

## Crystal structure of XMRV protease differs from the structures of other retropepsins

Mi Li<sup>1,2</sup>, Frank DiMaio<sup>3</sup>, Dongwen Zhou<sup>1</sup>, Alla Gustchina<sup>1</sup>, Jacek Lubkowski<sup>4</sup>, Zbigniew Dauter<sup>5</sup>, David Baker<sup>3</sup> & Alexander Wlodawer<sup>1</sup>

**Using energy and density guided Rosetta refinement to improve molecular replacement, we determined the crystal structure of the protease encoded by xenotropic murine leukemia virus-related virus (XMRV). Despite overall similarity of XMRV protease to other retropepsins, the topology of its dimer interface more closely resembles those of the monomeric, pepsin-like enzymes. Thus, XMRV protease may represent a distinct branch of the aspartic protease family.**

XMRV is a newly discovered human retrovirus and the first gamma-retrovirus shown to be associated with human diseases. It has been detected in prostate cancer cells<sup>1</sup> as well as in individuals with chronic fatigue syndrome<sup>2</sup>. Although the identification of XMRV as the causal agent for these diseases is still controversial<sup>3</sup>, it seems prudent to identify targets for drugs against this potential pathogen. Because XMRV is a retrovirus, inhibition of the three enzymes encoded in its genome (reverse transcriptase, integrase and protease) provides the most direct path to inactivation of the virus. It has already been shown that the integrase inhibitor raltegravir is a potent inhibitor of XMRV<sup>4</sup>. Enzyme inhibition has been a very successful route for developing therapeutic agents against human immunodeficiency virus (HIV). In particular, numerous drugs targeting HIV-1 protease have been developed in the last 20 years<sup>5</sup>. The success of these efforts depended very much on the availability of the structure of HIV-1 protease, both as an apoenzyme and in complexes with inhibitors<sup>6</sup>. Although all retroviral proteases studied to date are structurally similar<sup>7</sup>, the fine differences in their structures allow for the development of specific inhibitors. For example, although HTLV-1 protease<sup>8</sup> is similar to HIV-1 protease<sup>9</sup>, it is very poorly inhibited by most HIV-1 protease inhibitors. None of the clinical inhibitors of HIV-1 protease have EC<sub>50</sub> values below 35 μM against XMRV in cell culture, which is three to four orders of magnitude higher compared to HIV-1 (ref. 4).

Although XMRV protease has not been previously isolated or expressed and characterized on a molecular level, a closely related enzyme from Moloney murine leukemia virus (MoMLV) has been

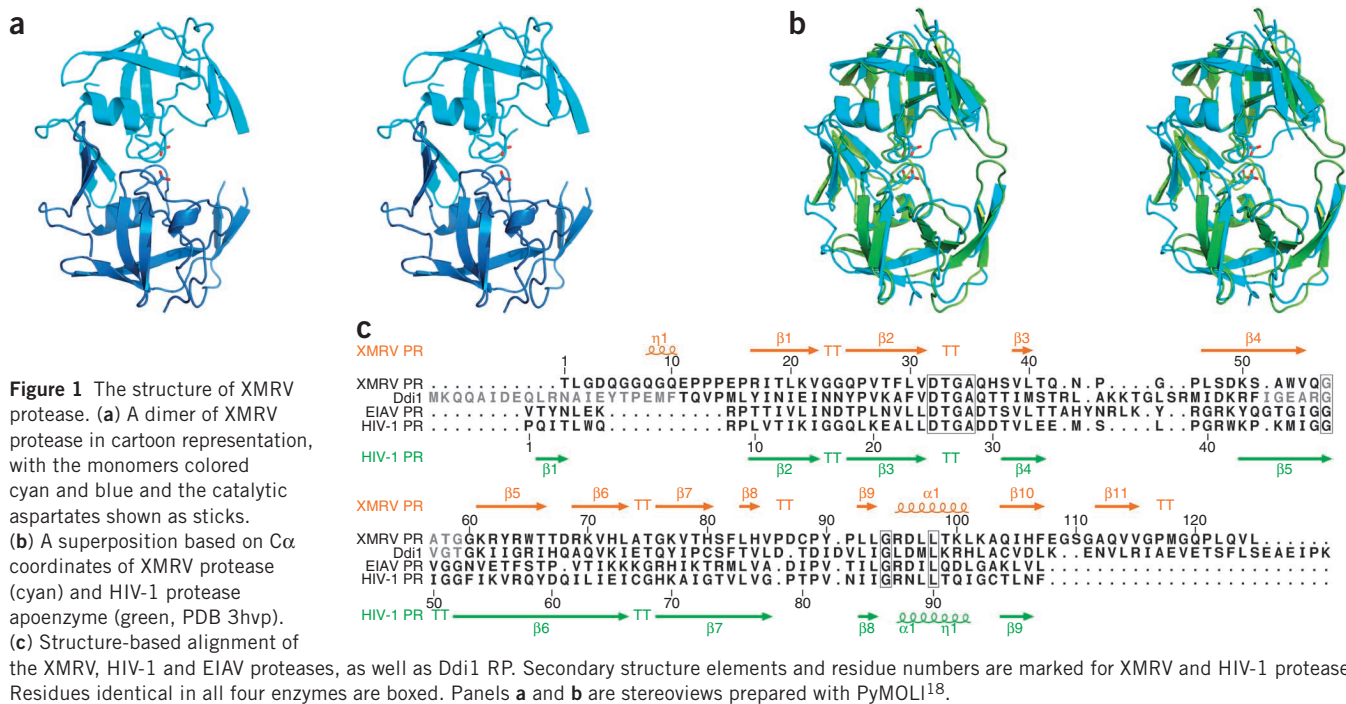
isolated and its amino acid sequence determined<sup>10</sup>. This information served as a guide in cloning XMRV protease (**Supplementary Methods**) and particularly in deciding the location of its probable termini. The expression construct contains 125 amino acids belonging to the enzyme, as well as a N-terminal hexahistidine tag preceded by a methionine. The enzyme migrates as a dimer on a gel-filtration column (data not shown). Its activity was demonstrated by extensive autolysis (**Supplementary Fig. 1a**) and by its cleavage of maltose-binding protein (MBP) in the MBP-XMRV fusion protein (**Supplementary Fig. 1b**). This autolysis was inhibited by TL-3, a broad-specificity retropepsin inhibitor (**Supplementary Methods** and **Supplementary Fig. 1**). This construct of XMRV protease was purified and crystallized, and diffraction data were collected to 1.97-Å resolution (**Supplementary Methods**).

Because XMRV protease contains only a single methionine, near its C terminus (Met118), phasing of diffraction data by using anomalous dispersion of selenomethionine seemed unlikely without the introduction of additional methionine residues. However, the structural similarity of all known retroviral proteases<sup>7</sup> suggested that molecular replacement should be sufficient for solving the structure of XMRV protease. We carried out extensive trials with models built on the basis of crystal structures of several retroviral proteases but found no refinable solutions (**Supplementary Methods**). We finally solved the structure of XMRV protease through a novel application of the Rosetta refinement<sup>11</sup> to several highest-scoring molecular replacement models. This application of the Rosetta refinement produced sufficient improvement of these structures to enhance the molecular replacement signal and resulted in a model that could be further refined by standard means (**Supplementary Methods** and **Table 1**).

A molecule of XMRV protease is a homodimer (**Fig. 1a**), with a two-fold symmetry axis that does not coincide with the symmetry elements of the crystal. Its fold generally resembles those of other retroviral proteases (**Fig. 1b**), although with several substantial differences, especially at the two termini. Both the N and C termini are longer in XMRV protease than in most other retropepsins. The N terminus contains a helical insertion before strand β1 (**Fig. 1c**). Instead of the interdigitated N and C termini (β1 and β9 strands, **Fig. 1c**) that create the dimer interfaces in all other structurally characterized retroviral proteases, the dimer interface of XMRV protease utilizes hairpins formed by strands β10 and β11, near the C termini of both monomers (**Figs. 1c** and **2a** and **Supplementary Figs. 2a** and **3**). The flaps of each protomer (residues 48–66) are partially disordered at their tips, a situation common for the apoenzymes of retropepsins<sup>12</sup>. However, the ordered parts of the flaps appear to represent the open conformation

<sup>1</sup>Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland, USA. <sup>2</sup>Basic Research Program, SAIC-Frederick, Frederick, Maryland, USA. <sup>3</sup>Department of Biochemistry, University of Washington, Seattle, Washington, USA. <sup>4</sup>Macromolecular Assembly Structure and Cell Signaling Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland, USA. <sup>5</sup>Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Argonne National Laboratory, Argonne, Illinois, USA. Correspondence should be addressed to A.W. (wlodawer@nih.gov).

Received 29 July; accepted 29 October 2010; published online 23 January 2011; doi:10.1038/nsmb.1964



seen in the apo form of HIV-1 protease<sup>13</sup>. The N-terminal fragment of XMRV protease is partially helical, with residues Gly6 through Glu111 disordered in monomer B, and is quite different from its counterparts in other retroviral enzymes (Supplementary Fig. 4).

Although the mode of dimerization of XMRV protease shows substantial differences from those of other retropepsins (Fig. 2b), it is much closer to that of the putative protease (RP) domain of the eukaryotic protein Ddi1 (ref. 14). The crystal structure of the isolated RP domain of *Saccharomyces cerevisiae* Ddi1 was solved and refined at 2.3-Å resolution (PDB code 2I1A; ref. 14), revealing similarity in the overall structural fold to retropepsins. However, to our knowledge, no enzymatic activity of Ddi1 RP has been reported. The overall structural similarity (Supplementary Fig. 5) of XMRV protease and Ddi1 RP is reflected by the r.m.s. deviations of 1.66 Å and 1.87 Å between the equivalent 85 C $\alpha$  atoms in the monomers and 174 C $\alpha$  atoms in the dimers of both proteins, respectively. By comparison, an analogous alignment of XMRV protease with the apo form of HIV-1 protease (PDB code 3HVP; ref. 13) yields r.m.s. deviations of 2.18 Å for the monomers and 2.35 Å for the dimers.

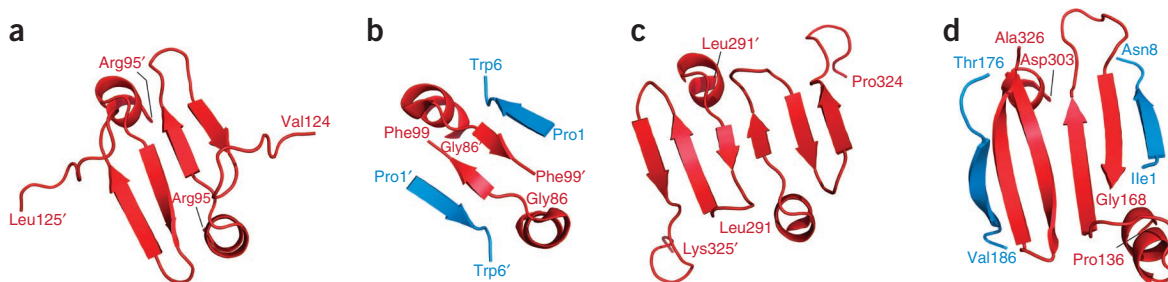
Like those of XMRV protease, the N and C termini of Ddi1 RP are substantially longer than in a majority of retropepsins. The dimer interface in Ddi1 RP is formed solely by the C-terminal part of the protomer (by three consecutive  $\beta$  strands,  $\beta$ 7– $\beta$ 9; Fig. 1c) and does not include the N terminus at all (Fig. 2c). A comparable situation is seen in XMRV protease, except that the interface uses only two  $\beta$  strands ( $\beta$ 10– $\beta$ 11). Residues Gly119 and Gln120 make a turn after  $\beta$ 11 and form hydrogen bonds with the O and N atoms of Gly116, thus extending the sheet, but the following segment of the C-terminal chain does not form any regular structure and points in a completely different direction (Fig. 2a and Supplementary Fig. 2a).

As noted in the description of the structure of Ddi1 RP<sup>14</sup>,  $\beta$  strands that form the dimer interface in that protein are rotated by  $\sim 45^\circ$  compared to their counterparts in HIV-1 protease and other retropepsins. Two of these strands in XMRV protease superimpose almost exactly on their counterparts in Ddi1 RP, retaining their angles, with only the residues at the turn between the interface strands following

a slightly different path in the two proteins, despite their identical length (Supplementary Fig. 6). The axis of the dimer interface  $\beta$ -sheet in XMRV protease is aligned roughly perpendicular to the long axis of the protease dimer. The direction of the interface strands and the lack of interdigitation resembles a situation seen in pepsin-like aspartic proteases, with the caveat that the dimerization interface in the latter enzymes is six-stranded (as in Ddi1 RP), in contrast to the four-stranded interface in XMRV protease (Fig. 2 and Supplementary Fig. 2a). This structure of the interface sheet results in a much smaller number of contacts with the opposite protomer in the dimer compared to other retropepsins, in which extensive intermolecular contacts are created by interdigitation of the C- and N-terminal  $\beta$  strands. Nonetheless, XMRV protease is dimeric in solution as well as in crystals.

The two  $\beta$ -strands that follow helix  $\alpha$ 1 in XMRV protease and Ddi1 RP and form the dimer interface are both topologically and structurally equivalent to the corresponding C-terminal loops of each domain of pepsin-like aspartic proteases (Fig. 2 and Supplementary Fig. 2a), whereas the third strand is missing in XMRV protease. In this respect, XMRV protease seems to be closer than the other retroviral proteases to the putative common ancestor of monomeric and dimeric aspartic proteases<sup>15</sup>, indicating divergence in their evolutionary paths.

A unique structural organization of N and C termini in XMRV protease leads to differences in the intersubunit interactions within the dimer interface compared to other retroviral enzymes. An important interaction stabilizing the dimers of retroviral proteases is created by an ion pair involving Arg8 of one protomer and Asp29' of the other one (HIV-1 protease numbering) (Supplementary Fig. 2b). In contrast to all other characterized retropepsins, in XMRV protease these two residues are not conserved. A residue equivalent to Arg8 is Glu15 (Fig. 1c), but its side chain faces an opposite direction because the following Pro16 adopts a *cis* conformation. Although Pro16 is conserved among retroviral proteases, the *trans* conformation of this residue in most of these enzymes leads to observed differences in topologies in the N-terminal strand. Gln36 in XMRV protease is equivalent to Asp29 in HIV-1 protease, and their respective side chains, in addition to differing in their ionic state, are also oriented differently. Although simian foamy



**Figure 2** Dimer interface regions in aspartic proteases. Strands belonging to the N-terminal regions of the molecules (or domains in pepsin) are blue, and the C-terminal regions are red. (a) XMRV protease. (b) HIV-1 protease. (c) Ddi1 RP. (d) Pepsin.

virus protease also lacks a corresponding ion pair, its structure has been characterized by NMR only for a monomer<sup>16</sup>, and thus its dimer interface cannot be analyzed. These intersubunit ionic interactions are substituted by hydrophobic contacts in XMRV protease (**Supplementary Fig. 2b**), thus modifying the network of interactions within the dimer interface. It must be pointed out, however, that mutation R8Q in HIV-1 protease, which replaces the ion pair with polar interactions, leads to only small differences in the activity of the enzyme<sup>17</sup>, indicating that the presence of an ion pair may not be necessary to stabilize the dimer.

As for other retropepsins crystallized in the absence of ligands, a water molecule bridges the two catalytic aspartates. The architecture of the active site in XMRV protease, particularly the hydrophobic lining of the binding site area, also resembles those of other retropepsins, suggesting that this enzyme might have similar substrate-recognition preferences. As an example, the loop Leu83–Leu92, equivalent to the so-called polyproline loop in HIV-1 protease (residues Leu76–Ile84), adopts a conformation in XMRV protease that is very similar to that in other retroviral enzymes, but contrasts with the one found in Ddi1 RP (**Supplementary Fig. 7**). As revealed by numerous structures of inhibitor complexes of retropepsins, residues of this loop are involved in extensive interactions with the ligands. Therefore, although the only structure of XMRV protease currently available is that of the apoenzyme form, overall conservation of the structural features of retropepsins in the active site area allows prediction of the putative subsites for the residues of substrates and/or peptidic inhibitors. The residues predicted to form subsites S1–S4 in the monomer of XMRV protease are compared with their equivalents in HIV-1 and EIAV proteases in **Supplementary Table 2**. Although the predominantly hydrophobic character of the binding sites is well preserved as a result of the conservative nature of a majority of substitutions, the presence in XMRV protease of unique polar residues such as His37 in S2 and S4, Tyr90 in S1 and S3, and Gln36 and Gln55 (presumably, since the fragment of the flap with this residue is disordered) in S3 and S5 provides clues for the design of specific inhibitors against XMRV protease. The other important difference observed in pocket S3 is due to the lack of conservation in XMRV protease of the previously mentioned Arg8 and Asp29 that form part of this pocket in the other retroviral enzymes. Further studies with substrates and inhibitors of XMRV protease will be necessary to define the specificity of this enzyme and to design more effective inhibitors.

**Accession codes.** Protein Data Bank: Coordinates and structure factors have been deposited with the accession code 3NR6.

*Note: An enhanced version of this article and supplementary information are available on the Nature Structural & Molecular Biology website.*

#### ACKNOWLEDGMENTS

We acknowledge the use of beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT), located at the Advanced Photon Source (APS), 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. This work was supported in part by the Intramural Research Program of the US National Institutes of Health, National Cancer Institute, Center for Cancer Research and with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. 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.

#### AUTHOR CONTRIBUTIONS

M.L. produced the protein and grew crystals; M.L. and Z.D. collected and processed crystallographic data; M.L., F.D.M., D.Z., A.G., J.L., and Z.D. performed calculations, structure refinement and analysis; D.B. and A.W. supervised the project. All authors discussed the results and participated in writing the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Schlager, R., Choe, D.J., Brown, K.R., Thaker, H.M. & Singh, I.R. *Proc. Natl. Acad. Sci. USA* **106**, 16351–16356 (2009).
- Lombardi, V.C. *et al. Science* **326**, 585–589 (2009).
- Groom, H.C. *et al. Retrovirology* **7**, 10 (2010).
- Singh, I.R., Gorzynski, J.E., Drobysheva, D., Bassit, L. & Schinazi, R.F. *PLoS ONE* **5**, e9948 (2010).
- Wlodawer, A. & Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).
- Wlodawer, A. & Erickson, J.W. *Annu. Rev. Biochem.* **62**, 543–585 (1993).
- Dunn, B.M., Goodenow, M.M., Gustchina, A. & Wlodawer, A. *Genome Biol.* **3**, S3006.1–S3006.7 (2002).
- Gustchina, A., Jaskolski, M. & Wlodawer, A. *Cell Cycle* **5**, 463–464 (2006).
- Li, M. *et al. Proc. Natl. Acad. Sci. USA* **102**, 18322–18337 (2005).
- Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. *Proc. Natl. Acad. Sci. USA* **82**, 1618–1622 (1985).
- DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W. & Baker, D. *J. Mol. Biol.* **392**, 181–190 (2009).
- Miller, M., Jaskólski, M., Rao, J.K.M., Leis, J. & Wlodawer, A. *Nature* **337**, 576–579 (1989).
- Wlodawer, A. *et al. Science* **245**, 616–621 (1989).
- Sirkis, R., Gerst, J.E. & Fass, D. *J. Mol. Biol.* **364**, 376–387 (2006).
- Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A. & Blundell, T.L. *Nature* **271**, 618–621 (1978).
- Hartl, M.J., Wohrl, B.M., Rosch, P. & Schweimer, K. *J. Mol. Biol.* **381**, 141–149 (2008).
- Louis, J.M., Clore, G.M. & Gronenborn, A.M. *Nat. Struct. Biol.* **6**, 868–875 (1999).
- DeLano, W.L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, USA, 2002).