amide groups of residues at the tips of the flaps acting as donors (Jaskolski *et al.*, 1991).

In this work, we crystallized two variants of M-PMV PR in its dimeric form. One of these variants is identical to that used for the crystallization of the monomeric protein (Gilski *et al.*, 2011) and has both of the cysteine residues present in the native sequence replaced by alanines (C7A/C106A) as a precaution against potential disulfide-locked dimerization. In addition, the active-site aspartate was mutated to asparagine (D26N) to prevent autodigestion. The second variant has no cysteine-to-alanine replacements but has the active-site D26N mutation. For the crystallization experiments of both variants, we incubated the protein with the same active-site inhibitor as was used during the crystallization of the monomeric protein (without its structural presence). For the D26N single mutant, an additional crystal form was obtained in a crystallization trial without inhibitor. All of the crystal forms of dimeric M-PMV PR presented in this work, *i.e.* the triple mutant with the inhibitor (3MI), the single mutant with the inhibitor (1MI) and the single mutant without inhibitor (1M), are isomorphous and are different from the crystals of the monomeric protein, despite the fact that one of the present crystals of dimeric M-PMV PR was obtained using nearly identical crystallization conditions to those that yielded the monomeric form.

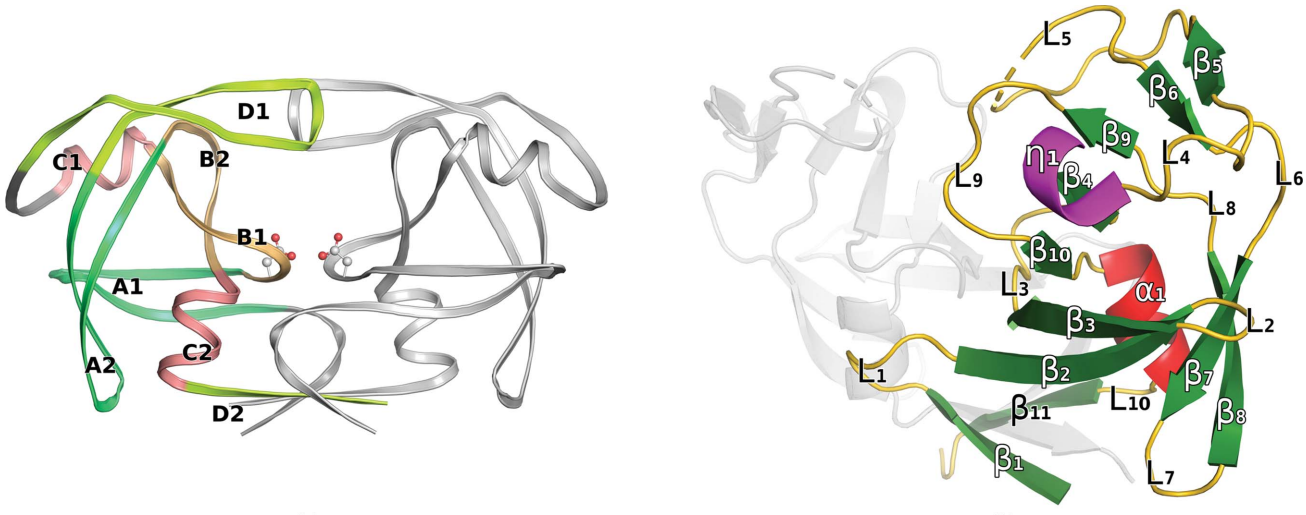
In contrast to the large number of existing efficient inhibitors of HIV PR, only a few chemical structures of M-PMV PR inhibitors have been published (Hrusková-Heidingsfeldová *et al.*, 1995). These inhibitors were originally designed for the protease from Myeloblastosis-associated virus (MAV), the helper virus of Avian myeloblastosis virus (AMV). The amino-acid sequence of MAV PR is highly similar to that of RSV PR. The best compound from this series, with the sequence Pro-(O-Me)Tyr-Val-PSA-Ala-Met-Thr, where (O-Me)Tyr is O^{η}-methylated tyrosine and PSA denotes (3*S*,4*S*)-3-hydroxy-4-amino-5-phenylpentanoic acid, inhibited M-PMV PR at a nanomolar concentration and displayed better inhibition parameters for M-PMV PR than for MAV PR (K_i values of 3 and 10 n*M*, respectively; Hrusková-Heidingsfeldová *et al.*, 1995). This inhibitor was used in the crystallization experiments in this work.

The history of struggles with the structure of retroviral proteases as the main goal is long and picturesque, both worldwide and in our laboratory (Jaskolski *et al.*, 2015). Yet, as the present account confirms, we can still find new, intriguing features within this fascinating family of key therapeutic targets.

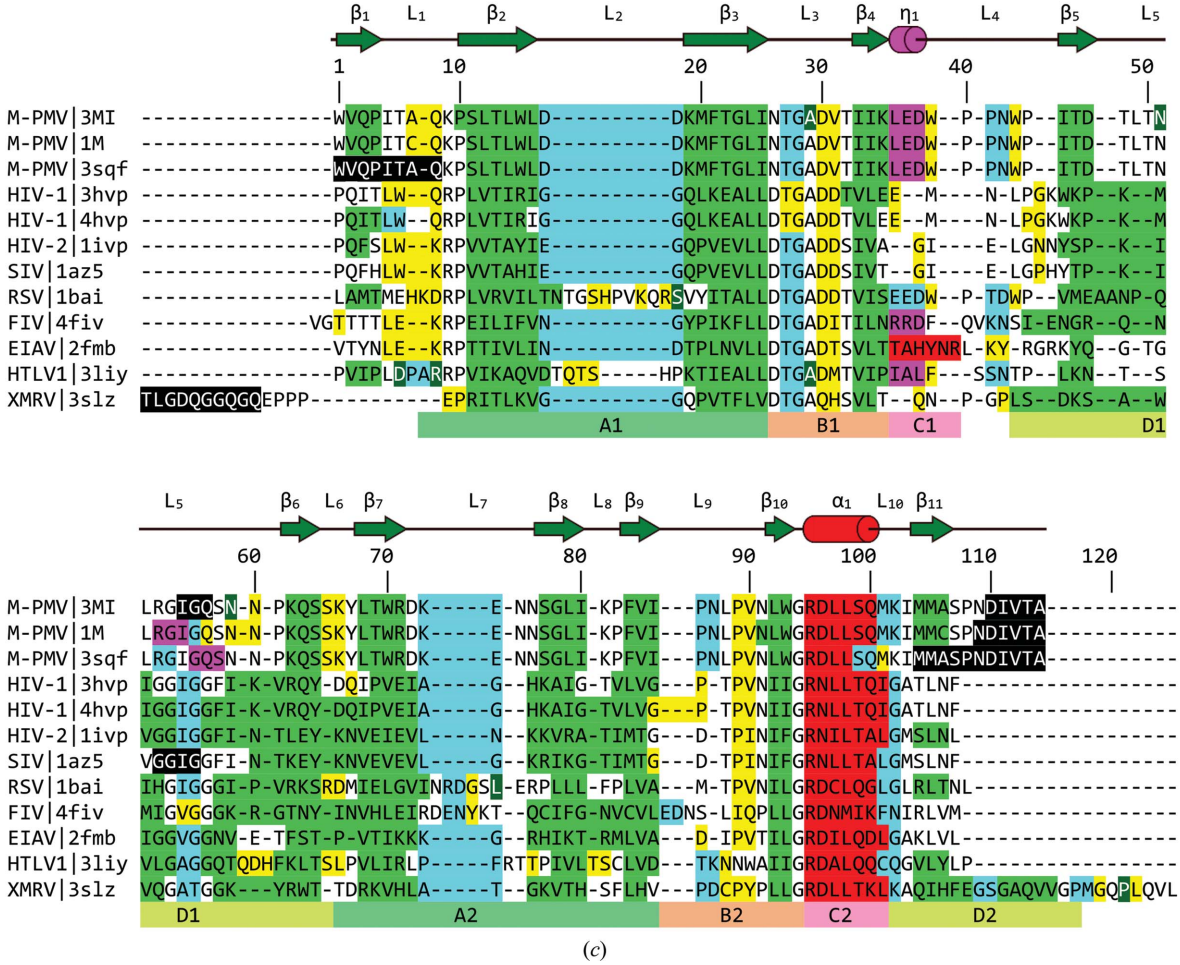
2. Materials and methods

2.1. Cloning, expression and purification

The D26N or C7A/D26N/C106A mutations were introduced into the previously described plasmid pBPS13ATG using the QuikChange Site-Directed Mutagenesis Kit (Zábranská *et al.*, 2007) and were verified by DNA sequencing. The expression of M-PMV PRs was performed in *Escherichia coli* BL21 (DE3) cells under the conditions described previously (Zábranská *et al.*, 2007). All protease



(*a*) (*b*)



(*c*)

Figure 1

(*a*) The canonical fold of retroviral protease based on EIAV PR (PDB entry 2fmb). Common structural elements in the N/C-terminal halves of the protein are indicated with colors: A1/A2 hairpins, green; B1/B2 loops, light orange; C1/C2 helices, pink; D1/D2 hairpins, lime. The side chains of the catalytic aspartates (from the DTG triads), shown in ball-and-stick representation, point towards the binding pocket. (*b*) Secondary-structure elements of the 3MI model shown in a slightly different view for better visibility. (*c*) Structural alignment of retroviral proteases. The color of the background indicates the secondary structure assigned by *DSSP*. Green denotes β-ladder, red denotes α-helix, magenta denotes 3_{10}-helix, blue denotes hydrogen-bonded turn, yellow denotes bend, dark green with a white font denotes isolated β-bridges and black denotes residues that are not present in a given model. The analysis is based on data retrieved from the Protein Data Bank (and identified by PDB code) for the HIV-1 PR apo structure (3hvp) and that with an inhibitor (4hvp), HIV-2 PR (1ivp), SIV PR (1az5), RSV PR (1bai), FIV PR (4fiv), EIAV PR (2fmb), HTLV-1 PR (3liy), XMRV PR (3slz) and M-PMV PR in a monomeric state (3sqf). M-PMV PR in a dimeric state is represented by chains *B* of 1M and 3MI. The last row represents the structural regions highlighted in (*a*). Secondary-structure elements (green arrows, β-strands, β; red cylinder, α-helix, α; magenta cylinder, 3_{10}-helix, η; black lines, loops, L) are shown above the sequences as pictograms for the present M-PMV PR 3MI model shown in (*b*).

Table 1
Data-collection and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

Structure	3MI	1MI	1M
Mutations	D26N, C7A, C106A	D26N	D26N
Inhibitor	Yes	Yes	No
Data collection			
Space group	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å)	51.60, 29.41, 85.53	49.95, 29.50, 85.39	49.99, 29.57, 85.20
β (°)	103.8	101.7	102.5
Temperature (K)	100	100	100
X-ray source	BL14.2, BESSY	BL14.1, BESSY	BL14.1, BESSY
Wavelength (Å)	0.91841	0.91841	0.91841
Resolution (Å)	48.30–1.90 (2.02–1.90)	48.91–1.64 (1.74–1.64)	48.91–1.98 (2.10–1.98)
Mosaicity (°)	0.12	0.13	0.21
$R_{\rm int}$†	0.108 (0.980)	0.043 (0.826)	0.119 (0.627)
$R_{\rm meas}$‡	0.124 (1.146)	0.050 (0.970)	0.139 (0.737)
$\langle I/\sigma(I)\rangle$	11.49 (1.53)	14.36 (1.37)	8.24 (2.03)
$CC_{1/2}$	99.7 (96.9)	99.9 (56.8)	99.7 (52.1)
No. of measured reflections	80465 (11316)	111109 (17159)	61992 (8832)
No. of unique reflections	20007 (3082)	30408 (4819)	17123 (2594)
Multiplicity	4.02	3.65	3.62
Completeness (%)	99.1 (95.9)	99.3 (98.3)	98.9 (94.3)
Wilson B factor (Å^2)	29.7	28.0	25.0
Solvent content (%)	48.0	46.5	48.2
Refinement			
No. of reflections, work/test set	19004/1001	29403/1004	16122/1000
$R/R_{\rm free}$§	0.1929/0.2354	0.1787/0.2202	0.1996/0.2606
Protein molecules in asymmetric unit	2 [one dimer]	2 [one dimer]	2 [one dimer]
No. of atoms			
Protein	1666	1701	1683
Inhibitor	60	60	—
Water	149	157	129
$\langle B\rangle$ (Å^2)			
Protein	39.69	41.83	36.01
Inhibitor	82.06	73.01	—
Water	45.07	45.31	42.57
R.m.s.d. from ideal			
Bond lengths (Å)	0.017	0.016	0.015
Bond angles (°)	1.74	1.70	1.84
Ramachandran statistics (%)			
Favored	97	96	96
Allowed	3	4	4
Outliers	0	0	0
PDB code	6s1u	6s1v	6s1w

† $R_{\rm int} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement of the intensity of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl. ‡ $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement of the intensity of reflection hkl, $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl and $N(hkl)$ is the number of observations of intensity $I(hkl)$ (the redundancy). § $R = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure factors, respectively. $R_{\rm free}$ was calculated analogously for around 1000 randomly selected reflections that were excluded from refinement.

forms were isolated from inclusion bodies by solubilization in 8 *M* urea and were renatured by stepwise dialysis against 50 m*M* Tris–HCl pH 7.0, 1 m*M* EDTA containing 0.05% β-mercaptoethanol (buffer *A*). Pure M-PMV PR monomer was obtained by batch ion-exchange chromatography on QAE-Sephadex A-25 equilibrated with buffer *A* and was characterized by sedimentation-equilibrium experiments using an analytical ultracentrifuge as described by Zábranská *et al.* (2007).

2.2. Crystallization

2.2.1. 3MI. Before the crystallization experiment, the protein was incubated for 2 h with a 1.2-fold molar excess (relative to the dimeric enzyme) of a peptidomimetic inhibitor with the sequence Pro-(O-Me)Tyr-Val-PSA-Ala-Met-Thr, where (O-Me)Tyr denotes O$^{\eta}$-methylated tyrosine and PSA denotes (3*S*,4*S*)-3-hydroxy-4-amino-5-phenylpentanoic acid. Crystallization trials were set up manually using the hanging-drop vapor-diffusion technique at 292 K by mixing 1 µl of 6.8 mg ml^{-1} protein solution (buffered with 10 m*M* Tris pH 7.4) and 1 µl reservoir solution composed of 0.1 *M* imidazole pH 6.5, 1.2 *M* sodium acetate. Single crystals appeared within a week and were harvested after six weeks.

2.2.2. 1MI. Prior to crystallization, protein solution at 5.0 mg ml^{-1} in 10 m*M* Tris pH 7.4 plus 5 m*M* TCEP was incubated for 2 h with a 1.4-fold molar excess of the inhibitor (the same as used for 3MI). Crystallization trials were set up manually using the hanging-drop vapor-diffusion technique at 292 K by mixing 3 µl protein solution and 1 µl reservoir solution. The reservoir solution consisted of 0.1 *M* sodium citrate pH 5.6, 25% propan-2-ol, 5 m*M* TCEP. The first single

crystals appeared after two days. The crystals used in the X-ray diffraction experiments were harvested after seven months.

2.2.3. 1M. Crystallization trials were set up manually using the hanging-drop vapor-diffusion technique at 292 K by mixing 3 µl protein solution (4.1 mg ml^{-1} protease in 10 m*M* Tris pH 7.4, 5 m*M* TCEP) and 1 µl reservoir solution. The reservoir solution consisted of 0.1 *M* sodium citrate pH 5.6, 15% propan-2-ol, 5 m*M* TCEP. After 2 h, the streak-seeding technique was used to transfer crystallization seeds from a crystal conglomerate to a fresh pre-equilibrated drop with the same conditions. Useful single crystals grew within two days and were harvested after four weeks.

2.3. X-ray data collection and processing

All crystals were cryoprotected in mother liquor supplemented with 25%(*v*/*v*) PEG 400. X-ray diffraction data were collected at 100 K using synchrotron radiation provided by beamlines 14.2 and 14.1 at BESSY, Berlin equipped with a MAR Research MX-225 detector (3MI) or a PILATUS 6M detector (1MI and 1M). The diffraction data were processed and scaled using the *XDS* package (Kabsch, 2010). Space groups and unit-cell and data-collection parameters are summarized in Table 1. Raw X-ray diffraction images have been deposited in the RepOD Repository at the Interdisciplinary Centre for Mathematical and Computational Modeling (ICM) of the University of Warsaw, Poland, and are available for download with the following digital object identifiers (DOIs): http://dx.doi.org/10.18150/repod.0005795 (3MI), http://dx.doi.org/10.18150/repod.6499856 (1MI) and http://dx.doi.org/10.18150/repod.7575322 (1M).

2.4. Structure solution and refinement

The structure of 3MI was determined by molecular replacement in *Phaser* (McCoy *et al.*, 2007) using the coordinates of monomeric M-PMV protease (PDB entry 3sqf, molecule *A*; Gilski *et al.*, 2011) as a search model. The program found two copies of the model in the asymmetric unit in space group $P2_1$ that were intertwined to form a tight homodimer. For the other two structures, 1MI and 1M, chain *A* from the refined structure of 3MI was used to create the starting model. Structural refinements were carried out in *REFMAC*5 (Murshudov *et al.*, 2011) from the *CCP*4 suite (Winn *et al.*, 2011). *Coot* (Emsley *et al.*, 2010) was used for manual rebuilding and adjustments of the models in electron-density maps between rounds of automatic refinement, as well as to build the inhibitor molecules and to validate the solvent structure. Since the starting model was truncated at both termini, the missing residues were added according to the electron-density maps. Fragments of the flap loops and several side chains had to be removed from the model because of a lack of evidence in the electron density (Supplementary Fig. S1). Six TLS segments per polypeptide chain were defined for each protein model according to analysis by the *TLSMD* server (Painter & Merritt, 2006). Stereochemical restraints for the amide bonds to the PSA residue (Val-PSA and PSA-Ala) of the inhibitor were created using *JLigand* (Lebedev *et al.*, 2012). The final models also included 129, 157 and 149 water molecules for the 1M, 1MI and 3MI structures, respectively. The refinement statistics are given in Table 1.

3. Results and discussion

3.1. Conformation of the M-PMV PR dimer

3.1.1. Overall conformation of the dimers. The polypeptides used in the present crystallization experiments correspond to the Trp1–Ala114 sequence of M-PMV PR (with the exception of the active-site D26N mutation and, where applicable, the C7A/C106A mutations) and are combined into dimers in all structures. All of the dimeric forms of M-PMV PR studied in this work have the characteristic fold of classical retroviral proteases. The terminal residues, which were missing in the monomeric structure, form the canonical antiparallel β-sheet (Fig. 2). This interdigitated β-sheet is the main part of the dimerization interface. There are still 5–6 C-terminal residues that do not participate in β-sheet formation and are not defined in the electron density.

The remaining β-strands form one β-sheet within the protomer core (Fig. 1). In both structures with the inhibitor (1MI and 3MI) the ligand molecule occupies the active-site cleft, running with one polarity across the approximately C_2-symmetric dimeric binding crevice between the active site and the flap arms. Despite the presence of the ligand, the flap loops are not well ordered, with some residues that are not defined in the electron density (Supplementary Fig. S1). Interestingly, the structure without the inhibitor (1M) has one entire flap loop well defined while the other one is missing (residues Leu49–Asn59).

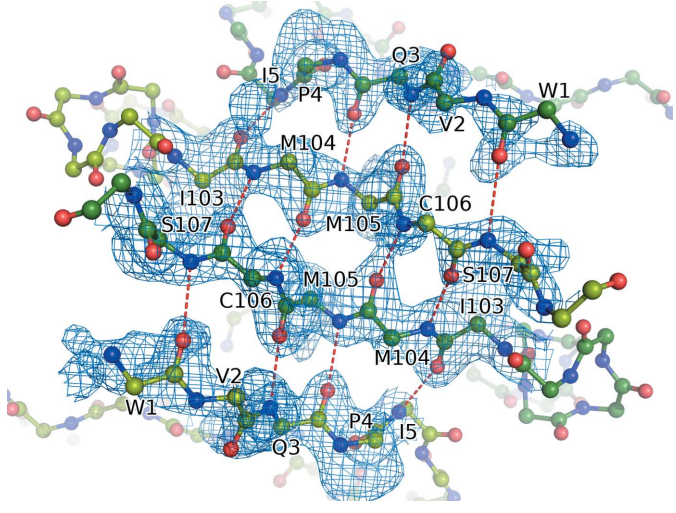


Figure 2
The antiparallel dimerization β-sheet of M-PMV PR consists of the C- and N-terminal peptides of chain *A* (green) and chain *B* (light green). The $2mF_o - DF_c$ electron-density map is contoured at 1.0σ for the indicated residues of both chains. Red dashed lines represent hydrogen bonds within the β-sheet. This illustration is based on the 1MI structure, with side chains omitted for clarity.

3.1.2. Intersubunit β-sheet. The most intimate interactions between the protomers within a dimer occur between the N- and C-termini of both chains forming the interface β-sheet (Fig. 2). This antiparallel pattern consists of alternating chains in the following order: Trp1–Ile5 of chain *A*, Ile103–Ser107 of chain *B*, Ile103–Ser107 of chain *A* and Trp1–Ile5 of chain *B*. According to *DSSP* (Touw *et al.*, 2015), the Val2–Pro4 and Met104–Ala/Cys106 residues are part of the β-ladder. However, the flanking residues Trp1 and Ile5 and Ile103 and Ser107, despite departing from strict geometrical requirements for the β conformation, still preserve the hydrogen-bond pattern characteristic of a β-sheet. The electron density of the remaining C-terminal residues becomes weaker beyond Pro108; for Asn109 only the main-chain atoms of molecule *A* are visible in 3MI, and there is no clear evidence for the remaining residues in all chains. The *B* factors for the residues building the β-sheet, with average values of 55.1, 65.3 and 54.5 Å^2 for 3MI, 1MI and 1M, respectively, are significantly higher compared with the whole protein, where the average *B* factors for all atoms are 39.7, 41.8 and 36.0 Å^2, respectively.

3.1.3. The potential for intersubunit disulfide-bond formation. In 3MI, the mutated C7A residue is located at the top of loop L_1 [Fig. 1(*b*)], which is exposed to solvent with no proximity to other (potential) sulfhydryl groups. The C106A residues of the two polypeptide chains are close to each other, but the distance between their C^β atoms of 5.8 Å is outside the range of typical C^β–$C^{\beta\prime}$ distances in cystine residues (Salam *et al.*, 2014). However, a small conformational adjustment would be sufficient to bring the –SH groups of unmutated Cys106 residues together for the formation of a disulfide bond.

In the 1MI and 1M structures, in which the cysteine residues are present in an unmutated form, the distances between the S atoms of the Cys106 residues are 3.3 and 3.4 Å, respectively; therefore, there is no disulfide bond formed. This was to be expected as a reducing agent (TCEP) was used during protein preparation and crystallization. The C^β–$C^{\beta\prime}$ distance in 1M is the same as in 3MI (5.8 Å) but is distinctly shorter in 1MI (5.1 Å). The mutation at the C7 position has no effect on the geometry of the L_1 loop. Additionally, in the 1MI model the side chains of these cysteine residues are modeled in two alternative conformations.

3.1.4. Conformation of the flap loops. The flaps in dimeric retropepsins are usually considered to make an important contribution to dimerization interactions (in addition to the N- and C-termini and the active-site loops), particularly in active-site complexes, where they are lowered and locked onto the bound substrate/inhibitor. In the present inhibitor complexes of M-PMV PR, however, this locking does not happen. In the two structures with inhibitor in the active site, the flaps are partly disordered and do not form any dimeric interactions. There are stretches of oligopeptides that are not visible in the electron density: Gly54–Ser58 of chain *A* and Ile55–Gln57 of chain *B* in 3MI, and Gly54–Ser58 of chain *A* and Gly54–Gly56 of chain *B* in 1MI. In addition, the side chains of Arg53 in both structures, as well as of Leu52 of chain *A* in 3MI, have no corresponding electron density and were therefore truncated at C^β. The structure 1M, without inhibitor, is even more peculiar because the entire flap loop of one protomer (*B*) is clearly visible in the electron density (Fig. 3) while the other one is absent (Leu49–Asn59) (Supplementary Fig. S1). The typical β-sheet interactions reinforcing the β-hairpin structure of flap B are disturbed at Thr48. The next juxtaposed pair (Asn51 and Asn59) form a hydrogen-bond pattern suitable for an antiparallel β-sheet, but the following residues (Arg53–Ile55) form a turn with a conformation corresponding to a

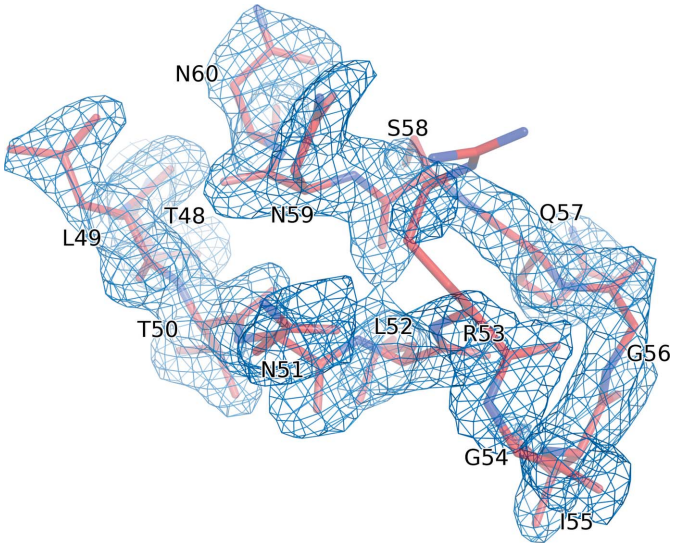


Figure 3
The flap of subunit *B* of the 1M structure shown in stick representation (red) in the corresponding $2mF_o - DF_c$ electron density contoured at 1.0σ.

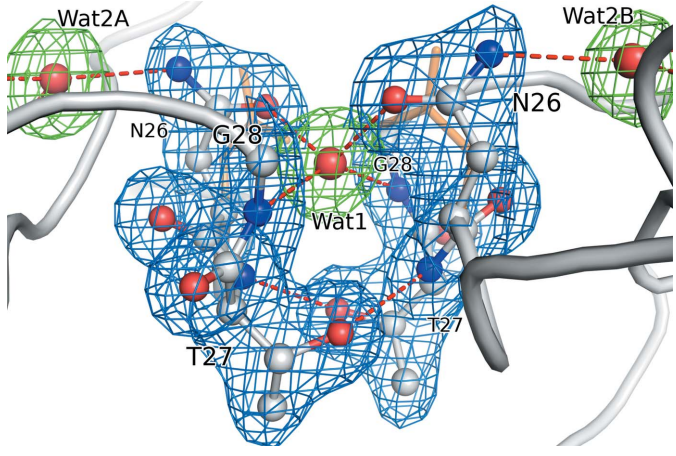


Figure 4
The catalytic loops of 3MI with a water molecule (Wat1) uniquely located between the Asn26 residues. The protein molecules are shown in $2mF_o - DF_c$ electron density (blue) contoured at 1.0σ. Water molecules Wat1, Wat2A and Wat2B are shown in $mF_o - DF_c$ OMIT electron density (green) contoured at 3.0σ. Red dashed lines indicate hydrogen bonds. The Wat1 molecule is tetrahedrally coordinated. The side chains of the Thr27 residues (at the bottom) form hydrogen bonds called the 'fireman's grip'. The inhibitor molecule has been omitted for clarity. Side chains of the catalytic aspartates from HIV-1 PR (orange, semi-transparent) are shown for comparison with a retropepsin active site without a Wat1 molecule (HIV-1; PDB entry 4hvp).

short 3_{10}-helix (Fig. 1). Despite being lowered over the active site, the flap does not form any new, specific interactions with the core of the protein. The interaction interface is created entirely by hydrophobic side chains. In the space where the absent flap of chain *A* could be expected, seven well defined water molecules are located, with no apparent correlation with the atom positions generated by pseudodyad symmetry from the other flap. These water molecules do not form any specific hydrogen bonds with the protein, in analogy to the situation described for the visible flap B.

3.1.5. Architecture of the active-site loops. The active site consists of the triad Asn26-Thr27-Gly28, in which the first asparagine is a point mutation of the catalytic aspartate. The characteristic 'fireman's grip' of hydrogen bonds (in which the threonine side chains and the main-chain NH groups are connected in a symmetric fashion across the dimer) is preserved in all structures. The rotamer of the Asn26 side chains was determined based on an unambiguous hydrogen-bond network, especially with a water molecule (Wat1) that is buried between them (Fig. 4). A list of hydrogen bonds formed by the Asn26 residues is presented in Supplementary Table S1. This network of hydrogen bonds in the active site of M-PMV PR (described in detail below) differs significantly from the characteristic pattern that is usually observed in retropepsins and indeed in all aspartic proteases. The most prevalent system of hydrogen bonds observed in the active site of retropepsins (excluding nonspecific water molecules in apo forms) includes interaction of the ('inner') $O^{\delta1}$ atom of the catalytic aspartate residue with the NH donor of the glycine residue (the last residue of the catalytic triad) and a bridge with the $O^{\delta1}$ atom of the complementary aspartate, formed with the participation of a nucleophilic water molecule (hereafter referred to as WatC) in the apo structures or a hydroxyl group or similar surrogate in inhibitor complexes. The side chains of the Asn26 residues are moved apart and twisted around the $C^{\beta}-C^{\gamma}$ bond out of the usual almost coplanar conformation. Thus, the usual hydrogen-bond link between the $O^{\delta1}$ atom of Asn26 (or aspartate in wild-type retropepsins) and NH from Gly28 is now bridged by the Wat1 water molecule in both protomers. As mentioned above, this very well defined water molecule (with $2mF_o - DF_c$ electron density visible up to 3.3σ, 2.6σ and 3.6σ with *B* factors of 28, 38 and 26 Å^2 in 3MI, 1MI and 1M, respectively) is perfectly coordinated by the main-chain NH donors of the Gly28 residues and by the side-chain $O^{\delta1}$ acceptors of the Asn26 residues at the very interface between the two subunits. The pseudodyad symmetry of the active site is therefore preserved. Such a water molecule has not previously been seen at this location in any retropepsin structure, as confirmed by a detailed analysis of active-site water molecules found in retropepsin structures in the Protein Data Bank (Berman *et al.*, 2000; supporting information, Section S1) or, even more generally, in any aspartic protease for that matter. A somewhat similar situation is observed in the monomeric structure of M-PMV PR, in which a similar water molecule bridges the interaction between the Asn26 $O^{\delta1}$ and Gly28 NH atoms. However, in the monomeric state the catalytic triad does not form the active site and is exposed to solvent; therefore, this water molecule (corresponding to Wat1) can be treated as part of the hydration layer.

The Asn26 active-site residues of M-PMV PR are engaged in a distinct hydrogen-bond network including a total of three water molecules: Wat1, as described above, and two additional molecules (Wat2A and Wat2B) symmetrically located at hydrogen-bonding distances from the NH_2 groups of the side-chain amides. These water molecules have excellent definition in the electron density (Fig. 4) and correspondingly good *B* factors (Wat2A, 26–41 Å^2; Wat2B, 22–25 Å^2). Each of the additional water molecules (Wat2A and Wat2B) is also coordinated by the

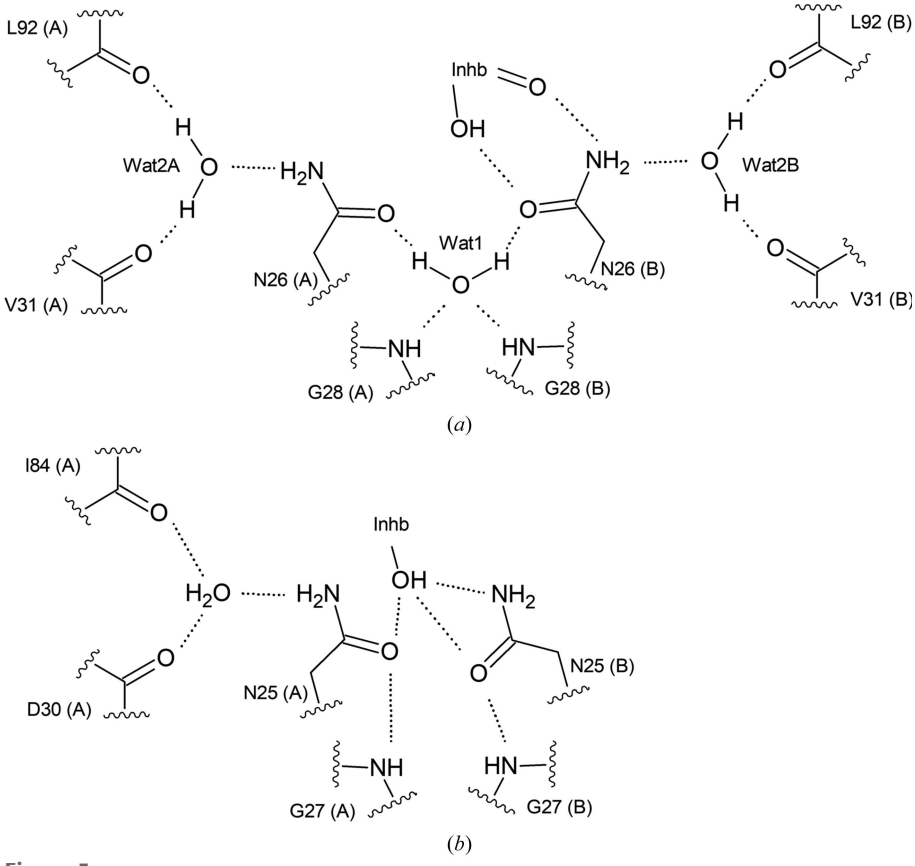


Figure 5
(*a*) A scheme showing the hydrogen-bond interactions around the active-site Asn26 (mutation of Asp26) residues in 1MI. This hydrogen-bond network is present in all M-PMV PR dimers studied in this work. Water molecule Wat1 at the bottom of the active site is absolutely unprecedented among all retropepsins. The water molecules Wat2A and Wat2B that complete the hydrogen-bond pattern have partial analogs in two HIV-1 PR structures, for example PDB entry 5kr1 (*b*).

carbonyl groups of Val31 and Leu92 (Fig. 5). Analogous water molecules with the same structural role are also present in the monomeric state of the protein. Among the 707 models of other retroviral proteases in the PDB as of January 2019, only two structures [PDB entries 5kr1 (Liu *et al.*, 2016) and 1n49 (Prabu-Jeyabalan *et al.*, 2003)], both of HIV-1 PR in complex with asymmetric inhibitors, show some similarity to the hydrogen-bonding network involving the Wat2A and/or Wat2B sites of M-PMV PR. In these examples, however, only one water molecule per dimer is present, coordinated by the active-site side chain and two carbonyl groups of the main chain (Fig. 5). The model with PDB code 5kr1 shares the active-site Asp→Asn mutation with the present M-PMV PR variants; however, in PDB entry 1n49 the active site is unchanged, which suggests that this unusual hydrogen-bonding pattern is not an artifact of the Asp→Asn mutation.

3.1.6. The interface water molecule. The apo forms of aspartic proteases, but not their complexes with peptidomimetic inhibitors, have a nucleophilic water molecule (WatC, catalytic water) tightly hydrogen-bonded between the carboxylic groups of the two catalytic aspartates. In inhibitor complexes this water molecule is absent, and its role is usually played by a structural element (for example a hydroxyl group) of the inhibitor. In complexes formed by retroviral aspartic proteases (but not by cellular proteases such as pepsin, which have only one prominent flap) with peptidic inhibitors, there is a different characteristic water molecule (WatI, inhibitory water) at the interface between the inhibitor and the flap loops. WatI is tetrahedrally coordinated by two C═O groups of the inhibitor (acting as hydrogen-bond donors in these interactions) and two main-chain NH groups from the flap loops (acting as hydrogen-bond acceptors in these interactions).

In both structures with the inhibitor (3MI and 1MI) there are water molecules in a similar position to WatI. However, the hydrogen-bonding pattern is different. The water molecule is coordinated by only one C═O group of the inhibitor: either that of Val3 in 3MI or that of Ala5 in 1MI. There are no visible flap residues to complete the tetrahedral hydrogen-bonding pattern in 3MI, and in 1MI the NH group that forms a hydrogen bond to this water molecule is the last modeled residue of the flap, flanking a gap, and thus is poorly defined by electron density. Because of the poor quality of the electron density in this area, treating these water molecules as the classical WatI interface molecules would be an over-interpretation. In fact, the electron density modeled as these water molecules could be even a trace of a fragment of the disordered flap loops. In the 3MI and 1MI structures the hydroxyl group of the PSA residue (the γ-amino acid that serves as the inhibitory hotspot of the inhibitor molecule) penetrates the active site, forming a hydrogen bond to one of the Asn26 residues (Supplementary Table S1). Therefore, there is no space for WatC in these structures. As described above, Wat1 is buried below the crest of the catalytic Asn26 side chains in all three structures and therefore cannot take the role of WatC. In 1M the side chains of Asn26 are coordinated by seven water molecules (including Wat1), but only Wat1 is hydrogen-bonded between these residues. However, for the same reasons as above, it does not meet the criteria for WatC.

Table 2
R.m.s.d. values (Å) for C$^{\alpha}$ alignments of M-PMV PR protomers, with the chain ID in parentheses.

The number of superposed C$^{\alpha}$ pairs is given in square brackets.

	1M (*B*)	1MI (*A*)	1MI (*B*)	3MI (*A*)	3MI (*B*)
1M (*A*)	0.15 [87]	0.20 [89]	0.25 [91]	0.18 [90]	0.36 [92]
1M (*B*)		0.19 [88]	0.15 [93]	0.21 [93]	0.35 [98]
1MI (*A*)			0.17 [88]	0.17 [96]	0.33 [99]
1MI (*B*)				0.21 [87]	0.32 [95]
3MI (*A*)					0.29 [101]

3.2. Structural comparisons

3.2.1. The present protomers versus monomeric M-PMV PR. There are two main differences between the previously reported monomeric form of M-PMV PR (Gilski *et al.*, 2011) and the present dimerized protomers: (i) the conformation of the flap and (ii) the involvement of the N- and C-termini in quaternary-structure formation. The monomeric model starts with Leu9 and ends with Ile103, while the dimeric models comprise all of the N-terminal residues and end at Pro109 (or Asn110) at the C-terminus; the last six (five) residues are still disordered. The terminating residues of the monomeric model (Leu9, Lys102 and Ile103) diverge from the positions occupied in the dimeric forms.

Counterintuitively, the flap of the monomer is better defined in the electron-density maps than in the dimer, where additional stabilizing factors should be provided by the inhibitor molecule and/or mutual flap–flap interactions. The visible flap stems of 3MI and 1MI seem to follow the C$^{\alpha}$ traces of other inhibitor complexes of retropepsins, with an r.m.s. deviation of 3.13 Å for 16 C$^{\alpha}$ pairs in 1MI flap B and the most similar retropepsin (similarity assessed upon r.m.s. deviation for all-C$^{\alpha}$ superposition; Table 3) from Rous sarcoma virus (PDB entry 1bai, chain *B*; Wu *et al.*, 1998). In comparison with the flap found in the monomeric state of M-PMV PR (PDB entry 3sqf, chain *B*; Gilski *et al.*, 2011), the same flap B of the 1MI model shows a higher r.m.s. deviation of 3.63 Å (17 C$^{\alpha}$ pairs).

The flap conformation of the 1M model is exceptional as there is high asymmetry between the two flap loops. One of the loops (A) is almost completely disordered (residues Leu49–Asn59 are missing), while the other is fully ordered despite the absence of an inhibitor. The conformation of flap B includes a 3_{10}-helix, Arg53–Ile55, which serves as part of the turn of the flap. It is different from the 3_{10}-helix reported for the monomeric form of M-PMV PR (Gln57–Asn59), in which the helix contracts the length of the flap. The overall r.m.s. deviation for 20 superposed C$^{\alpha}$ atoms of flap B of model 1M and flap A of the monomeric molecule is 4.28 Å.

The superior definition in the electron density of the flap element of protomer B (especially in the 1M structure) is related to crystal-packing effects, as discussed in Section 3.4.

Table 3
R.m.s.d. values (Å) for C^{α} superpositions of selected retroviral protease models in the PDB (identified by virus acronym and PDB code) onto the present M-PMV PR dimers.

The row marked 1M (*B*) presents r.m.s. deviations for one subunit only (chain *A* of the corresponding proteases was used). The number of superposed C^{α} pairs is given in square brackets.

	HIV-1	HIV-1	HIV-2	SIV	SIV	RSV	RSV	FIV	EIAV	HTLV-1	XMRV	M-PMV
	3hvp	4hvp	1ivp	1az5	1yth	2rsp	1bai	4fiv	2fmb	3liy	3slz	3sqf
1M (*B*)	2.00 [79]	2.24 [87]	1.95 [79]	2.25 [87]	2.16 [87]	1.11 [81]	2.09 [92]	2.46 [94]	1.39 [73]	1.66 [82]	1.17 [69]	0.246 [70]
1M	2.58 [156]	2.21 [157]	2.10 [146]	2.23 [129]	1.74 [133]	1.21 [163]	1.50 [147]	1.58 [136]	1.46 [142]	1.73 [163]	1.22 [115]	—
1MI	2.57 [163]	1.45 [139]	1.99 [139]	2.00 [138]	1.49 [140]	1.20 [153]	1.48 [158]	1.44 [143]	1.62 [156]	1.90 [172]	1.04 [126]	—
3MI	2.14 [147]	1.96 [152]	2.28 [147]	2.95 [142]	1.87 [149]	1.15 [168]	1.49 [151]	1.46 [140]	1.31 [142]	1.64 [161]	1.05 [126]	—

3.2.2. Comparison of the present protomers. The results of C^{α} alignment of the six protomers from the three dimers described in this work are shown in Table 2. The most distinct protomer is chain *B* of 3MI, with r.m.s. deviations in the range 0.29–0.36 Å and the highest similarity to chain *A* of the same dimer. The lowest r.m.s. deviation (0.15 Å) is calculated for the subunits of 1M; however, this value is artificially lowered by the artificially low atom count (the flap of one protomer is missing). The largest distances between equivalent C^{α} atoms are invariably found within the flap structures.

3.2.3. M-PMV PR dimers versus dimers of other retropepsins. The overall fold of the M-PMV PR dimers is similar to that found in other retroviral PRs (Fig. 6), with a high similarity of the core part (A1 except for the β-turn, B1, B2 and C2; see Fig. 1) and the largest deviations in the outer layer of secondary-structure elements (A2, C1 and D1). The C^{α} r.m.s.d. values (Table 3) show the highest similarity to Rous sarcoma virus (RSV) PR and (with a low C^{α} count) to Xenotropic murine leukemia virus-related virus (XMRV) PR. Structure-based sequence alignment (Fig. 1) also demonstrates this similarity. The secondary structures were assigned according to *DSSP* (Touw *et al.*, 2015). The models for comparison were selected as follows (PDB codes are given in parentheses): HIV-1, apo form (3hvp; Wlodawer *et al.*, 1989), HIV-1, PR inhibitor complex (4hvp; Miller, Schneider*et al.*, 1989), HIV-2 PR (1ivp; Mulichak *et al.*, 1993), SIV PR, apo form (1az5; Rose *et al.*, 1998), RSV PR (1bai; Wu *et al.*, 1998), FIV PR (4fiv; Kervinen *et al.*, 1998), EIAV PR (2fmb; Kervinen *et al.*, 1998), HTLV-1 PR (3liy; Satoh *et al.*, 2010) and XMRV PR (3slz; Li, Gustchina *et al.*, 2011). The first short β-ladder β_1 is part of the dimerization interface. The following loop L_1 is extended in M-PMV PR by Cys7 compared with other retropepsins, except for RSV PR and HTLV-1 PR, which also have longer loops L_1. Cys7 is artificially mutated to alanine in 3MI. The highly conserved RP pair is replaced by Lys9–Pro10 in the M-PMV PR sequence, preserving the chemical character of this dipeptide. A mutation in this region is also present in XMRV PR as EP, with a complete change of the electrostatic charge of the first element. The A1 hairpin shares its conformation with other retropepsins, except for RSV PR and HTLV-1 PR, where an extended loop is formed instead of the turn. The L_3 loop contains the catalytic triad and is highly conserved in all retropepsins. The D26N mutation introduced into the M-PMV PR sequence causes no significant changes to the main chain compared with other retroviral PRs. The C1 helix of the common template is present in M-PMV PR as a short 3_{10}-helix, similar to those found in FIV PR and HTLV-1 PR. The conformation of the EED turn found in RSV PR (PDB entry 1bai, in complex with an inhibitor) is also close to a 3_{10}-helix. This feature is absent in the proteases from HIV, SIV and XMRV, while in EIAV PR there is a fully developed α-helix. The following element of the template, the D1 hairpin, corresponding to the flap loops, is an area in which M-PMV PR exhibits a conformational singularity. For the structures with inhibitor (1MI and 3MI), the stems of the flaps that are visible in the electron-density maps follow a similar trace as in other PR complexes, in contrast to the flap of the monomeric form, which is very different. The β-ladder of the flap hairpin is barely preserved, but again this may be associated with the low quality of the electron-density maps in this area. The conformation of the flap in 1M, which has an empty active site, differs from that in the remaining M-PMV PR structures and in other retroviral PRs (Fig. 7). It reaches deeper into the active site (Fig. 6) than is typical of PR complexes with inhibitors, not to mention the apo forms [for example HIV-1 PR (PDB entry 3hvp), RSV PR (PDB entry 2rsp) and SIV PR (PDB entry 1az5)], in which the flaps (if not disordered) are elevated in order to open access to the active site. The flap β-ladder (β_5–β_6) is preserved only at the base of the D1 fragment. The turn of the flap loop, which is usually composed of the conserved residues Ile55 and Gly56, is extended (preceded) by the 3_{10}-helix (Arg53–Ile55). A 3_{10}-helix within the flap is also present in the monomeric M-PMV PR structure, but it is shifted to Gly57–Ser59, *i.e.* to the C-terminal strand of the loop, and plays a different role, not as part of the turn but as an element compacting the geometrical extent of the flap. Loop L_6 separates the D1 and A2 hairpins. A similar structure is found in RSV PR, HTLV-1 PR and XMRV PR. HIV-1 PR contains this feature in the apo form but not in inhibitor complexes. When this loop is absent, the β-strand (β_6) continuously turns into β_7. The next segment of the template, the A2 hairpin, which in M-PMV PR is mainly represented by the β_7–β_8 hairpin, is similar to analogous hairpins in other retropepsins. Only RSV PR and FIV PR differ in this respect, with longer loops connecting the β-strands. The B2 loop clearly replicates the structural elements of B1: a turn, a loop (L_3) and a short β-strand (β_{10}) that leads to C2. The canonical C-terminal α-helix (α_1) is longer than in the monomeric state, thus making M-PMV PR dimers even more similar to other retropepsins in this respect.

It is most likely that this change is caused by the stabilization effect of the dimerization β-ladder formed by the C-terminal peptide (β_{11}), which is disordered in the monomeric state. The ladder is common to all retropepsins and is a remnant of a D2 hairpin, which is fully formed only in XMRV PR.

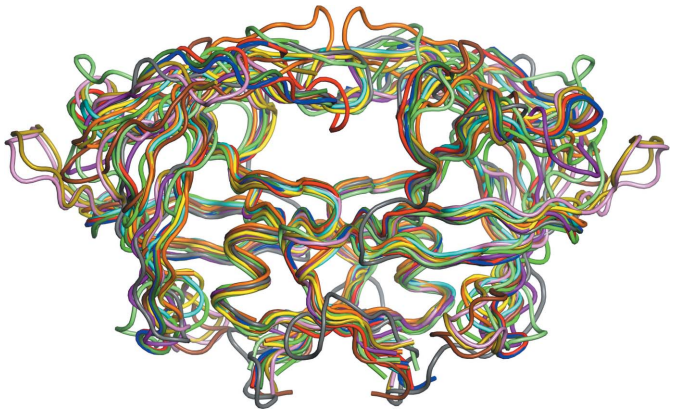


Figure 6
Structural superposition of dimeric retroviral proteases: red, M-PMV, 1M; deep green, M-PMV, 1MI; blue, M-PMV, 3MI; orange, HIV-1, apo form (PDB entry 3hvp); yellow, HIV-1, inhibitor complex (PDB entry 4hvp); cyan, HIV-2, inhibitor complex (PDB entry 1ivp); green, SIV, apo form (PDB entry 1az5); olive, RSV, apo form (PDB entry 2rsp); pink, RSV, inhibitor complex (PDB entry 1bai); brown, FIV, inhibitor complex (PDB entry 4fiv); purple, EIAV, inhibitor complex (PDB entry 2fmb); lime, HTLV-1, inhibitor complex (PDB entry 3liy); gray, XMRV, inhibitor complex (PDB entry 3slz).

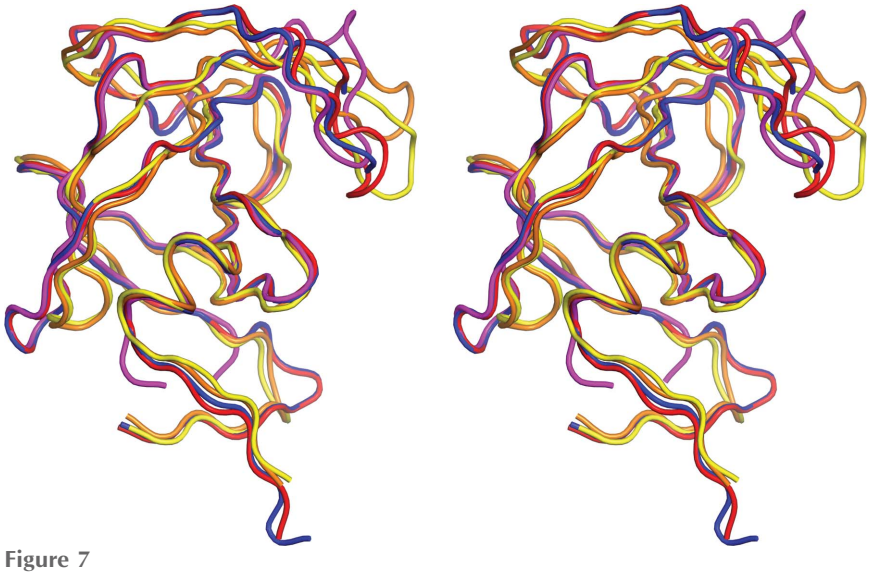


Figure 7
Stereoview of superposed chains of M-PMV PR subunits [magenta, monomeric form (PDB entry 3sqf, chain *A*); red, dimeric apo form (1M, chain *B*); blue, dimeric form, inhibitor complex (3MI, chain *B*)] together with subunits of HIV-1 PR [yellow, inhibitor complex (PDB entry 4hvp, chain *A*); orange, apo form (PDB entry 3hvp)]. The most prominent divergence is found within the D1 hairpin fragment, *i.e.* the flap. Despite a gap at the tip of the flap, the 3MI model follows the trace of HIV-1 PR, which represents the classic fold of retroviral proteases in inhibitor complexes. The 1M model is significantly different; there is a kink in the flap, formed by a 3_{10}-helix, that points it down towards the active site. The 3_{10}-helix of monomeric M-PMV PR (magenta) is also clearly visible; however, it is shifted to a different segment of the flap sequence, pointing up in this illustration. The *A* chain of 1MI is not shown as it is practically identical to the 3MI structure.

3.3. Conformation and enzyme interactions of the inhibitor molecules

3.3.1. Description of inhibitor conformation. The electron-density maps give clear evidence of the presence of the inhibitor; however, a precise conformational analysis is not fully warranted. Polder maps calculated according to Liebschner *et al.* (2017) (Fig. 8) confirm the presence of the inhibitor in the 1MI model beyond any doubt, with correlation coefficients CC(1,2) = 0.57, CC(1,3) = 0.85 and CC(2,3) = 0.51, *i.e.* with CC(1,3) distinctly larger than 0.8 and the remaining two coefficients. The results for the 3MI model, while still indicating the presence of a ligand in the active site, are less conclusive, with CC(1,2) = 0.67, CC(1,3) = 0.80 and CC(2,3) = 0.68. The appearance of bulky difference electron density and the overall similarity to the 1MI model led us to the decision to build the inhibitor in 3MI in the same manner as in 1MI.

The highly asymmetric and characteristic sequence of the inhibitor molecule, Pro-(O-Me)Tyr-Val-PSA-Ala-Met-Thr, was used to guide the proper position and orientation of the entire ligand molecule, especially in the 1MI complex. There are three residues with distinctly bulky side chains: (O-Me)Tyr with a *p*-methoxybenzyl group, Met with a 2-(methylsulfanyl)-ethyl group and PSA with a benzyl group. The distance between the two phenyl-containing groups is significantly shorter than the distance from the sulfur-containing Met side chain to the PSA side chain. These assumptions guided the modeling of the inhibitor with one polarity only. Attempts to build the inhibitor in the opposite direction (or as a superposition of two orientations) invariably resulted in a worse fit to the electron density, larger deviations from ideal geometry and a higher $R_{\rm free}$. In the final models it was not possible to build the side chains of the N-terminal proline residue. The overall quality of the electron density around the inhibitor molecule is significantly better in the 1MI structure than in that of 3MI. It is important to stress that while the general placement of the inhibitor in the active site is unquestionable, the conformation of the individual inhibitor residues, especially away from its central part, should be treated as tentative, with less satisfactory support from experimental electron density.

The main interactions between the inhibitor and the protein involve hydrogen bonds between their main-chain atoms. These sequence-independent interactions are common in inhibitor complexes of retroviral PRs (Wlodawer & Gustchina, 2000). The list of inhibitor–protein hydrogen bonds also includes the side chains of the active-site Asn26 residues (which form contacts with the OH/C=O groups of the central part of the inhibitor) and the side chains of the Asp30 residues [Asp30/*A*

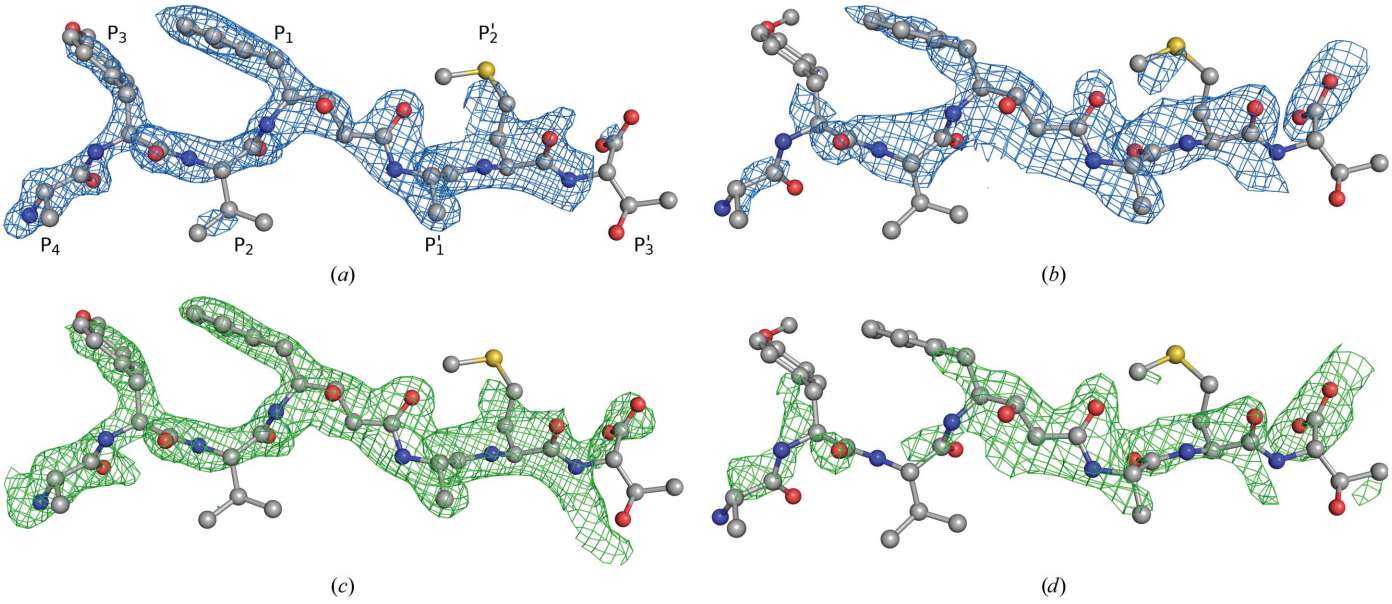


Figure 8
Conformation of the inhibitor molecule in 1MI (*a*) and 3MI (*b*) in $2mF_o - DF_c$ electron-density maps (blue) contoured at the 1.0σ level. For reference, $mF_o - DF_c$ polder electron-density maps (green) are shown at the 2.7σ contour level for 1MI (*c*) and 3MI (*d*).

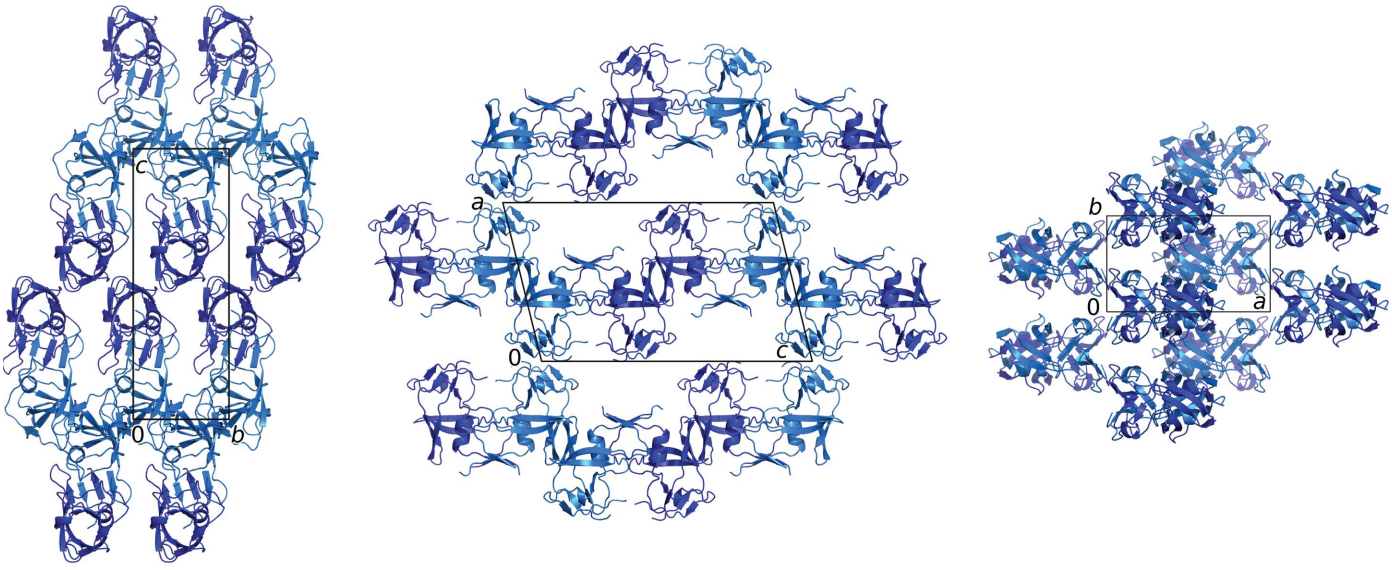
interacting with the NH group of (O-Me)Tyr and Asp30/*B* interacting with the C═O group of the threonine].

To describe the positions of peptidic substrate/inhibitor residues and their corresponding binding pockets in a protease framework, the nomenclature of Schechter & Berger (1967) is usually used. The inhibitor residues are counted from the scissile bond towards the N-terminus as P_1, P_2, ..., P_n and towards the C-terminus with primed labels as P_1', P_2', ..., P_n'. The corresponding enzyme subsites are labeled S_1, S_2, ..., S_n and S_1', S_2', ..., S_n'. The first three residues of the inhibitor (P_4, P_3 and P_2) maintain contacts with the main chain of subunit *B* (Asn51, Arg53 and Asp30). The PSA residue is located, as designed, close to the active site. The hydroxyl group at the nonscissile junction is positioned within hydrogen-bonding distance of the Asn26 side chain of protomer *B*. The PSA residue is denoted as P_1 because its side chain, the benzyl group, is located on the N-terminal side of the nonscissile surrogate (C^{α}—C^{β}) of the peptide bond and fits into site S_1. Owing to the length of the main chain of the PSA residue, which is longer than a common amino-acid residue, P_1' does not fit into the S_1' pocket, and thus alanine is used in this position. In the canonical model of retropepsin–peptidomimetic inhibitor interactions, the carbonyl groups of P_2 and P_1' coordinate the interface water molecule WatI. However, in the present inhibitor the distance between these groups (~6.8 Å) is too long and only one of them can maintain this interaction: either alanine (P_1') in 1MI or valine (P_2) in 3MI. The last two residues of the inhibitor (P_2' and P_3') interact with the main chain of subunit *A* (Gly28, Asp30, Arg53 and Asn51). Both the inhibitor and flap residues show high temperature factors. The average (main-chain) *B* factors in the 1MI structure are 68 Å^2 for the inhibitor, 74 Å^2 for flap A and 41 Å^2 for flap B. The corresponding values in the 3MI structure are 76, 77 and 54 Å^2, respectively.

3.3.2. Polarity of inhibitor binding and the question of twofold disorder of the protein dimer. In both complexes the inhibitor was fitted in one direction in the electron-density maps, as justified above. The unique polarity of the inhibitor molecule implies asymmetry of the subunits, *i.e.* deviations from perfect C_2 symmetry of the protease dimer. The asymmetry is clear especially in the flaps, which have visibly different electron density, and in the crystal packing, where the contacts with symmetry-related molecules differ in each subunit of the protein. The apo structure without inhibitor (1M) crystallized isomorphously in the same space group and with the same unit-cell parameters. It also exhibits clear asymmetry between the subunits. For instance, one flap is fully visible (chain *B*) while the other one is absent.

3.4. Crystal packing and molecular interactions

All of the crystals studied in this work exhibit the same crystal packing, with a tight homodimer in the asymmetric unit but otherwise with limited intermolecular contacts. Overall, the two protomers have a similar pattern of contacts with symmetry-related molecules, with one exception: an additional contact between chain *B* (in all models), where the base of the flap (Thr46 and Asp47) is in close proximity to loop L_9 (Asn87) from a symmetry-related molecule [Fig. 9(*b*)]. This exception correlates with the better definition of the flaps in chains *B* compared with chains *A* in all described dimers. According to *PISA* calculations (Krissinel & Henrick, 2007), the surface area of the dimerization interfaces is very similar for the two models with the inhibitor (1381 Å^2 for 1MI and 1372 Å^2 for 3MI) and is slightly larger for 1M at 1425 Å^2. These values are significantly larger than the intermolecular interaction area of the monomeric form of M-PMV PR (PDB entry 3sqf), which was calculated as 796.6 Å^2. A very low



(*a*) (*b*) (*c*)

Figure 9
Crystal packing of 3MI viewed down [100] (*a*), [010] (*b*) and [001] (*c*). Blue and light blue represent chains *A* and *B*, respectively.

Matthews volume (1.71 Å^3 Da^{-1}) and solvent content, of merely 28.1%, was reported for the crystals of monomeric M-PMV PR. In contrast, the crystals of dimeric M-PMV PR have a rather typical Matthews volume of ∼2.4 Å^3 Da^{-1} and solvent content of about 48%. The interface surface areas between the closest symmetry-related protein molecules are 1086, 953 and 1027 Å^2 for 1M, 1MI and 3MI, respectively. The crystal packing of M-PMV PR dimers is illustrated diagrammatically in Fig. 9.

4. Conclusions and outlook

The presented crystal structures of dimeric Mason–Pfizer monkey virus protease, both with and without an inhibitor, reveal overall structural similarity to the canonical fold of retroviral proteases. However, the characteristic and symmetric hairpin loops (Ile45–Ser64 in M-PMV PR) called flaps exhibit a large conformational variability in both the apo and holo structures of M-PMV PR. In retroviral proteases, the flap arms generally function to open access to the active site (when elevated or disordered) or seal a substrate/inhibitor molecule in the active-site cleft (when lowered), usually with the recruitment of an interface water molecule (WatI). The repertoire of M-PMV PR flap conformations in the known crystal structures is quite diverse. In the monomeric form (PDB entry 3sqf), the flap is compacted by curling itself on the exposed core of the protein. Part of this compaction is a short 3_{10}-helix. A similar 3_{10}-helix is also seen, albeit with a sequence shift and a different structural role, in the present 1M structure. These 3_{10}-helical intrusions are found near the tip of the flap and are different from the N-terminal template helix C1, which in M-PMV PR also adopts a 3_{10}-helical form. This 'N-terminal' helix is an important token confirming the hypothesis that the sequences of retropepsins may have evolved via gene duplication and divergence. The C2 α-helix in the 'C-terminal' part of the protein has a much clearer presence and is longer in the dimeric structure of M-PMV PR than in the monomeric form, most likely because of the conformational change of the following C-terminal oligopeptide. In the dimeric apo form 1M, one flap is lowered into the active site, although there is no inhibitor molecule to lock it in the 'down' position, while the other flap of the same dimer is disordered. Also intriguingly, the tips of all of the flap loops in the inhibitor complexes (1MI and 3MI) are disordered and are not present in the electron density, although their visible stems seem to trace the direction found in inhibitor complexes of other dimeric retropepsins. Consistent with the poor stability of the flaps in the inhibitor complexes of M-PMV PR, the characteristic interface water molecule WatI, which is invariant in retropepsin complexes with peptidomimetic inhibitors, is also only poorly defined in the electron density. The absence of a clear flap-lock sealed with WatI over the inhibitor molecule suggests that the complexes described in this work are not in the canonical closed form, but rather represent some kind of a 'breathing state' in which the space over the inhibitor is still accessible. The dimeric structures of M-PMV PR, however, contain an unusual water molecule deep in the active-site structure, coordinated in a symmetric tetrahedral fashion by the side chains of the catalytic residues (mutated D26N in these structures) and by the NH amide groups of the Gly28 residues from the catalytic triads. This water molecule corresponds neither to the tetrahedral interface molecule WatI (found between the flaps and substrate/inhibitor) nor to the nucleophilic water molecule WatC (poised between the substrate and active-site aspartates) but is nested much deeper in the active site in a manner that is unique to M-PMV PR.

The N- and C-termini of both subunits are engaged in the formation of the antiparallel β-sheet at the dimeric interface as in other retropepsins. However, in the present construct of the M-PMV PR protein (Trp1–Ala114), the C-terminal

peptides Asn109–Ala114 extend beyond the intersubunit β-structure and are disordered. This suggests that a shorter M-PMV PR construct, Trp1–Pro108 or Trp1–Ser107, corresponding to the 12 kDa truncated form of the protein, could be more suitable for future crystallographic studies.

An interesting issue is related to the presence of Cys106 in the native M-PMV PR sequence, which is posed close enough to its counterpart from the complementary subunit across the dimer interface for disulfide-bond formation. The use of a reducing agent (TCEP) in the crystallization buffer for 1M and 1MI precluded the oxidation of this cysteine residue, so the preventive C7A/C106A mutations in the 3MI variant were not necessary. On the other hand, it might be interesting to oxidize the crystals of 1M(I) to determine whether a covalently (S—S) linked dimer could be produced in a solid-state reaction within the crystal.

Although the inhibitor is unambiguously present in the 1MI structure (and most likely in the 3MI structure as well) in one orientation, its relatively poor electron density prohibits any deeper analysis of its conformation and enzyme interactions. The quality of the electron-density maps of the inhibitor cannot be explained by poor binding of the Pro-(O-Me)Tyr-Val-PSA-Ala-Met-Thr molecule in the binding site of M-PMV PR, as previous tests of phestatine-type inhibitors with this protease (Hrusková-Heidingsfeldová *et al.*, 1995) indicated that an extension or sequence variation of this structure resulted in a decrease of the original K_i (3 n*M*) by one order of magnitude. Rather, the quality of the electron density seems to be a consequence of the peculiar behavior of the flap loops of this protease. It might be necessary to provide a molecular mechanism, such as a disulfide S—S bond or ionic salt bridge, to lock the flaps firmly upon the inhibitor molecule in order to achieve better stability of these two important structural elements.

An obvious question arises from the presented results which underlies the general reason for working with M-PMV PR. What are the structural peculiarities of M-PMV PR that also make it stable in the monomeric form, in contrast to all other retroviral proteases? The overall conformation of dimeric M-PMV PR is similar to other retroviral proteases and does not provide a particular reason for this behavior, with one notable exception regarding the flap element, which is particularly conspicuous in the apo form 1M. In this inhibitor-free structure, flap B is folded in a way that is still more reminiscent of inhibitor complexes than of other dimeric retropepsins without inhibitor. The other flap of 1M (flap A) again differs significantly owing to its complete disorder. These observations suggest that during dimerization one flap has to unfold from the curled conformation and thus has to overcome an energetic barrier. Secondly, the flaps of dimeric M-PMV PR (especially in its inhibitor complexes 3MI and 1MI) never come together to form a dimer-stabilizing lock at their tips. Thirdly, there are reasons to believe that the dimer interface β-sheet has decreased stability compared with other retropepsins. For instance, the *B* factors in this area are high, suggesting that the β-sheet is not very stable. Moreover, a loose, unstructured C-terminus may additionally destabilize this secondary structure. Additionally, there are solvent molecules penetrating the dimerization interface in the region of the active site, suggesting a lower affinity of the protomers for each other. In conclusion, it may be the lowered stability of the M-PMV PR dimer that makes the monomeric form more abundant.

Acknowledgements

The diffraction data were collected on beamlines BL14.1 and BL14.2 of the BESSY II electron-storage ring operated by the Helmholtz-Zentrum Berlin (Mueller *et al.*, 2015). We would like to acknowledge the help and support of the local staff and of our colleagues Dr Milosz Ruszkowski and Dr Jakub Barciszewski.

Funding information

This work was supported by a grant (RVO 61388963) from the ASCR to IP.

References

Andreeva, N. S. (1991). *Adv. Exp. Med. Biol.* **306**, 559–572.
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.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Gilski, M., Kazmierczyk, M., Krzywda, S., Zábranská, H., Cooper, S., Popović, Z., Khatib, F., DiMaio, F., Thompson, J., Baker, D., Pichová, I. & Jaskolski, M. (2011). *Acta Cryst.* D**67**, 907–914.
Gustchina, A., Kervinen, J., Powell, D. J., Zdanov, A., Kay, J. & Wlodawer, A. (1996). *Protein Sci.* **5**, 1453–1465.
Hrusková-Heidingsfeldová, O., Andreansky, M., Fábry, M., Bláha, I., Strop, P. & Hunter, E. (1995). *J. Biol. Chem.* **270**, 15053–15058.
Jaskolski, M., Miller, M., Rao, J. K. M., Gustchina, A. & Wlodawer, A. (2015). *FEBS J.* **282**, 4059–4066.
Jaskolski, M., Miller, M., Rao, J. K. M., Leis, J. & Wlodawer, A. (1990). *Biochemistry*, **29**, 5889–5898.
Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B. H. & Wlodawer, A. (1991). *Biochemistry*, **30**, 1600–1609.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kervinen, J., Lubkowski, J., Zdanov, A., Bhatt, D., Dunn, B. M., Hui, K. Y., Powell, D. J., Kay, J., Wlodawer, A. & Gustchina, A. (1998). *Protein Sci.* **7**, 2314–2323.
Khatib, F., DiMaio, F., Foldit Contenders Group, Foldit Void Crushers Group, Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., Jaskolski, M. & Baker, D. (2011). *Nature Struct. Mol. Biol.* **18**, 1175–1177.
Kleffner, R., Flatten, J., Leaver-Fay, A., Baker, D., Siegel, J. B., Khatib, F. & Cooper, S. (2017). *Bioinformatics*, **33**, 2765–2767.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. & Murshudov, G. N. (2012). *Acta Cryst.* D**68**, 431–440.
Li, M., DiMaio, F., Zhou, D., Gustchina, A., Lubkowski, J., Dauter, Z., Baker, D. & Wlodawer, A. (2011). *Nature Struct. Mol. Biol.* **18**, 227–229.
Li, M., Gustchina, A., Materúz, K., Tözsér, J., Namwong, S., Goldfarb, N. E., Dunn, B. M. & Wlodawer, A. (2011). *FEBS J.* **278**, 4413–4424.
Li, M., Laco, G. S., Jaskolski, M., Rozycki, J., Alexandratos, J., Wlodawer, A. & Gustchina, A. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18332–18337.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). *Acta Cryst.* D**73**, 148–157.
Liu, Z., Huang, X., Hu, L., Pham, L., Poole, K. M., Tang, Y., Mahon, B. P., Tang, W., Li, K., Goldfarb, N. E., Dunn, B. M., McKenna, R. & Fanucci, G. E. (2016). *J. Biol. Chem.* **291**, 22741–22756.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
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. & Wlodawer, A. (1989). *Science*, **246**, 1149–1152.
Mueller, U., Förster, R., Hellmig, M., Huschmann, F., Kastner, A., Malecki, P., Pühringer, S., Röwer, M., Sparta, K., Steffien, M., Ühlein, M., Wilk, P. & Weiss, M. (2015). *Eur. Phys. J. Plus*, **130**, 141.
Mulichak, A. M., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Curry, K. A., Tomich, C.-S., Thaisrivongs, S., Sawyer, K., Watenpaughq, D., Sawyer, T. K. & Watenpaugh, K. D. (1993). *J. Biol. Chem.* **268**, 13103–13109.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Painter, J. & Merritt, E. A. (2006). *J. Appl. Cryst.* **39**, 109–111.
Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M. & Schiffer, C. A. (2003). *J. Virol.* **77**, 1306–1315.
Rose, R. B., Craik, C. S., Douglas, N. L. & Stroud, R. M. (1996). *Biochemistry*, **35**, 12933–12944.
Rose, R. B., Craik, C. S. & Stroud, R. M. (1998). *Biochemistry*, **37**, 2607–2621.
Salam, N. K., Adzhigirey, M., Sherman, W. & Pearlman, D. A. (2014). *Protein Eng. Des. Sel.* **27**, 365–374.
Satoh, T., Li, M., Nguyen, J. T., Kiso, Y., Gustchina, A. & Wlodawer, A. (2010). *J. Mol. Biol.* **401**, 626–641.
Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P. & Vriend, G. (2015). *Nucleic Acids Res.* **43**, D364–D368.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.
Wlodawer, A. & Gustchina, A. (2000). *Biochim. Biophys. Acta*, **1477**, 16–34.
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. Mol. Biol.* **2**, 480–488.
Wlodawer, A. & Jaskolski, M. (2016). *Encyclopedia of Cell Biology*, edited by R. A. Bradshaw & P. D. Stahl, Vol. 1, pp. 738–745. Waltham: Academic Press.
Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. (1989). *Science*, **245**, 616–621.
Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W. & Weber, I. T. (1998). *Biochemistry*, **37**, 4518–4526.
Zábranská, H., Tůma, R., Kluh, I., Svatoš, A., Ruml, T., Hrabal, R. & Pichová, I. (2007). *J. Mol. Biol.* **365**, 1493–1504.
Zábranský, A., Andreánsky, M., Hrušková-Heidingsfeldová, O., Havlíček, V., Hunter, E., Ruml, T. & Pichová, I. (1998). *Virology*, **245**, 250–256.