

# The endopeptidase of the maize-affecting *Marafivirus* type member maize rayado fino virus doubles as a deubiquitinase

Received for publication, May 11, 2021, and in revised form, July 5, 2021 Published, Papers in Press, July 12, 2021,  
<https://doi.org/10.1016/j.jbc.2021.100957>

Ankoo Patel<sup>1</sup>, Jessica A. M. McBride, and Brian L. Mark\*

From the Department of Microbiology, University of Manitoba, Winnipeg, Canada

Edited by Craig Cameron

*Marafiviruses* are capable of persistent infection in a range of plants that have importance to the agriculture and biofuel industries. Although the genomes of a few of these viruses have been studied in-depth, the composition and processing of the polyproteins produced from their main ORFs have not. The *Marafivirus* polyprotein consists of essential proteins that form the viral replicase, as well as structural proteins for virus assembly. It has been proposed that *Marafiviruses* code for cysteine proteases within their polyproteins, which act as endopeptidases to autocatalytically cleave the polyprotein into functional domains. Furthermore, it has also been suggested that *Marafivirus* endopeptidases may have deubiquitinating activity, which has been shown to enhance viral replication by downregulating viral protein degradation by the ubiquitin (Ub) proteasomal pathway as well as tampering with cell signaling associated with innate antiviral responses in other positive-sense ssRNA viruses. Here, we provide the first evidence of cysteine proteases from six different *Marafiviruses* that harbor deubiquitinating activity and reveal intragenus differences toward Ub linkage types. We also examine the structural basis of the endopeptidase/deubiquitinase from the *Marafivirus* type member, maize rayado fino virus. Structures of the enzyme alone and bound to Ub reveal marked structural rearrangements that occur upon binding of Ub and provide insights into substrate specificity and differences that set it apart from other viral cysteine proteases.

Positive-sense ssRNA (+ssRNA) viruses have evolved remarkable polycistronic translational mechanisms that maximize genomic coding capacity to produce the viral proteins needed for replication and packaging (1, 2). These include programmed ribosomal frameshifting, polyprotein expression and processing, and the presence of subgenomic mRNAs (2, 3). Unlike the highly divergent DNA viruses that typically use alternative splicing of viral mRNA to produce their full arsenal of proteins (4, 5), +ssRNA viruses most commonly have their nonstructural genes translated directly from their genomes *via* large ORFs (3, 5). Upon translation, these ORFs give rise to precursor polyproteins that can be processed into functional units by endopeptidases found within the polyprotein or by cellular proteases that achieve the

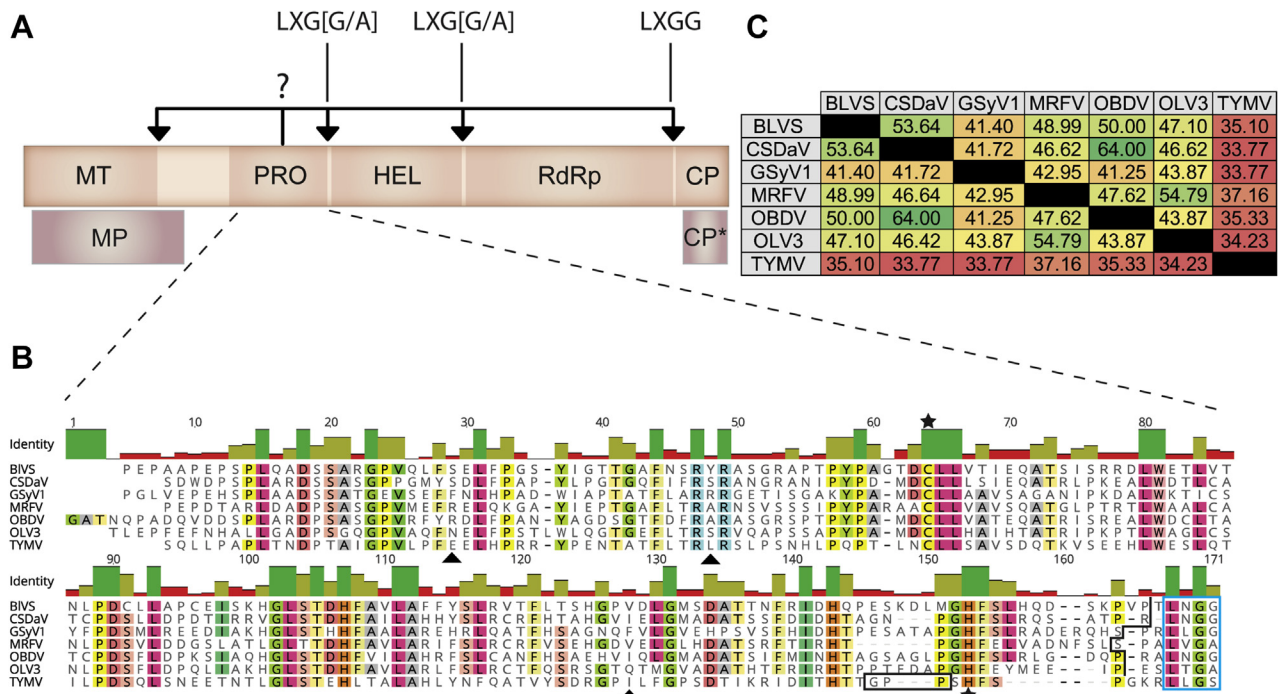
same purpose (5). This expression mechanism allows viruses to produce a full set of proteins from their genomes without needing to encode for additional genetic features that would typically direct and regulate translation. +ssRNA viruses infect virtually all forms of life (6). Significant attention has been paid to human- and animal-affecting +ssRNA viruses that are capable of cross-species transmission; however, the +ssRNA virome affecting plants appears to be far more diverse and abundant (6, 7). With the threat of food shortage becoming more of a reality in a matter of decades because of climate change and the quickly increasing global population (8, 9), understanding the relationship between crops important to global food security and their viral pathogens will aid in developing sustainable agricultural practices.

*Marafiviruses* are +ssRNA viruses, which cause persistent infection in diverse plant species that are agriculturally relevant to food and biofuels production (10–12). Currently, there are ten classified members of the genus *Marafivirus* (13), and additional *Marafivirus* candidates that affect different plant species have been identified recently (14, 15). *Marafiviruses* are persistent propagative viruses, meaning that they are also able to replicate in the leafhopper insect vectors that transmit them to their final plant host (16). *Marafiviruses* have a single-stranded genome with an average size of 6 to 7 kb, which is translated into a single polyprotein that is roughly 2000 amino acids in length (Fig. 1A). The polyprotein characteristically contains four nonstructural domains that comprise the viral replicase (methyltransferase, protease (PRO), helicase (HEL), and RNA-dependent RNA polymerase (RdRp)) as well as the major structural coat protein (CP) for genome packing (17). Maize rayado fino virus (MRFV) is the type member of the genus *Marafivirus* within the family Tymoviridae (10). The virus was first described in 1969 from Costa Rica and has since been found as far south as Brazil and as far north as the United States (10, 18, 19). MRFV infection causes reduction in the height and ear development with few to no seeds within certain corn species and can lead to severe agricultural loss with decimation to nearly 100% in some cultivars (20).

MRFV shares remarkable sequence similarity to the *Tymovirus* type member, turnip yellow mosaic virus (TYMV) (~43% at the amino acid level); however, the amount of foundational research that exists between the two is disparate, with TYMV being far better studied. The papain-like cysteine protease of TYMV (TYMV PRO) is currently the most extensively

\* For correspondence: Brian L. Mark, [brian.mark@umanitoba.ca](mailto:brian.mark@umanitoba.ca).

## Marafivirus endopeptidases double as deubiquitinases



**Figure 1. Marafivirus polyprotein arrangement and endopeptidase comparison.** **A**, simplified schematic of the typical arrangement of the sole polyprotein produced by *Marafiviruses* and two additional proteins produced from genomic RNA. Arrows indicate junction points between each domain and the possible cleavage sites targeted by the endopeptidase. The four amino acid residues at each putative cleavage sites are indicated where "X" is any amino acid. **B**, sequence alignment of PRO domains from six *Marafiviruses* and one *Tymovirus*. Highlighted residues indicate agreement of  $\geq 65\%$ . Stars indicate the catalytic cysteine and histidine of the active sites, and the triangle indicates a potential triad residue. Triangles indicate key residues. The box around the C terminus GPP of TYMV is the mobile loop characteristic to *Tymoviruses*. The stepwise line at the C terminus of the *Marafivirus* PRO domains indicate where the expression constructs used herein were terminated. Alignment was done in Geneious v.11.1.5. **C**, percent identity matrix of all six *Marafivirus* PRO domains and the PRO domain of TYMV. Values are presented in the heat map format, where green indicates a higher degree of sequence similarity. Multiple sequence alignment matrix values were determined using Clustal Omega (70). BIVS, blackberry virus S; CP, major coat protein; CP\*, minor coat protein; CSDaV, citrus sudden death-associated virus; GSYV1, Grapevine Syrah virus 1; HEL, helicase; MP, movement protein; MRFV, maize rayado fino virus; MT, methyl transferase; OBDV, oat blue dwarf virus; OLV3, Olive latent virus 3; PRO, protease; RdRp, RNA-dependent RNA polymerase; TYMV, turnip yellow mosaic virus.

characterized plant deubiquitinase (DUB) (21–25). Over 67% of plant viral proteases, like all the *Marafivirus* PROs, are cysteine proteases (26). The main function of PRO is to act as a polyprotein processor, but it has been shown to have auxiliary function as a ubiquitin (Ub) hydrolase to aid in bypassing the host innate immune system by removing Ub from the RdRp, preventing its degradation by the Ub proteasome (21, 24, 25, 27). Notably, the junctions between the replicase domains within the *Marafivirus* polyprotein (Fig. 1A, arrows) closely resemble the C-terminal <sup>73</sup>LRGG<sup>76</sup> tail of Ub (the cleavage site for Ub hydrolases) (28), suggesting *Marafivirus* endopeptidases may also have deubiquitinating activity.

Protein ubiquitination is a highly conserved post-translational modification process that occurs in eukaryotes including plants, which regulates the function, trafficking, and fate of protein substrates in the cell (28, 29). Ubiquitination involves the tethering of a Ub molecule(s) to a protein substrate by the combined efforts of Ub-activating enzymes, Ub-conjugating enzymes, and Ub ligases (28, 30). Ub is typically conjugated to the substrate protein as a polyubiquitin chain (28). The first Ub molecule is covalently attached to a lysine on the surface of a substrate protein via the carboxy-terminal glycine of Ub, forming a covalent

isopeptide bond with the lysine side chain  $\epsilon$ -amino group (28, 30). Additional Ub molecules can be attached to the first conjugated Ub through linkages formed between one of seven lysine residues on the surface of Ub (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, or Lys63) and the terminal Gly of the newly added Ub molecule (29). Met1-linked chains occur in mammalian systems, although they have not been identified in plant cells (31). Typically, Ub chains are of the same linkage type, but mixed poly-Ub chains are known to exist (28). The nature of the conjugation type and/or length ultimately determines the function or fate of the protein substrate (30, 32).

In the model plant organism, *Arabidopsis thaliana*, 6% of all protein-encoding genes are purportedly linked to some facet of Ub modification (33), and 12 genes have been identified to date that code for functional Ub or Ub-like proteins such as RUB proteins (31), illustrating the large role that the Ub system has in plants. Lys48 and Lys63 are the most well-understood poly-Ub chain types and the two most abundant forms in plants (34). Lys48 polyubiquitination (and Lys63 to a lesser extent) marks a substrate protein for degradation in plants by the Ub-proteasome system (30, 35, 36), whereas Lys63 has many different roles in plant cells such as DNA replication/repair,

iron homeostasis, endocytosis, nutrient transport, vacuolar sorting, protein synthesis, and immunity (31, 37).

Here, we provide the first structural and functional insight into the endopeptidases from the genus *Marafivirus*. We present data illustrating that the PRO domains from six different species within this genus all have *bona fide* deubiquitinating activity. We compare the differences and similarities of all six PROs toward recognition of different types of Ub substrates and contrast our findings with those of the *Tymoviruses*. We also discuss the endopeptidase activity of MRFV PRO that is required for polyprotein processing to generate the viral replicase complex. X-ray structures of MRFV PRO alone and bound to Ub reveal unique structural characteristics that differ from TYMV PRO and notable conformational changes that occur in response to Ub substrate. Together, our data provide new insight into how *Marafiviruses* generate their replicase machinery through polyprotein processing and that they exhibit DUB activity that may corrupt the immune response of their hosts.

## Results and discussion

### Comparative sequence analysis of the PRO domains of six Marafiviruses and one Tymovirus

As shown in Figure 1A, the majority of the *Marafivirus* proteome exists within one major polyprotein (~200 kDa) consisting of the core replication proteins needed for the virus to replicate its genome within the host (10, 38). Structural proteins are also present within the polyprotein as well as the sole endopeptidase PRO (38). The methyltransferase, PRO, HEL, RdRp, and CP domains are consistently found in this arrangement within the polyprotein. The PRO domains are believed to be involved with the processing of the polyproteins into individual functional subunits by cleavage of the putative scissile bond(s) as observed for *Tymoviruses* such as TYMV (22, 39–41).

An amino acid sequence alignment of the six *Marafivirus* PRO domains we examined reveals considerable variability in the percent identity between the *Marafivirus* enzymes and consistently low identity to TYMV PRO although they appear to serve the same purpose (Fig. 1, B and C). Regardless, the similarity between the *Marafivirus* and TYMV PRO sequences was sufficient to identify *Marafivirus* PRO domains based on alignments against the TYMV PRO sequence for which an X-ray crystal structure has been determined (Fig. 1B). The PRO domains from six *Marafiviruses* were explored: BIVS, CSDaV, GSyV1, MRFV, OBDV, and OLV3. The DNA sequence for each PRO was codon-optimized for expression in *E. coli*. Amino acid sequences for the *Marafiviruses* had been derived from RNA sequences deposited in the NCBI (Table S1). Figure 1B shows the amino acid sequences of each PRO domain that was expressed. As determined here, functionally active *Marafivirus* PROs are compact with an average sequence length and molecular weight of ~160 residues and ~17 kDa, respectively. The *Marafivirus* PRO enzymes we studied share at least 40% sequence identity, with CSDaV and OBDV having a remarkable 64% sequence identity (Fig. 1C).

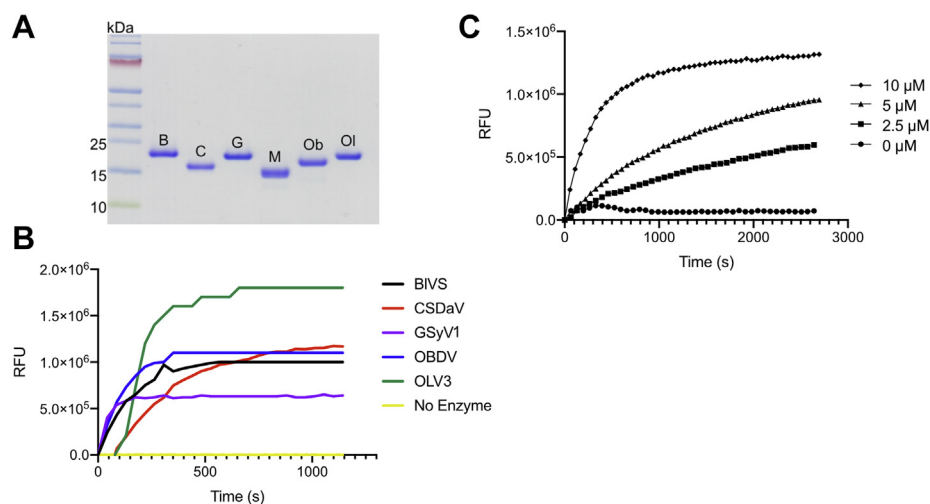
Comparing the sole *Tymovirus* PRO (TYMV) with the sequences of the PROs from the *Marafiviruses* consistently reveals, as expected, the highest degree of disagreement. Interestingly, GSyV1 PRO has the least sequence similarity to any of the PRO domains of all *Marafivirus* endopeptidases analyzed and also shares the least similarity with TYMV (along with CSDaV) at 33.77%. These variations in sequence similarity can potentially be attributed to many factors from wide differences in hosts, host climate/ecosystem, and viral vectors.

### Marafivirus endopeptidases have auxiliary deubiquitinating activity

Importantly, the process of ubiquitination is reversible, allowing Ub molecules conjugated to various substrate proteins to be uncoupled after the cellular function(s) dictated by ubiquitination is complete (29). This reversibility of the Ub system is carried out by cellular DUBs, of which, there has been approximately 50 identified in *A. thaliana* alone (29, 31, 37, 42). Cellular DUBs are important in reversing Ub linkages to target proteins but also are essential in processing Ub precursor proteins (31, 42). Not surprisingly, viruses have acquired the ability to exploit the Ub system to their advantage by encoding for multifunctional proteolytic enzymes (often cysteine proteases) that not only assist with viral replication by processing the viral polyprotein but also acting as DUBs to shut down Ub-dependent host antiviral mechanisms (21, 43–45).

When it was discovered that TYMV PRO had *bona fide* DUB activity, it was suggested that additional plant viruses may possess this function as well, including *Marafiviruses* (11, 19, 21). To assay for this potential DUB activity in *Marafiviruses*, we chose to study the enzymatic activity of six PRO enzymes from a range of *Marafivirus* species (Fig. 1B). Each PRO domain examined was based on the region known to exhibit endopeptidase and DUB activity from TMYV PRO and excluded the putative cleavage site at the putative PRO|HEL junction of each viral polyprotein as shown in Figure 1B. Each construct contained the conserved Cys and His residues (Fig. 1B, denoted with stars) that form the papain-like cysteine PRO active site, forming the catalytic dyad in which a cysteine nucleophile and histidine base work in concert to hydrolyze the scissile bond (29, 45). The recombinant PRO domains were stable *in vitro*, and each could be purified to homogeneity (Fig. 2A). The fluorogenic substrate Ub-AMC was used to assess the DUB activity for each enzyme (46). The fluorogenic substrate on its own has a relatively low signal; however, hydrolysis by DUB enzymes liberates the AMC fluorophore from Ub, which dequenches the fluorophore and results in a measurable signal (46). As seen in Figure 2A, the PRO domains of BIVS, CSDaV, GSyV1, OBDV, and OLV3 (black, red, purple, blue, and green, respectively) were incubated with Ub-AMC at a constant concentration and fluorescence measurements were taken over a 20-min period at an ambient temperature. A clear increase in fluorescence was observed for all five of the viral enzymes over time compared with a control lacking enzyme, which showed no appreciable Ub-AMC hydrolysis. These data

## Marafivirus endopeptidases double as deubiquitinases



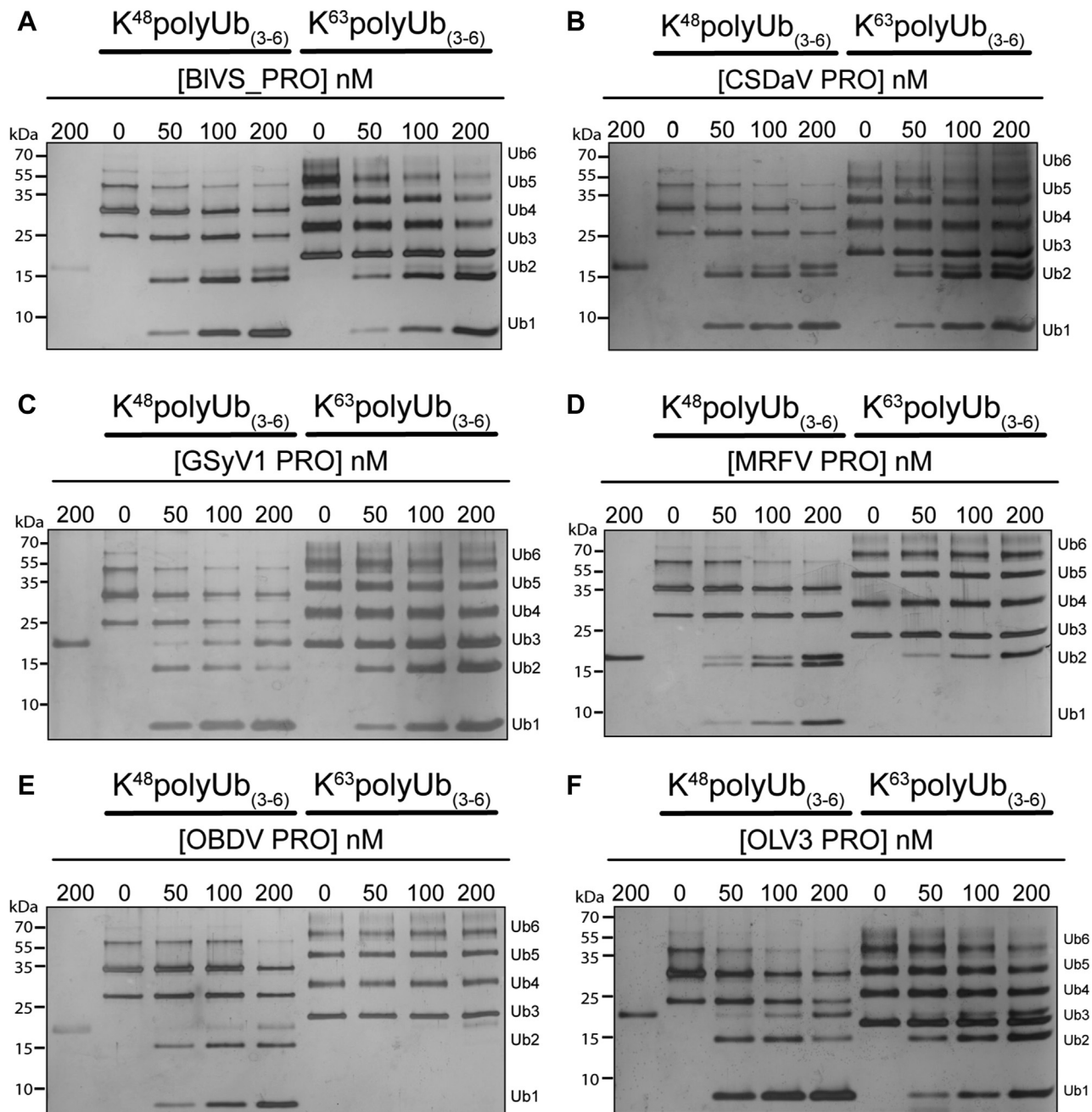
**Figure 2. Ubiquitin hydrolase activity of *Marafivirus* proteases.** A, 10% TRIS Tricine SDS-PAGE gel of purified PRO domains of six *Marafiviruses* (blackberry virus S (B), citrus sudden death-associated virus (C), Grapevine Syrah virus 1 (G), maize rayado fino virus (M), oat blue dwarf virus (Ob), and olive latent virus 3 (Ol)). B, fluorescence versus time progress curve of *Marafivirus* proteases with Ub-AMC. Substrate was used at a final concentration of 200 nM and enzyme at 5  $\mu$ M. Curves are colored with respect to enzyme (black–blackberry virus S (BIVS); red–citrus sudden death-associated virus (CSDaV); purple–Grapevine Syrah virus 1 (GSyV1); blue–oat blue dwarf virus (OBDV); green–Olive latent virus 3 (OLV3); yellow–substrate only). C, fluorescence versus time progress curve of MRFV PRO with Ub-AMC. The substrate was used at a final concentration of 200 nM, and the enzyme concentration was varied as indicated. Kinetic plots were designed in GraphPad Prism v.8.4.3. AMC, 7-amino-4-methylcoumarin; MRFV, maize rayado fino virus; PROs, proteases; Ub, ubiquitin.

confirm that *Marafivirus* PRO enzymes do exhibit DUB activity, further expanding the number of known viruses known to encode this activity. As all of the structural data we present here were for the *Marafivirus* type member PRO domain (MRFV PRO), a more thorough progress curve was generated to show enzyme concentration dependence (Fig. 2B). It can be seen that there is a clear concentration-dependent rate of increase in the fluorescent signal directly proportional to the relative amount of enzyme present per reaction. These findings confirm that *Marafivirus* PRO domains are also DUB enzymes.

As mentioned above, the cellular outcome of ubiquitination is often dictated by the topology of the Ub scaffold tethered to the target protein in the form of poly-Ub chains (29). To evaluate the substrate recognition ability of these PROs toward poly-Ub chains, each enzyme was coincubated with either  $K^{48}/K^{63}$  poly-Ub chains (3–6 Ub molecules in length) to determine their substrate specificity. Upon mixing enzyme with each Ub chain type, the presence of di-/mono-Ub should accumulate, and higher molecular weight species diminish if Ub chain hydrolysis is occurring. As seen in Figure 3, A–C and F, BIVS, CSDaV, GSyV1, and OLV3 PRO all appear to hydrolyze both  $K^{48}/K^{63}$  poly-Ub chains. A clear accumulation of di-/mono-Ub is present for both substrate types and increases with higher enzyme concentration. Interestingly, MRFV and OBDV PRO (Fig. 3, D and E) only appear to act on  $K^{48}$  poly-Ub chains. When comparisons are drawn between the PROs at the sequence level, MRFV PRO is more similar in sequence to BIVS and OLV3 PRO (~49 and 55%, respectively—Fig. 1C) than in comparison with OBDV PRO (~48%), whereas OBDV PRO is more similar to BIVS PRO (50%) and CSDaV PRO (64%); however, the preferences of MRFV and OBDV PRO toward Ub-substrate are shared. The preference of MRFV and OBDV PRO for

$K^{48}$  poly-Ub chains illustrates a surprising difference between these *Marafivirus* endopeptidases. It is possible that BIVS, CSDaV, GSyV1, and OLV3 PRO share a common structural feature that allows for broader substrate specificity and is absent from MRFV and OBDV PRO. Unfortunately, there are too many differences between the enzymes at the primary sequence level to identify a region or motif that could be responsible for the difference in substrate specificity, and three-dimensional structural analysis would be needed to gain further insight.

By way of comparison, earlier studies illustrated that TYMV PRO deubiquitinates the RdRp of the TYMV replicase, rescuing the polymerase from proteasomal degradation via K48-polyubiquitination (21, 27). The DUB activity of TYMV PRO thus decreases RdRp turnover and appears to contribute to viral infectivity in plant cells (21, 27). Interestingly, it was also previously found that TYMV PRO is also able to hydrolyze K63 poly-Ub chains (21), but its intended target substrate in the cell remains unknown. The role of K63 uncoupling by viral plant DUBs is not well understood compared with mammalian system-affecting viral DUBs, which degrade K63 linkages to suppress innate immune signaling cascades (43, 45, 47, 48). We found that CSDaV, BIVS, GSyV1, and OLV3 also exhibit activity against K48 and K63-poly-Ub (Fig. 3), so it is possible that these enzymes uncouple Ub from a variety of cellular targets to promote viral replication, not just the viral RdRp. This is true for mammalian +RNA viruses that encode DUB enzymes (45), and additional studies may reveal this to also be the case for plant +RNA viruses. In contrast, we found that MRFV and OBDV PRO show a finer degree of substrate recognition than the other viral PRO enzymes, including TYMV PRO, showing activity only against K48 linkages. This suggests preference toward preventing RdRp



**Figure 3. Qualitative analysis of poly-Ub chain hydrolysis by Marafivirus proteases.** A–F, K<sup>48</sup>/K<sup>63</sup> poly-Ub chains in the presence or absence of variable concentrations of the indicated Marafivirus proteases/DUBs. Lane 1 for each assay contains protease alone as a reference. Ub chain lengths after hydrolysis are indicated. DUBs, deubiquitinases; Ub, ubiquitin.

degradation by the 26S proteasome or other aspects of K48 polyubiquitination in plants, such as plant development, hormone signaling, and cell cycle mediation (49).

#### MRFV PRO facilitates polyprotein processing

Autocatalytic viral polyprotein processing by a PRO domain(s) encoded within the polyprotein represents a remarkably efficient mechanism of protein expression by +RNA viruses (22, 50–53). Previous findings confirmed that TYMV’s polyprotein has at least two cleavage sites, which is carried out by the PRO domain (22, 40, 54). They are located between the

PRO|HEL and HEL|RdRp junctions. We used MRFV PRO to gain insight into marafaviral polyprotein cleavage events and all subsequent experiments because this enzyme also proved amenable to structural studies and has considerable agricultural relevance. Furthermore, being the type member of the genus, we believe that it is most representative. To begin exploring the polyprotein processing by the MRFV PRO domain, two *E. coli* protein expression constructs were designed for the recombinant production of a subsection of the MRFV polyprotein spanning the PRO–HEL region. Two versions of the region were generated, one that contained a

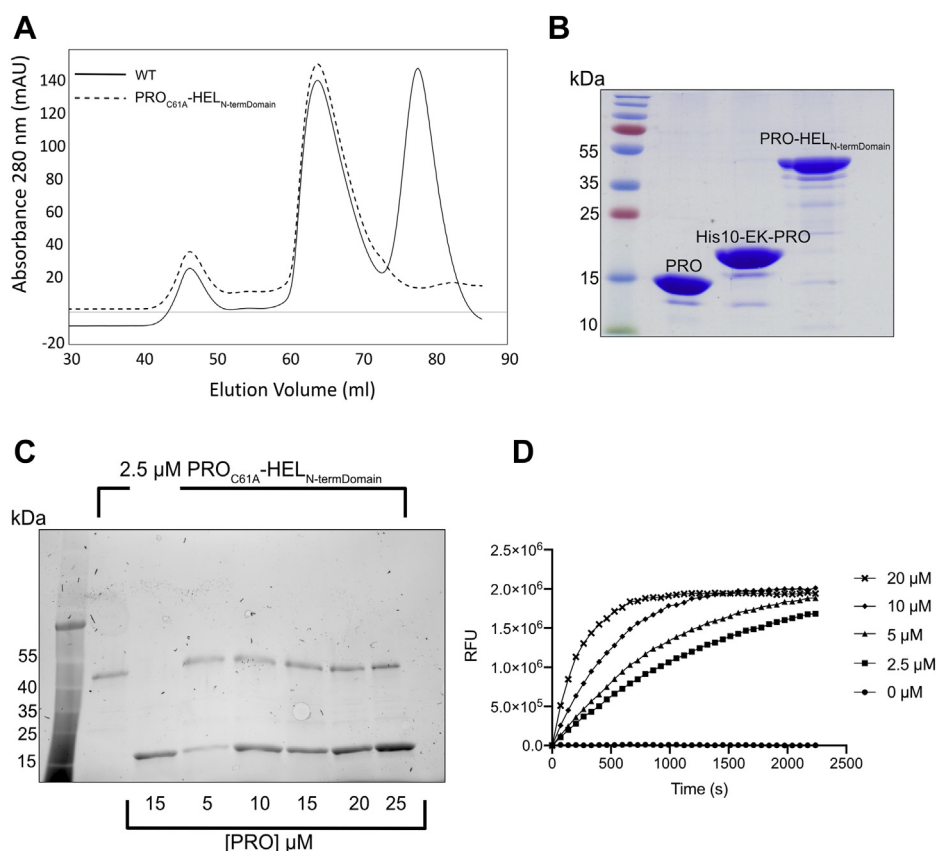
## Marafivirus endopeptidases double as deubiquitinases

catalytically active PRO domain (WT) and another where the active site cysteine of PRO had been mutated (C61A). For both proteins, the PRO domain was expressed in its entirety, whereas only the N-terminal domain (ATP-binding domain) of the HEL was included. Importantly, however, the region contained the putative LVGA recognition site at the PRO|HEL junction (Fig. 1, A and B; cyan box). We predicted the site would be cleaved by the PRO domain although it contained a GA motif at the C-terminus, which is atypical for a DUB, which usually cleaves after a diglycine motif (GG). The complete PRO-HEL fusion would have been ~60 kDa, while the truncated form is ~44 kDa, which proved amenable for expression in *E. coli*. Cleavage at the predicted site by PRO would result in a ~19-kDa N-terminal, His<sub>10</sub>-tagged PRO (with an enterokinase site in between the affinity tag and PRO), as well as an untagged version of the ATP-binding domain of the MRFV HEL (~25 kDa) that would not be captured through affinity purification if cleavage by PRO were to occur.

Figure 4A depicts the gel filtration chromatograms of the WT and C61A PRO-HEL<sub>N-TermDomain</sub> proteins. As shown for the active site mutant (dashed line), only one significant species is present at an elution volume of ~65 ml, whereas in the WT trace (solid line), two species are present, suggesting

autocatalytic cleavage of the PRO-HEL<sub>N-TermDomain</sub> by PRO. For the WT protein, the larger species shares a nearly identical elution volume as the species containing the PRO active site mutant and likely represented the intact PRO-HEL<sub>N-TermDomain</sub>. Furthermore, for the WT protein, a second, lower molecular weight species appears at an elution volume of ~80 ml. As stated above, cleavage at the proposed LVGA PRO|HEL junction would generate two additional proteins, only one of which (PRO) would be retained on a nickel affinity column as it is the only one that retains a His<sub>10</sub> affinity tag. Figure 4B shows an SDS-PAGE gel of the species obtained from the WT gel filtration experiment. The intact full-length protein would have a theoretical molecular weight of ~44 kDa, whereas the proteins generated by PRO-mediated cleavage would be ~19 kDa (PRO domain) and ~25 kDa (HEL ATPase domain). The processed HEL portion of the WT protein was lost in the purification process. Figure 4, A and B demonstrate that the WT PRO-HEL<sub>N-TermDomain</sub> is processed at the predicted junction, however, not to completion.

To assess the nature of how PRO is acting to process the PRO|HEL junction, be it in *cis* and/or *trans*, a cleavage assay was carried out using the C61A mutant. As seen in Figure 4C, a constant concentration of the C61A mutant of PRO-



**Figure 4. MRFV PRO cleaves at the PRO-HEL junction in *cis* and recognizes a cleavage site reminiscent of the C terminus of ubiquitin.** A, gel filtration chromatograms of the purification of MRFV PRO<sub>C61A</sub>-HEL<sub>N-TermDomain</sub> and WT. B, 12% Coomassie-stained SDS PAGE gel of MRFV PRO from pGEX-6P-1 (~16.4 kDa), in *cis*-cleaved MRFV PRO from WT MRFV PRO-HEL<sub>N-TermDomain</sub> (~19.4 kDa) and uncleaved WT MRFV PRO-HEL<sub>N-TermDomain</sub> (~44.1 kDa). C, 2.5 μM of purified MRFV PRO<sub>C61A</sub>-HEL<sub>N-TermDomain</sub> was incubated with increasing concentrations of MRFV PRO for 30 min at 25 °C. Each reaction was subsequently loaded onto a 12% stain-free SDS-PAGE gel (Bio-Rad). D, fluorescence versus time progress curve of MRFV PRO with LRGG-AMC. The substrate was used at a concentration of 25 μM, and the enzyme concentration was varied as indicated. Kinetic plot were designed in GraphPad Prism v.8.4.3. AMC, 7-amino-4-methylcoumarin; HEL, helicase; MRFV, maize rayado fino virus; PROs, proteases.

HEL<sub>N-TermDomain</sub> was incubated with an increasing concentration of active MRFV PRO domain. The PRO domain used in the experiment was identical to that used in the Ub hydrolase assays (Fig. 2B). The putative PRO cleavage site in the C61A mutant of PRO-HEL<sub>N-TermDomain</sub> was not altered and thus remained cleavable. The results from Figure 4C indicate that even in a 10-fold molar excess of catalytically active PRO, the C61A mutant fusion protein was not processed in *trans* by the PRO domain; no species smaller than the full-length protein, excluding the added PRO domain (~16 kDa), are visible. Should processing have occurred, both the larger PRO portion (~19 kDa) of the PRO-HEL<sub>N-TermDomain</sub> and ATP-binding domain of the HEL (~25 kDa) would have appeared. Taken together, these results indicate that the *Marafivirus* PRO domain extracts itself from the viral polyprotein by *cis* cleavage of the PRO|HEL junction. This differs from previous observations in TYMV in which cleavage at the PRO|HEL junction appears to occur both in *cis* and *trans* (22). However, exclusive in *cis* cleavage is not uncommon in plant-affecting +ssRNA viruses, as is seen in members of the families *Potyviridae* and *Closteroviridae*, whose endopeptidases have been shown to also act solely in *cis* (55, 56). It could be that the release of PRO from the HEL domain enables PRO to adopt a fold that allows for in *trans* cleavage at the HEL|RdRp and RdRp|CP junctions, but this would require further studies.

As mentioned above, many PROs from +ssRNA viruses have auxiliary functions that aid in viral replication, including DUB activity. The C terminus of Ub is composed of the four amino acid motif LRGG. As demonstrated above and consistent with a number of other viral DUBs, *Marafivirus* DUBs can recognize the C terminus (LRGG) motif of Ub and cleave the bond downstream of the diglycine motif (29). Interestingly, the sequences at the PRO|HEL, HEL|RdRp, and RdRp|CP junctions all mimic the C terminus of Ub (Fig. 1A). Indeed, the predicted RdRp|CP junction in MRFV has an exact LRGG sequence and is very likely a PRO cleavage site. Although we did not carry out a cleavage assay as detailed as the PRO|HEL assay described above, Figure 4D illustrates the ability of MRFV PRO to hydrolyze the synthetic fluorogenic peptide LRGG-AMC at the scissile bond between the terminal Gly and AMC. A clear concentration-dependent trend is seen with increasing amounts of enzyme over time. In light of the X-ray structure of MRFV PRO described below, it is not surprising that LRGG-AMC is a poor substrate compared with Ub-AMC, as the peptide would have minimal interactions with the enzyme compared with Ub. Nevertheless, the LRGG-AMC assay indicates that *Marafivirus* PRO domains are able to recognize LRGG alone and very likely cleave at the RdRp|CP junction to liberate the major CP from the replicase proteins. Given our cleavage data, the HEL|RdRp junction is also most likely processed by PRO. The Ub-like LXG[G/A] sequence at the HEL|RdRp junction is conserved among the marafiviral polyproteins (Fig. 1A) and previous results of *Tymoviruses* processing at this junction have been demonstrated (40, 41, 54). Whether HEL|RdRp and RdRp|CP of the *Marafivirus* polyprotein are also processed exclusively in *cis* and the temporal nature of these events remain to be determined.

### Crystal structure of MRFV PRO

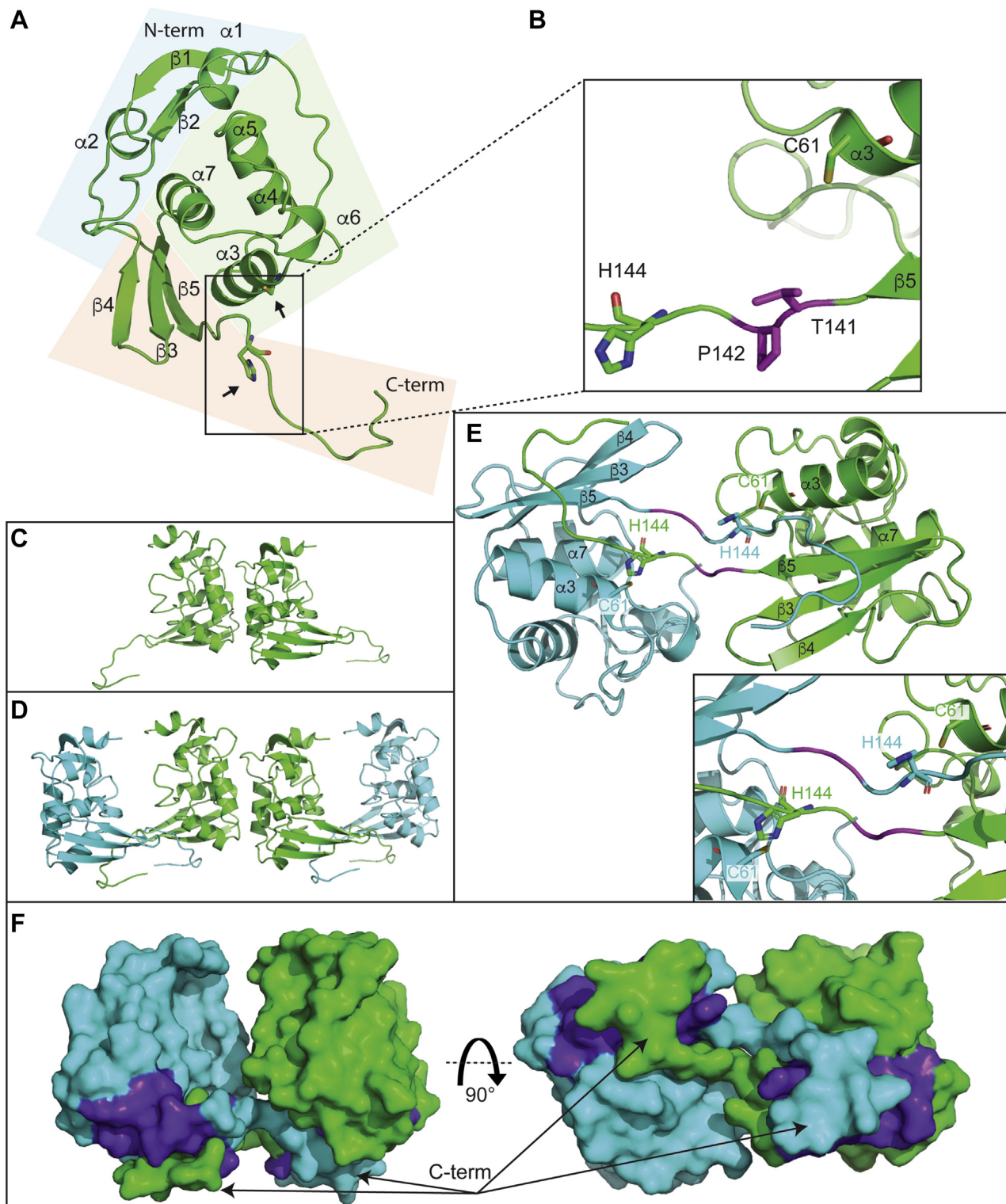
We were able to crystallize MRFV PRO in space group P2<sub>1</sub>, and its three-dimensional structure was determined to a resolution of 1.9 Å using the TYMV PRO X-ray structure (PDB code: 4A5U) as an MR search model. PRO adopts a compact, three-domain fold (Fig. 5A), although its C-terminal conformation is held in place by a neighboring copy of the enzyme (Fig. 5, D and E). As seen in Figure 5A, the first domain (blue panel) is composed of a β-sheet (β1↑ β2↓) and α-helices α1 and α2. The second domain (green panel) is predominantly a well-ordered five-helix bundle (α3-α7). Helix α3 contains the catalytic cysteine nucleophile (C61) with its solvent-exposed thiol group at the N-terminal end of α3 (Fig. 5A (arrows) and Fig. 5B). The third domain (orange panel) is comprised of a three-stranded β-sheet (β4↑ β3↓ β5↓) whose curved, open face packs against helices α3 and α7 of domain II. Domain III terminates with what appears to be a flexible loop that is solvent exposed until it inserts itself into the active site of a neighboring copy of the enzyme. This loop contains the active site histidine (H144) that is ~17 Å from the catalytic cysteine. Based on sequence alignments and a successful MR experiment, MRFV PRO is believed to have a simple Cys/His catalytic dyad similar to TYMV PRO (29).

The ~17 Å distance between the side chains of C61 and H144 does not comprise a functional active site. The asymmetric unit (ASU) is composed of two copies of MRFV PRO (Fig. 5C) with most interactions occurring between the alpha helical bundles of domain II of each molecule. As seen in Figure 5D, the C-terminal tail of each copy of protein in the ASU traverses into its neighbor and appears to complete the active site of the neighboring molecule (Fig. 5E). The C-terminal tails of each symmetry mate jut into the cleft that exists between helices α3 and α7 as well as strand β5 of their neighbor (Fig. 5E). The catalytic C61 and incoming His144 residue of the neighboring symmetry mate arrange themselves to within a few angstroms of each other and appear coordinated for catalysis (inset to Fig. 5E). The active site arrangement of the crystallographic symmetry mates more closely resembles the structured active site of TYMV PRO (Fig. S1). Finally, surface representations (Fig. 5F) reveal the close association of the two MRFV PRO molecules that comprise Figure 5, D and E. The midsection of the C-terminal tails nestles itself into deep grooves formed between helices α3 and α7 of domain II, as well as strand β5 of the sole β-sheet of domain III. The ends of the C-terminal tails seamlessly pack up against the convex face of the β-sheet (purple). MRFV PRO is monomeric in solution according to size-exclusion chromatography (Fig. S2), and this dimeric interaction between the monomers of the ASU is most likely a crystallographic artifact.

### Comparing the PROs of MRFV and TYMV

MRFV PRO and TYMV PRO share a similar three-dimensional fold (Fig. 6A) outside of variability in loops. Loops often play a large role in protein–protein interactions and substrate recognition (57), so it follows that different

## Marafivirus endopeptidases double as deubiquitinases

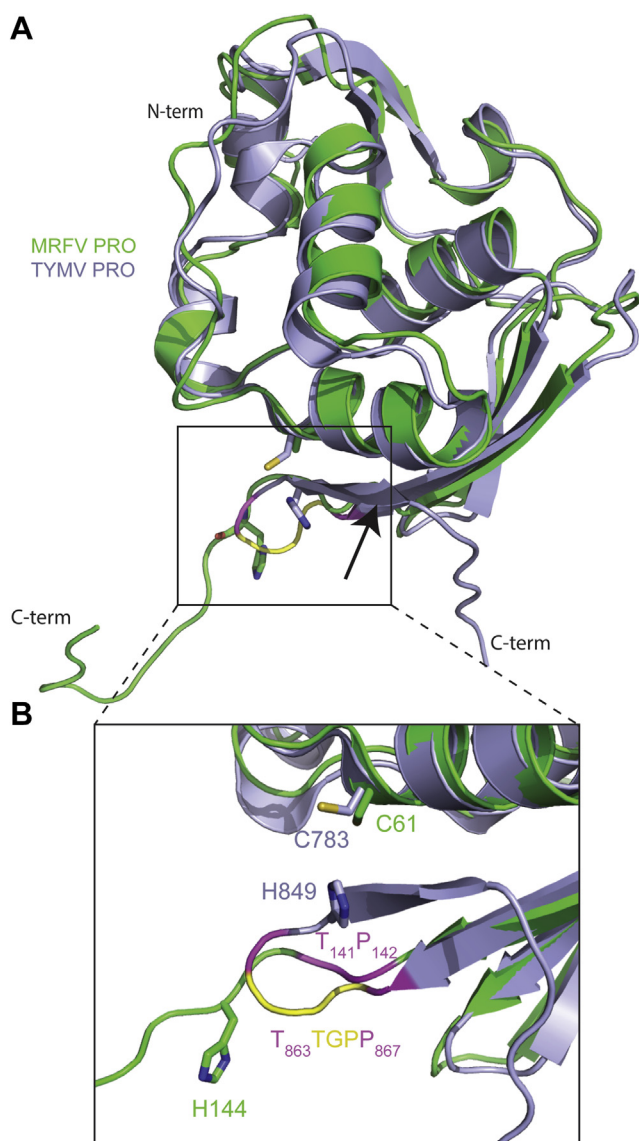


**Figure 5. Crystal structure of MRFV PRO.** *A*, overall structure of MRFV PRO determined at 1.9 Å resolution. Individual domains of PRO are shown in *blocked colors* (domains 1, 2, and 3; *blue, green, and orange*, respectively). The catalytic residues are indicated with *arrows*. *B*, close-up of the active site architecture of MRFV PRO. *C*, asymmetric unit of MRFV PRO crystals. *D*, interaction of MRFV PRO symmetry mates. *E*, close-up on interaction of MRFV PRO symmetry mates, one of each is shown in *green and cyan*. Labeling is identical to that in panels *A* and *B*. *Inset* is a close-up of the symmetry mate active site residues. *F*, surface representation of MRFV PRO symmetry mates at two different angles. Figures were generated in PyMOL (71). MRFV, maize rayado fino virus; PROs, proteases.

substrate/interaction requirements would manifest in loop variability between the MRFV and TYMV PROs. The MRFV PRO structure was compared with known three-dimensional

protein structures using the DALI server (58), which revealed TYMV PRO to be the closest structural homologue, with a Z-score of 20.4 (PDB code: 4A5U). The next





**Figure 6. Crystal structure of MRFV PRO superposed with TYMV PRO.** Overview (A) and close-up (B) of MRFV PRO overlaid with TYMV PRO (PDB code: 4A5U). MRFV PRO is shown in green, and TYMV PRO is shown in pale purple. T141 and P142 (MRFV) and the corresponding T867 and P867 (TYMV) are shown in magenta. The TYMV PRO loop composed of G<sup>864</sup>P<sup>865</sup>P<sup>866</sup> is shown in yellow. Figures were generated in PyMOL (71). MRFV, maizayado fino virus; PROs, proteases; TYMV, turnip yellow mosaic virus.

closest structural homologue was the ovarian tumor domain-containing protein 3 (OTUD3) from *Homo sapiens* with a Z-score of 5.9 (PDB code: 4BOU), followed by OTUD1 from *Saccharomyces cerevisiae* with a Z-score of 5.5 (PDB code: 3C0R). TYMV PRO itself has been characterized as a viral OTU DUB based on its overall core fold (24, 59) and appears to have more homology with OTUD3 and OTUD1 with Z-scores of 7.6 and 7.4, respectively. The difference in Z-scores of each PRO domain with OTUD1/3 can primarily be attributed to large variations in the C termini of MRFV and TYMV PRO. As mentioned previously, the C-terminal of the MRFV PRO structure differs compared with TYMV. For this reason, TYMV PRO has a more organized active site with its catalytic Cys783 and His849 residues in close coordinating

distance, even in the absence of Ub substrate (Fig. 6B). Despite their differences, the structural homology of MRFV PRO shared with TYMV PRO, OTUD1, and OTUD3 clearly classifies it as a viral OTU DUB.

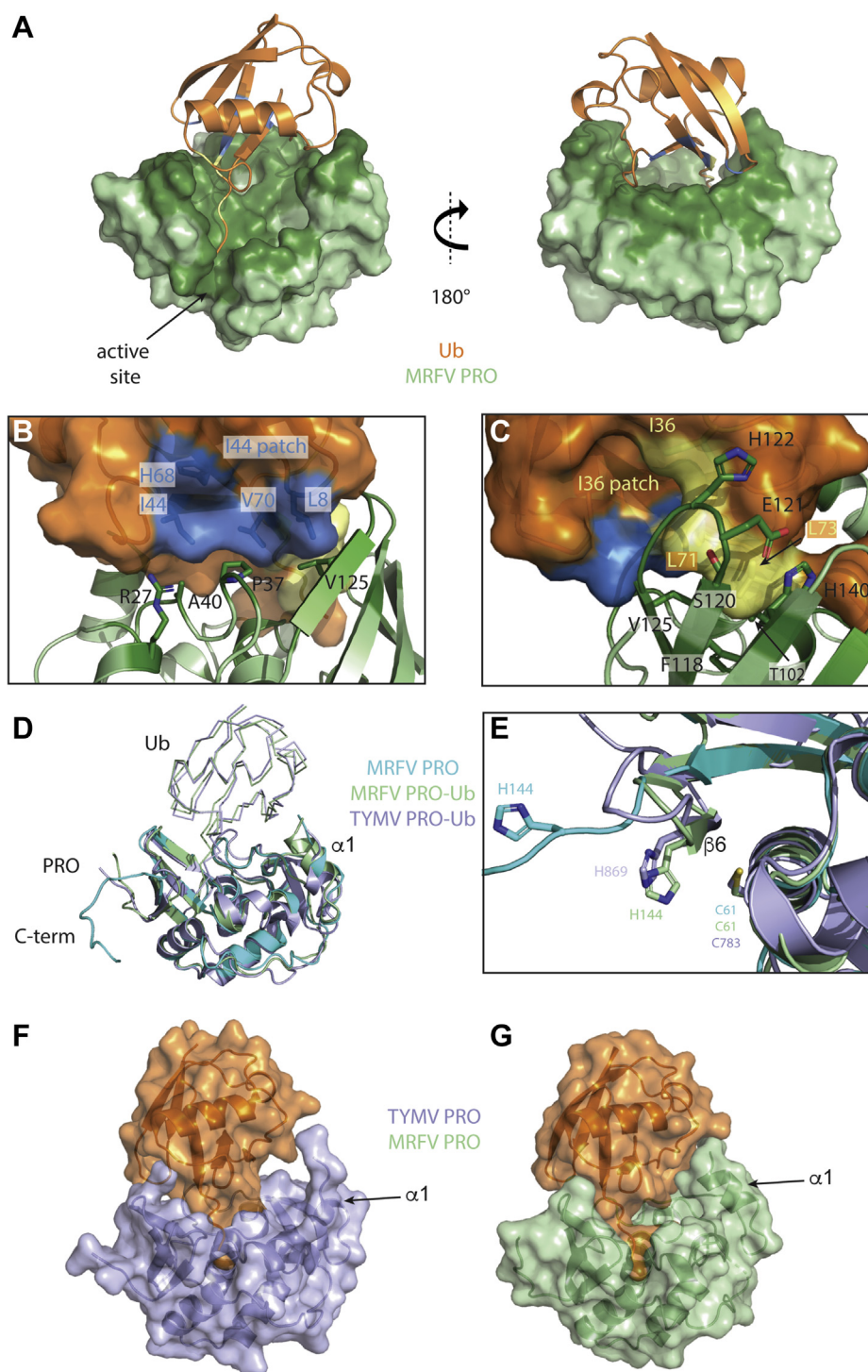
As seen in the sequence of TYMV PRO, there is a region at the C-terminal upstream of the catalytic His known as the “GPP flap” (outlined in black) (Fig. 1B). This motif has been shown to be essential for the PRO to toggle between endopeptidase and DUB activity (60). Mutations in the region decreased DUB activity but did not appear to hinder polyprotein processing at the PRO|HEL junction, indicating the importance of the flap to modulate Ub-dependent antiviral responses but not polyprotein processing (60). Interestingly, aside from the terminal proline (P142), our crystal structure of MRFV PRO reveals that the enzyme lacks a complete GPP motif (Figs. 1B and 6B). Glycine and proline are key residues in  $\beta$ -turns, and their absence may explain why MRFV PRO fails to form a four-stranded  $\beta$ -sheet following  $\beta$ 5 as seen in the previously determined crystal structure of TYMV PRO (Fig. 6) (24). Interestingly, with the exception of MRFV PRO, all the *Marafivirus* PRO domains in our study appear to contain a loop region similar to the GPP loop of TYMV PRO, yet we found them to be fully capable of cleaving the viral polyprotein and poly-Ub chains. Indeed, as described below, we found that interactions of the enzyme with Ub prompts MRFV PRO to adopt a conformation that brings H144 close to C61 and generate a complete active site within a single monomer of MRFV PRO that would turnover Ub.

#### The structure of MRFV PRO in complex with Ub

MRFV PRO was covalently linked to Ub. Specifically, Ub-3Br is a suicide substrate of deubiquitinating enzymes that is a modified form of WT Ub in that its C-terminus is modified to harbor a reactive C-terminal tail, which can irreversibly bind to the active site cysteines of DUBs through a covalent linkage (Fig. S3). The covalently linked protein complex was crystallized in space group I4. Figure 7A shows a remarkably large binding interface between MRFV PRO and Ub. The C-terminal tail of Ub nestles deep into the active site channel of PRO and numerous additional interactions of the enzyme with the beta-grasp fold of Ub (Fig. 7, A and B). As determined through the PISA server (61), there are 38 residues of PRO that are involved with interactions with Ub, which is over 25% of the residues, covering 948 Å<sup>2</sup> (~13%) of the accessible surface area.

Ub has four key surface features that are typically recognized by DUBs and Ub-binding proteins (28, 62). Two of these features are essential in how PRO recognizes Ub, which are the hydrophobic Ile44 and Ile36 patches (Fig. 7, A–C). The Ile44 patch is composed of residues L8, I44, H68, and V70. Figure 7B shows the interactions that occur at the Ile44 patch between PRO and Ub. Four key residues of PRO partake in the stabilizing of Ub on the surface of PRO. Hydrophobic residues P37 and A40 of the long loop connecting  $\alpha$ 1 and  $\alpha$ 2 grip against the hydrophobic Ile44 patch. The pyrrolidine side chain of P37 quite efficiently burrows into the groove in Ub

## Marafivirus endopeptidases double as deubiquitinases



**Figure 7. Crystal structure of MRFV PRO bound to Ub and compared with TYMV PRO-Ub.** *A*, MRFV PRO is shown as a surface representation and in green and Ub as a cartoon in orange. Dark green regions are interacting residues of PRO. *B*, interactions of the Ile44 patch of Ub (shown in blue) with key residues of PRO. *C*, interactions of the Ile36 patch of Ub (shown in yellow) with key residues of PRO. *D*, superposition of MRFV PRO, MRFV PRO-Ub, and TYMV PRO-Ub (PDB code: 6YPT). Ub molecules are shown in ribbons, and PRO domains are shown as cartoons. *E*, superposition of the active sites of MRFV PRO, MRFV PRO-Ub, and TYMV PRO-Ub. *F* and *G* surface representations of TYMV PRO-Ub and MRFV PRO-Ub. Figures were generated in PyMOL (71). MRFV, maize rayado fino virus; PROs, proteases; TYMV, turnip yellow mosaic virus; Ub, ubiquitin.

created by I44 and V70. V125 of PRO also uses its hydrophobic isopropyl side chain to facilitate interactions with the lobe of Ub that arises from L8. Although not hydrophobic, R27 is a key residue for stabilizing Ub binding at this region because its guanidino group hydrogen bonds with Ub's main-chain

carbonyl group of G47 (Fig. S4A). This hydrogen bond orients the guanidino group to press up against the I44 patch and impart another degree of stabilization. Interestingly, only MRFV and OBDV have Arg residues at this position (Fig. 1B, triangle at position ~30). R27, along with A45 (Fig. 1B, triangle

at position ~50), are the only two residues exclusively shared by OBDV and MRFV; however, A45 has no interactions with Ub. R27 is clearly important in MRFV PRO recognition of Ub and could be a factor that we can attribute into MRFV and OBDV lacking the ability to recognize and subsequently process Ub-K<sup>63</sup> chains.

The Ile36 patch of Ub also has many interactions with PRO as seen in Figure 7C. V125, F118, and S120 are all directly involved with hydrophobic contacts with the Ile36 patch, helping to stabilize Ub on the surface of PRO while also facilitating entry and guidance of the C-terminal tail of Ub down into the channel that terminates with the active site. V125 is involved with hydrophobic interactions with both patches, showing its importance in recognizing Ub. Similarly, TYMV PRO has an Ile847 (highlighted by a triangle in Fig. 1B) at this position, and previous experimental findings have shown the importance of this residue in DUB activity (24). Finally, T102 of MRFV PRO forms two hydrogen bonds with L73 of Ub. The first is between the main-chain amide of T102 and the main-chain carbonyl of L73. The second is between the side-chain hydroxyl of T102 and the L73 main-chain amide (Fig. S4, B and C). H122 of PRO forms a hydrogen bond between N<sup>δ1</sup> of its side chain and the side-chain amino group of Q40 of Ub (Fig. S4D). Finally, H140 of PRO forms a critical hydrogen bond between its side-chain N<sup>ε2</sup> and the main-chain carbonyl of R74 of Ub (Fig. S4E). Although not all of these interactions are directly with Ile36 patch residues, they are key in this general vicinity and help stabilize Ub globally when considered together. Furthermore, the interactions of T102 and H140 interact directly through hydrogen bonding with the “LRGG” tail of Ub, suggesting that they also participate in polyprotein substrate recognition.

Overall, the crystal structures of MRFV PRO–Ub and TYMV PRO–Ub (PDB code: 6YPT) are similar (Fig. 7D). The core secondary structure and folds are maintained with only subtle variation within loop regions. Ub binds to both proteins in a similar orientation. Interestingly, in the presence of Ub, the C-terminal tail of MRFV PRO adopts a conformation highly similar to the GPP flap of TYMV PRO (unliganded and liganded) as there now is a turn following β5, which allows for the formation of a 4-stranded β-sheet and a more canonical, rigid active site (Fig. 7, D and E). Furthermore, MRFV PRO forms a much more extensive interaction with Ub than TYMV PRO, with a complex formation significance score (CSS) of 1.000 (scale being 0–1) as determined from the PISA server (61) (Fig. 7, F and G). TYMV PRO–Ub has a CSS of 0.822. Much of this is due to α1 of MRFV PRO, which is more fully formed in the MRFV PRO–Ub structure than TYMV PRO–Ub and may also contribute to poly-Ub chain-type specificity. Interestingly, MRFV PRO (along with SARS-CoV-2 PLpro) have the highest CSS score of viral DUBs that have been structurally characterized bound to Ub, which includes Crimean–Congo hemorrhagic fever virus, equine arteritis virus, murine cytomegalovirus, Dugbe virus, Hazara orthonairovirus, mouse hepatitis virus, and Middle East respiratory and severe acute respiratory syndrome coronaviruses.

## Conclusion

Together, our results provide new structural and biochemical insights into the papain-like cysteine PROs present in the polyproteins of *Marafiviruses*. We demonstrate for the first time that these enzymes have deubiquitinating activity in addition to acting as endopeptidases that process the viral polyprotein. Polyprotein processing assays have provided first insights into how *Marafiviruses* process their polyproteins, which appear to have features distinct from their *Tymovirus* relatives. Our structural findings of the maize-affecting type member MRFV reveal that the enzyme has regions that appear to be highly dynamic, which assist in recognizing different viral and cellular substrates. Unexplored nuances exist that can still be investigated to understand how certain marafiviral DUBs selectively process only K48 poly-Ub chains and not K63 poly-Ub chains. Structural analysis also reveals that MRFV PRO has one of the most extensive interaction surfaces with Ub. Collectively, our results lay the groundwork in biochemically understanding this class of DUBs and sets the stage for future studies to exploit these enzymes.

## Experimental procedures

### DNA constructs

Synthetic DNA (Integrated DNA technologies) coding for ORFs of PRO domains from MRFV and citrus sudden death-associated virus (CSDaV) was amplified by PCR using primers listed in Table S1 and ligated into pGEX-6P-1 (GE Healthcare) using BamHI and XhoI restriction sites as shown in Table S1 (5' and 3', respectively). The remaining PRO ORFs from blackberry virus S (BIVS), Grapevine Syrah virus 1 (GSyV1), Olive latent virus 3 (OLV3), and oat blue dwarf virus (OBDV) were constructed by GenScript using the same cloning strategy.

The MRFV PRO<sub>C61A</sub>–HEL<sub>N-termDomain</sub> fusion protein expression construct (residues 667–1038 from its polyprotein; UniProtKB-Q91TW9) was generated from codon-optimized synthetic DNA (Integrated DNA technologies) using primers FW 5' GATATACATATGCCGGAACCCGATACC 3'/REV 5' TATATCGGATCCTTAGCAATAA AAGTCTACATAGG 3' (NdeI and BamHI restriction sites shown in bold, respectively) and cloned into pET19b (Novagen) in frame with the native N-terminal His<sub>10</sub> tag. The WT version of this protein (MRFV PRO–HEL<sub>N-termDomain</sub>), which retains the active site cysteine of PRO, was constructed using phosphorylated primers, and site-directed mutagenesis was used to reintroduce the cysteine that was formerly an alanine; FW 5' P-CCGTGCCGCTTGCTTACTGGTCG 3'/REV 5'P-GTTG GATAAGGGATAGAG 3' directly into the pET19b vector containing the insert.

### Protein expression and purification

Expression plasmids for the *Marafivirus* PRO domains were used to transform *Escherichia coli* BL21 (DE3) GOLD cells (Stratagene) for protein production. Transformed *E. coli* were grown overnight at 37 °C in LB containing 150 μg/ml

## Marafivirus endopeptidases double as deubiquitinases

ampicillin. The overnight culture was then used to inoculate 500 ml or 1 l of fresh ampicillin-containing LB (1:50 dilution) and was subsequently grown at 37 °C with shaking to an  $A_{600}$  of 0.7 to 0.8. Expression of the GST-tagged PRO enzymes (from pGEX-6P-1 constructs) or His<sub>10</sub>-tagged MRFV PRO<sub>C61A-HEL<sub>N-termDomain</sub></sub> was induced by the addition of 0.5 mM IPTG and left to incubate with shaking at 16 °C for an additional 18 h. Cells were then pelleted by centrifugation and either immediately used or stored at -80 °C.

All *Marafivirus* PRO domains were purified as follows. Cell pellets were resuspended in ice-cold lysis buffer (50 mM TRIS-HCl pH 8.0, 300 mM NaCl and 2 mM DTT) and lysed using an Avestin Emulsiflex C3 high-pressure cell homogenizer (ATA Scientific Instruments). Cell lysates were clarified by centrifugation (17,211g at 4 °C), and the supernatant containing GST-*Marafivirus* \_PRO was mixed end-over-end for 1 h at 4 °C with GST-Bind resin (Millipore) that had been pre-equilibrated in the lysis buffer. The lysate/resin slurry was poured into a gravity column and washed with ~20 column volumes of the lysis buffer, followed by elution of the fusion protein with the lysis buffer supplemented with 10 mM reduced glutathione (adjusted to pH 8.0 with NaOH). The GST tag was removed from each PRO domain using GST-tagged HRV 3C PreScission Protease (GE Healthcare), which was incubated with the eluted fusion protein in dialysis tubing overnight at 4 °C against 2 l of dialysis/gel filtration buffer (20 mM Tris HCl, pH 8.0, 150 mM NaCl, and 2 mM DTT). Tag-free PRO domains were separated from free GST and HRV 3C PreScission Protease using a Superdex 75 (GE Healthcare) gel filtration column. The concentration of each purified *Marafivirus* PRO was quantified using a NanoDrop (Thermo Fisher Scientific) instrument ( $A_{280}$ ,  $\epsilon/1000$  BIVS 11.46, CSDaV 16.96, GSyV1 13.98, MRFV 8.48, OBDV 9.97, OLV3 13.98 M<sup>-1</sup> cm<sup>-1</sup>).

MRFV\_PRO<sub>C61A-HEL<sub>N-termDomain</sub></sub> and the WT version-containing cell pellets were resuspended in ice-cold lysis buffer (50 mM Tris HCl, pH 8.0, 300 mM NaCl, 2 mM DTT, and 5 mM imidazole) and lysed identically to the PRO domains. Cell lysates were clarified by centrifugation (17,211g at 4 °C), and supernatants containing either protein were mixed end-over-end for 1 h at 4 °C with nickel-nitrilotriacetic acid resin (Qiagen) that had been pre-equilibrated in the lysis buffer. The lysate/resin slurry was then poured into a gravity column and washed with ~20 column volumes of the lysis buffer, followed by ten column volumes of the lysis buffer supplemented with 15 mM imidazole, followed by ten column volumes of the lysis buffer supplemented with 30 mM imidazole and finally eluted with the lysis buffer supplemented with 250 mM imidazole. The eluted proteins were dialyzed against 2 l of dialysis/gel filtration buffer (20 mM TRIS-HCl pH 8.0, 150 mM NaCl and 2 mM DTT) overnight at 4 °C and then further purified using a Superdex 75 gel filtration column. The concentrations of purified MRFV\_PRO<sub>C61A-HEL<sub>N-termDomain</sub></sub> or WT were quantified using a NanoDrop instrument ( $A_{280}$ ,  $\epsilon/1000$  = 39.42 M<sup>-1</sup> cm<sup>-1</sup>).

Ubiquitin(1–75)–3-bromopropylamine (Ub-3Br) was prepared and purified as previously described (47, 63) for covalent

coupling to MRFV PRO. This is a version of human Ub (UniProt P62987) lacking the terminal Gly. Purified Ub-3Br was dialyzed in 20 mM Tris HCl, pH 8.0, 150 mM NaCl, and 2 mM DTT, quantified by a Bradford protein assay and coupled with MRFV PRO in a 2-fold molar excess at 4 °C for 16 h with the addition of Tris(2-carboxyethyl)phosphine hydrochloride to a final concentration of 5 mM. The resulting MRFV PRO–Ub complex was separated from excess Ub-3Br by gel filtration.

### Enzyme assays

All six purified *Marafivirus* PROs were assayed against the fluorogenic substrate analogue 7-amino-4-methylcoumarin (AMC)–Ub (Ub-AMC) (Boston Biochem) or the synthetic peptide analogue LRGG-AMC (GenScript). The latter substrate represents the C-terminal motif of Ub as well as the linker between the RdRp domain and major CP of the MRFV polyprotein. Reaction buffer for all assays consisted of 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 2 mM DTT. Substrates in the reaction buffer was placed in a black, flat-bottom 96-well microplate (Corning Life Sciences), and the enzyme was added immediately before readings. The final reaction volume was 100  $\mu$ l. Time-course kinetics assays were carried out using a SpectraMax iD5 microplate reader (Molecular Devices). The instrument's monochromators were set to excitation of 345 nm and emission of 445 nm. The PMT gain was set to medium, and reads were taken every 9 to 11 s.

### Polyubiquitin chain hydrolysis assays

Two hundred nanograms of the substrate (K48- or K63-linked poly-ubiquitin chains [Ub<sub>3</sub>-Ub<sub>6</sub>; Boston Biochem]) was incubated with 50 to 200 nM of each *Marafivirus* DUB in a reaction buffer identical to the abovementioned kinetics assay. Each reaction was incubated for 30 min at 25 °C. Reactions were terminated with the addition of 2X SDS-PAGE loading buffer. Reactions were visualized by carrying out TRIS-Tricine PAGE (10%) and subsequent detection using a Pierce Silver Stain Kit (Thermo Fisher Scientific).

### Protein crystallization

MRFV PRO was crystallized using the vapor diffusion method at 15 mg/ml in a condition that contained 100 mM Bis-Tris propane (pH 7.5), 200 mM sodium acetate, and 20% PEG 3350. Crystals appeared after ~30 days at 4 °C. Crystals of the MRFV PRO–Ub complex were also grown using the vapor diffusion method in 100 mM phosphate citrate buffer (pH 3.8), 200 mM lithium sulfate, and 25% PEG 1000, which appeared after 1 day at 4 °C. Immediately before X-ray data collection, single crystals of both PRO and Ub-bound PRO were swept through a cryoprotectant composed of the initial crystallization condition supplemented with 25 or 15% glycerol (PRO and Ub-bound PRO, respectively) and subsequently flash-cooled in liquid nitrogen.

**Table 1**  
Crystallographic and refinement statistics for MRFV PRO and MRFV PRO–Ub structures

	MRFV PRO	MRFV PRO–Ub
X-ray source	Rigaku R-AXIS IV++	Rigaku R-AXIS IV++
Crystal geometry		
Space group	P2 <sub>1</sub>	I4
Unit cell (Å)	$a = 43.20$ $b = 73.26$ $c = 54.34$ ; $\alpha = \beta = \gamma = 90^\circ$	$a = b = 75.72$ $c = 79.56$ ; $\alpha = \beta = \gamma = 90^\circ$
Crystallographic data		
Wavelength (Å)	1.5418	1.5418
Resolution range (Å)	30.90–1.90 (1.97–1.90) <sup>a</sup>	27.42–2.09 (2.15–2.09) <sup>a</sup>
Total observations	89,460 (5639)	193,427 (15,741)
Unique reflections	27,471 (1718)	13,345 (1109)
Multiplicity	3.4 (3.3)	14.5 (14.2)
Completeness (%)	98.9 (99.6)	100 (100)
<i>R</i> <sub>merge</sub>	0.070 (0.41)	0.091 (0.36)
CC1/2	0.99 (0.74)	0.99 (0.97)
I/ $\sigma$ I	7.03 (2.06)	26.3(8.0)
Wilson B-factor (Å <sup>2</sup> )	22.91	14.5
Refinement statistics		
Reflections in test set	2366 (235)	1296 (138)
Protein atoms	2175	1683
Ligands	0	16
Solvent molecules	169	239
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.20/0.24	0.20/0.26
RMSDs		
Bond lengths/ angles (Å/°)	0.007/0.84	0.002/0.55
Ramachandran plot		
Favored/allowed (%)	97.23/2.42	97.65/1.41
Average B factor (Å <sup>2</sup> )	25.77	22.63
Macromolecules	25.34	21.76
Ligands	-	27.96
Solvent	31.38	28.42

<sup>a</sup> Values in parentheses refer to the highest resolution shell.

### X-ray data collection and structure determination

X-ray diffraction data were collected in-house at 100 K using a Rigaku MicroMax HF X-ray generator and R-AXIS IV++ image plate detector. Data were indexed using XDS for MRFV PRO (64) and iMOSFLM for MRFV PRO–Ub (65). Scaling was done using Aimless (66) as a part of the CCP4 i2 program suite (67). For the unliganded MRFV PRO structure determination, molecular replacement (MR) was carried out using the crystal structure of TYMV PRO with its cocrystallized contaminant excluded (PDB code: 4A5U). MR was done using PHENIX.PHASER (68) and was followed by model building using PHENIX.AUTOBUILD (68). Iterative model building and refinement was done using COOT (69) and PHENIX.REFINE (68). Structure determination for MRFV PRO bound to Ub was carried out almost identically; however, a multicomponent MR search was carried out using the structures of the unliganded form of MRFV PRO (determined herein) and Ub (PDB code: 1UBQ). Crystallographic and refinement statistics are provided in Table 1.

### Data availability

X-ray structures reported in this article are available in the Protein Data Bank (PDB) under PDB IDs 7MIA and 7MIC. All other data are presented in the article.

**Supporting information**—This article contains supporting information.

**Author contributions**—A. P. and B. L. M. conceptualization; A. P. and B. L. M. validation; A. P., J. A. M. M., and B. L. M. investigation; A. P. and B. L. M. visualization; A. P. and B. L. M. methodology; A. P. writing—original draft; A. P. and B. L. M. writing—review and editing; B. L. M. resources; B. L. M. data curation; B. L. M. formal analysis; B. L. M. supervision; B. L. M. funding acquisition; B. L. M. project administration.

**Funding and additional information**—This work was funded by a Discovery Grant to B. L. M. from the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2020-05682 and RGPIN-2015-05310).

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: +ssRNA, positive-sense ssRNA; AMC, 7-amino-4-methylcoumarin; ASU, asymmetric unit; BLVS, blackberry virus S; CP, coat protein; CSDaV, citrus sudden death-associated virus; CSS, complex formation significance score; DUB, deubiquitinase; GSyV1, Grapevine Syrah virus 1; HEL, helicase; MR, molecular replacement; MRFV, maize rayado fino virus; OBDV, oat blue dwarf virus; OLV3, Olive latent virus 3; OTUD3, ovarian tumor domain-containing protein 3; PROs, proteases; RdRp, RNA-dependent RNA polymerase; TYMV, turnip yellow mosaic virus; TYMV PRO, papain-like cysteine protease of TYMV; Ub, ubiquitin; Ub-3Br, ubiquitin(1–75)–3-bromopropylamine.

### References

- Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1994) Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol.* **68**, 5045–5055
- Firth, A. E., and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409
- Spall, V. E., Shanks, M., and Lomonosoff, G. P. (1997) Polyprotein processing as a strategy for gene expression in RNA viruses. *Semin. Virol.* **8**, 15–23
- Boudreault, S., Roy, P., Lemay, G., and Bisailon, M. (2019) Viral modulation of cellular RNA alternative splicing: A new key player in virus–host interactions? *Wiley Interdiscip. Rev. RNA* **10**, e1543
- Yost, S. A., and Marcotrigiano, J. (2013) Viral precursor polyproteins: Keys of regulation from replication to maturation. *Curr. Opin. Virol.* **3**, 137–142
- Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., and Koonin, E. V. (2018) Origins and evolution of the global RNA virome. *mBio* **9**, e02329-18
- Dolja, V. V., and Koonin, E. V. (2011) Common origins and host-dependent diversity of plant and animal viromes. *Curr. Opin. Virol.* **1**, 322–331
- Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., and Mortensen, D. A. (2017) Agriculture in 2050: Recalibrating targets for sustainable intensification. *Bioscience* **67**, 386–391
- Calil, I. P., and Fontes, E. P. B. (2017) Plant immunity against viruses: Antiviral immune receptors in focus. *Ann. Bot.* **119**, 711–723
- Hammond, R. W., and Ramirez, P. (2001) Molecular characterization of the genome of maize rayado fino virus, the type member of the genus Marafivirus. *Virology* **282**, 338–347
- Edwards, M. C., and Weiland, J. J. (2014) Coat protein expression strategy of oat blue dwarf virus. *Virology* **450–451**, 290–296
- Ruiz-García, A. B., Sabaté, J., Lloria, O., Laviña, A., Batlle, A., and Olmos, A. (2017) First report of grapevine syrah virus-1 in grapevine in Spain. *Plant Dis.* <https://doi.org/10.1094/PDIS-05-17-0700-PDN>
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018) Virus taxonomy: The database of the

## Marafivirus endopeptidases double as deubiquitinases

- International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717
- Nemchinov, L. G., François, S., Roumagnac, P., Ogliastro, M., Hammond, R. W., Mollov, D. S., and Filloux, D. (2018) Characterization of alfalfa virus F, a new member of the genus Marafivirus. *PLoS One* **13**, 1–11
  - Kim, H., Park, D., and Hahn, Y. (2018) Identification of novel RNA viruses in alfalfa (*Medicago sativa*): An alphapartitivirus, a deltapartitivirus, and a marafivirus. *Gene* **638**, 7–12
  - Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., and Redinbaugh, M. G. (2008) Insect vector interactions with persistently transmitted viruses. *Annu. Rev. Phytopathol.* **46**, 327–359
  - Hammond, R. W., Edwards, M. C., and Ramirez, P. (2011). In: Tidona, C., Darai, G., eds. *The Springer Index of Viruses*, Springer New York, New York, NY
  - Nault, L. R. (1980) Leafhopper transmission and host range of maize rayado fino virus. *Phytopathology* **70**, 709
  - Edwards, M. C., Weiland, J. J., Todd, J., and Stewart, L. R. (2015) Infectious maize rayado fino virus from cloned cDNA. *Phytopathology* **105**, 833–839
  - Zambrano, J. L. (2013) Identification of resistance to maize rayado fino virus in maize inbred lines. *Plant Dis.* **97**, 1418–1823
  - Chenon, M., Camborde, L., Cheminant, S., and Jupin, I. (2012) A viral deubiquitylating enzyme targets viral RNA-dependent RNA polymerase and affects viral infectivity. *EMBO J.* **31**, 741–753
  - Jakubiec, A., Drugeon, G., Camborde, L., and Jupin, I. (2007) Proteolytic processing of turnip yellow mosaic virus replication proteins and functional impact on infectivity. *J. Virol.* **81**, 11402–11412
  - Moriceau, L., Jomat, L., Bressanelli, S., Alcaide-Loridan, C., and Jupin, I. (2017) Identification and molecular characterization of the chloroplast targeting domain of turnip yellow mosaic virus replication proteins. *Front. Plant Sci.* **8**, 1–19
  - Lombardi, C., Ayach, M., Beaurepaire, L., Chenon, M., Andreani, J., Guerois, R., Jupin, I., and Bressanelli, S. (2013) A compact viral processing proteinase/ubiquitin hydrolase from the OTU family. *PLoS Pathog.* **9**, e1003560
  - Fieulaine, S., Witte, M. D., Theile, C. S., Ayach, M., Ploegh, H. L., Jupin, I., and Bressanelli, S. (2020) Turnip yellow mosaic virus protease binds ubiquitin suboptimally to fine-tune its deubiquitinase activity. *J. Biol. Chem.* **295**, 13769–13783
  - Rodamilans, B., Shan, H., Pasin, F., and García, J. A. (2018) Plant viral proteases: Beyond the role of peptide cutters. *Front. Plant Sci.* **9**, 1–11
  - Camborde, L., Jupin, I., Planchais, S., Tournier, V., Jakubiec, A., Drugeon, G., Lacassagne, E., Pflieger, S., Chenon, M., and Jupin, I. (2010) The ubiquitin-proteasome system regulates the accumulation of Turnip yellow mosaic virus RNA-dependent RNA polymerase during viral infection. *Plant Cell* **22**, 3142–3152
  - Komander, D., and Rape, M. (2012) The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229
  - Komander, D., Clague, M. J., and Urbé, S. (2009) Breaking the chains: Structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **10**, 550–563
  - Hua, Z., and Vierstra, R. D. (2011) The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **62**, 299–334
  - Miricescu, A., Goslin, K., and Graciet, E. (2018) Ubiquitylation in plants: Signaling hub for the integration of environmental signals. *J. Exp. Bot.* **69**, 4511–4527
  - Xu, P., and Peng, J. (2006) Dissecting the ubiquitin pathway by mass spectrometry. *Biochim. Biophys. Acta* **1764**, 1940–1947
  - Bartel, B., and Citovsky, V. (2012) Focus on ubiquitin in plant biology. *Plant Physiol.* **160**, 1
  - Johnson, A., and Vert, G. (2016) Unraveling K63 polyubiquitination networks by sensor-based proteomics. *Plant Physiol.* **171**, 1808–1820
  - Braten, O., Livneh, I., Ziv, T., Admon, A., Kehat, I., Caspi, L. H., Gonen, H., Bercovich, B., Godzik, A., Jahandideh, S., Jaroszewski, L., Sommer, T., Kwon, Y. T., Guharoy, M., Tompa, P., et al. (2016) Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4639–E4647
  - Saeki, Y., Kudo, T., Sone, T., Kikuchi, Y., Yokosawa, H., Toh-e, A., and Tanaka, K. (2009) Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. *EMBO J.* **28**, 359–371
  - Zhou, B., and Zeng, L. (2017) Conventional and unconventional ubiquitination in plant immunity. *Mol. Plant Pathol.* **18**, 1313–1330
  - Martelli, G. P., Sabanadzovic, S., Sabanadzovic, N. A. G., Edwards, M. C., and Dreher, T. (2002) The family Tymoviridae. *Arch. Virol.* **147**, 1837–1846
  - Jakubiec, A., Notaise, J., Tournier, V., Hericourt, F., Block, M. A., Drugeon, G., van Aelst, L., and Jupin, I. (2004) Assembly of turnip yellow mosaic virus replication complexes: Interaction between the proteinase and polymerase domains of the replication proteins. *J. Virol.* **78**, 7945–7957
  - Kadare, G., Rozanov, M., and Haenni, A. L. (1995) Expression of the turnip yellow mosaic virus proteinase in *Escherichia coli* and determination of the cleavage site within the 206 kDa protein. *J. Gen. Virol.* **76**, 2853–2857
  - Bransom, K. L., Wallace, S. E., and Dreher, T. W. (1996) Identification of the cleavage site recognized by the turnip yellow mosaic virus protease. *Virology* **217**, 404–406
  - Isono, E., and Nagel, M. K. (2014) Deubiquitylating enzymes and their emerging role in plant biology. *Front. Plant Sci.* **5**, 1–6
  - Bailey-Elkin, B. A., Knaap, R. C. M., Johnson, G. G., Dalebout, T. J., Ninaber, D. K., Van Kasteren, P. B., Bredenbeek, P. J., Snijder, E. J., Kikkert, M., and Mark, B. L. (2014) Crystal structure of the middle east respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity to demonstrate its role in innate immune suppression. *J. Biol. Chem.* **289**, 34667–34682
  - Van Kasteren, P. B., Bailey-Elkin, B. A., James, T. W., Ninaber, D. K., Beugeling, C., Khajehpour, M., Snijder, E. J., Mark, B. L., and Kikkert, M. (2013) Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E838–E847
  - Bailey-Elkin, B. A., Knaap, R. C. M., Kikkert, M., and Mark, B. L. (2017) Structure and function of viral deubiquitinating enzymes. *J. Mol. Biol.* **429**, 3441–3470
  - Dang, L. C., Melandri, F. D., and Stein, R. L. (1998) Kinetic and mechanistic studies on the hydrolysis of ubiquitin C-terminal 7-amido-4-methylcoumarin by deubiquitinating enzymes. *Biochemistry* **37**, 1868–1879
  - Messick, T. E., Russell, N. S., Iwata, A. J., Sarachan, K. L., Shiekhattar, R., Shanks, J. R., Reyes-Turcu, F. E., Wilkinson, K. D., and Marmorstein, R. (2008) Structural basis for ubiquitin recognition by the Otul ovarian tumor domain protein. *J. Biol. Chem.* **283**, 11038–11049
  - Capodagli, G. C., Deaton, M. K., Baker, E. A., Lumpkin, R. J., and Pegan, S. D. (2013) Diversity of ubiquitin and ISG15 specificity among Nairoviruses' viral ovarian tumor domain proteases. *J. Virol.* **87**, 3815–3827
  - Walsh, C. K., and Sadanandom, A. (2014) Ubiquitin chain topology in plant cell signaling: A new facet to an evergreen story. *Front. Plant Sci.* **5**, 122
  - Snijder, E. J., Kikkert, M., and Fang, Y. (2013) Arterivirus molecular biology and pathogenesis. *J. Gen. Virol.* **94**, 2141–2163
  - Emmott, E., de Rougemont, A., Hosmillo, M., Lu, J., Fitzmaurice, T., Haas, J., and Goodfellow, I. (2019) Polyprotein processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity. *J. Biol. Chem.* **294**, 4259–4271
  - Rausalu, K., Utt, A., Quirin, T., Varghese, F. S., Žusinaite, E., Das, P. K., Ahola, T., and Merits, A. (2016) Chikungunya virus infectivity, RNA replication and non-structural polyprotein processing depend on the nsP2 protease's active site cysteine residue. *Sci. Rep.* **6**, 37124
  - Gildenhuis, S. (2020) Expanding our understanding of the role polyprotein conformation plays in the coronavirus life cycle. *Biochem. J.* **477**, 1479–1482
  - Bransom, K. L., Weiland, J. J., and Dreher, T. W. (1991) Proteolytic maturation of the 206-kDa nonstructural protein encoded by turnip yellow mosaic virus RNA. *Virology* **184**, 351–358

55. Carrington, J. C., Freed, D. D., and Sanders, T. C. (1989) Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and *in vitro*. *J. Virol.* **63**, 4459–4463
56. Peng, C.-W., Peremyslov, V. V., Mushegian, A. R., Dawson, W. O., and Dolja, V. V. (2001) Functional specialization and evolution of leader proteinases in the family Closteroviridae. *J. Virol.* **75**, 12153–12160
57. Espadaler, J., Querol, E., Aviles, F. X., and Oliva, B. (2006) Identification of function-associated loop motifs and application to protein function prediction. *Bioinformatics* **22**, 2237–2243
58. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
59. Bailey-Elkin, B. A., van Kasteren, P. B., Snijder, E. J., Kikkert, M., and Mark, B. L. (2014) Viral OTU deubiquitinases: A structural and functional comparison. *PLoS Pathog.* **10**, e1003894
60. Jupin, I., Ayach, M., Jomat, L., Fieulaine, S., and Bressanelli, S. (2017) A mobile loop near the active site acts as a switch between the dual activities of a viral protease/deubiquitinase. *PLoS Pathog.* **13**, e1006714
61. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
62. Dikic, I., Wakatsuki, S., and Walters, K. J. (2009) Ubiquitin-binding domains from structures to functions. *Nat. Rev. Mol. Cell Biol.* **10**, 659–671
63. Borodovsky, A., Ova, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159
64. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
65. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
66. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214
67. Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., *et al.* (2018) CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. D Struct. Biol.* **74**, 68–84
68. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877
69. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
70. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
71. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.