JBC RESEARCH ARTICLE

Crystal structure of the Rubella virus protease reveals a unique papain-like protease fold

Received for publication, May 8, 2022, and in revised form, July 1, 2022 Published, Papers in Press, July 11, 2022,
https://doi.org/10.1016/j.jbc.2022.102250

Ezekiel Ze Ken Cheong[1,2,‡], Jun Ping Quek[1,3,‡], Liu Xin[4], Chaoqiang Li[5], Jing Yi Chan[6], Chong Wai Liew[2,3], Yuguang Mu[2], Jie Zheng[4,5,7], and Dahai Luo[1,2,3,*]

From the [1]Lee Kong Chian School of Medicine, [2]School of Biological Sciences, and [3]NTU Institute of Structural Biology, Nanyang Technological University, Singapore, Singapore; [4]Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China; [5]School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China; [6]School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore, Singapore; [7]School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China

Edited by Craig Cameron

Rubella, a viral disease characterized by a red skin rash, is well controlled because of an effective vaccine, but outbreaks are still occurring in the absence of available antiviral treatments. The Rubella virus (RUBV) papain-like protease (RubPro) is crucial for RUBV replication, cleaving the nonstructural polyprotein p200 into two multifunctional proteins, p150 and p90. This protease could represent a potential drug target, but structural and mechanistic details important for the inhibition of this enzyme are unclear. Here, we report a novel crystal structure of RubPro at a resolution of 1.64 Å. The RubPro adopts a unique papain-like protease fold, with a similar catalytic core to that of proteases from Severe acute respiratory syndrome coronavirus 2 and foot-and-mouth disease virus while having a distinctive N-terminal fingers domain. RubPro has well-conserved sequence motifs that are also found in its newly discovered *Rubivirus* relatives. In addition, we show that the RubPro construct has protease activity in *trans* against a construct of RUBV protease–helicase and fluorogenic peptides. A protease–helicase construct, exogenously expressed in *Escherichia coli*, was also cleaved at the p150–p90 cleavage junction, demonstrating protease activity of the protease–helicase protein. We also demonstrate that RubPro possesses deubiquitylation activity, suggesting a potential role of RubPro in modulating the host's innate immune responses. We anticipate that these structural and functional insights of RubPro will advance our current understanding of its function and help facilitate more structure-based research into the RUBV replication machinery, in hopes of developing antiviral therapeutics against RUBV.

Rubella is an infectious disease that is well characterized by rashes (1). It is often confused with measles that also cause rashes. Confoundingly, rubella is also known as German measles, and measles is also known as rubeola (2). However, rubella and measles are different diseases caused by different viruses. The Rubella virus (RUBV) is the etiological agent of the Rubella disease and belongs to the genus, *Rubivirus*, of the newly created family, *Matonaviridae* (3). Rubella infection during the first trimester of pregnancy can result in miscarriage or congenital rubella syndrome (CRS). CRS is characterized by fetal cataracts, deafness, heart defects, and global developmental delay (4). Infection at the early stages of pregnancy typically has the worst prognosis (5). There is currently no treatment for CRS apart from symptomatic treatment (6). Before the rubella vaccine was developed in 1969, rubella epidemics occurred every 6 to 9 years (7). Modern rubella vaccines utilize the RA27/3 strain (8), a strain of RUBV obtained from an aborted fetus infected with the virus (9). The vaccine is typically administered in a combination of measles, mumps, and rubella vaccines. With 97% effectiveness (10), the vaccine has virtually eliminated rubella in more than 130 countries (11).

RUBV is a group VI virus with a single-stranded positive-sense RNA genome enclosed by an icosahedral capsid (4, 12). The viral genome is around 10 kb in size and has the highest GC content of RNA viruses, at 70% (13). The genome has a 5′ cap structure with a poly(A) tail at its 3′end. The 5′-proximal ORF encodes the nonstructural polypeptide p200, whereas the 3′-proximal ORF encodes the structural proteins, capsid, and surface glycoproteins E1 and E2 (Fig. 1*A*) (14). The nonstructural polyprotein p200 is then processed into two nonstructural proteins, p150 and p90 (15). The p150 protein consists of a methyltransferase and protease domains (16), whereas the p90 protein has both helicase and RNA-dependent RNA polymerase domain (17). These nonstructural proteins are crucial to RNA viruses for replication (18) and polyprotein processing (19).

The RUBV protease, RubPro, while in the p200, cleaves the polyprotein between residues G1301 and G1302 (20). This cleavage, at SRGG/GTCA, is found at the C-terminal region of RubPro in p150 and the N-terminal region of the helicase domain of p90 (Fig. 1*B*) (21). Based on computer alignments, RubPro is predicted to be a papain-like cysteine protease (PCP). PCPs form a large family among cysteine proteases and can be found in plants, viruses, and parasites (22). Coronaviruses and alphaviruses utilize PCPs for polypeptide processing and immune evasion mechanisms (23, 24). Severe acute respiratory

‡ These authors contributed equally to this work.
* For correspondence: Dahai Luo, luodahai@ntu.edu.sg.
Present address for Ezekiel Ze Ken Cheong: Duke-NUS Medical School, Singapore.

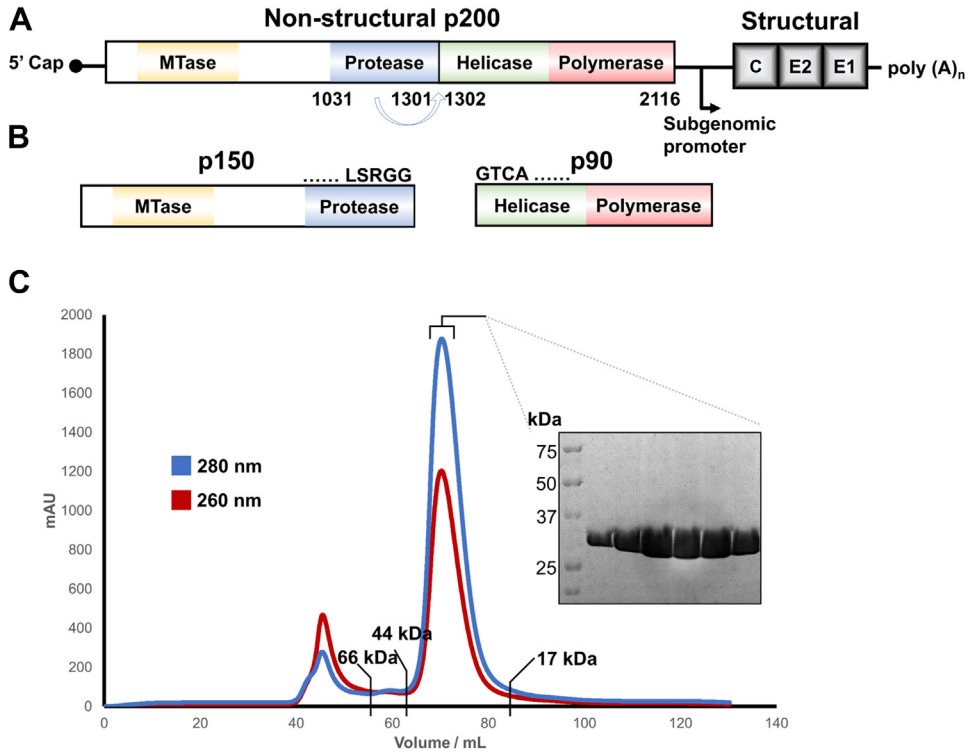

Figure 1. RUBV genome and RUBV protease function and expression. *A*, schematic representation of the RUBV genome. The 5′ ORF encodes for the nonstructural protein p200, whereas the 3′ ORF encodes the structural proteins, capsid (C), and surface glycoproteins E1 and E2. The enzymatic domains of the nonstructural polypeptide are labeled, with the protease domain highlighted in *blue*. *B*, the protease domain located within p150 will cleave the nonstructural p200 into p150 and p90. The cleavage site is at SRGG/GTCA. *C*, the gel filtration profile of RubPro on Superdex 75 16/60 column (GE Healthcare). SDS-PAGE analysis of the RubPro with an expected molecular weight of 30.3 kDa. The elution volumes of three other standard proteins are labeled as reference. RUBV, Rubella virus; RubPro, RUBV papain-like protease.

syndrome coronavirus 2 (SARS-CoV-2) papain-like protease (PLpro) (25) and foot-and-mouth disease (FMDV) PCP (26, 27) cleaves inflammatory ubiquitin (Ub) and other Ub-like antiviral proteins, such as interferon-stimulated gene 15 (ISG15), to modulate host innate immunity responses (28). The functional importance of PCPs for viral replication and survival makes them viable drug targets for antiviral treatment (29). RubPro has residues C1152 and H1273 as the catalytic dyad (30), confirmed *via* site-directed mutagenesis (31). RubPro requires divalent ions like Zn^{2+}, Co^{2+}, and Cd^{2+} for protease activity (32). In addition, the Ca^{2+}-dependent association of calmodulin with RubPro is necessary for the proteolytic activity of RubPro (33). Using FMDV PCP as a reference, homology modeling has been used to partially model the RubPro structure (34). RubPro contains a Ca^{2+}-binding EF-hand domain, which plays a role in maintaining the structure of RubPro (34). The cysteine residues are arranged close to Zn^{2+}, and this Zn^{2+} coordination also contributes to structural stability (35). The structure of RubPro has not been solved, and its enzymatic activity or inhibition has not been characterized quantitatively.

Scientific knowledge of RUBV is relatively superficial compared with other viruses, likely because of the effective vaccine reducing the need for deeper research. The only proteins of RUBV with solved structures are the structural proteins E1 (36) and capsid protein (37). The enzymatic and essential nonstructural proteins' structures and interactions are not well understood or adequately characterized. While only RUBV is found naturally in humans (38), two newly discovered *Rubivirus*, Ruhugu virus (RUHV) and Rustrela virus (RUSV), have crossed barriers across other mammalian host species. This raised concerns about the potential zoonotic transmission of RUBV-like viruses into human hosts (39). Furthermore, despite an effective RUBV vaccine, there are still rubella outbreaks occurring because of gaps in national immunization programs. Japan experienced three rubella outbreaks during the last decades (40, 41). There were outbreaks in Poland (42) and Romania (43) in 2012 as well. Once an outbreak manages to break the first line of defense, which is vaccine coverage, there are no antiviral treatments available to combat the virus and the disease. Effective antiviral therapeutics requires a fundamental understanding of viral replication processes. Many successful drugs inhibit proteins of the replication machinery, such as proteases (44, 45) and RNA polymerases (46–48). The lack of knowledge about RUBV highlights a need for deeper research into its replication to potentially develop antiviral treatments as well as to expand the knowledge of virology that could enlighten the replication and pathogenesis mechanisms of other viruses.

In this study, the crystal structure of RubPro to a resolution of 1.64 Å is reported. This novel structure provides the molecular basis for the substrate recognition and characterization of the active site for potential inhibitor studies. Next, sequence alignment was conducted with RUHV and RUSV proteases. RubPro was also structurally aligned with other viral PCPs deposited in the Protein Data Bank (PDB) for further

structural analysis. The protease's activity was also characterized using RUBV protease–helicase as the natural substrate as well as using fluorogenic peptides. RubPro was shown to have K48-linkage–specific deubiquitylation activity. This study provides structural and functional insight into RubPro to facilitate antiviral development against RUBV as well as to expand the field of knowledge of viral proteins.

Results

Structure determination of RubPro

RubPro domain of the P150 (residues 1021–1301) was purified as a monomeric protein with an expected molecular weight of 30.3 kDa (Fig. 1*C*). The novel crystal structure of RubPro was determined at a resolution of 1.64 Å collected at Zn K absorption edge of 9.7 KeV (Figs. 2 and S1). The RubPro adopts a unique right-hand architecture with two Zn^{2+}-binding sites and a catalytic Cys–His dyad. The fingers, palm, and thumb domains correspond to the N-terminal, middle, and C-terminal regions of RubPro, respectively. The RubPro fingers domain consists of three short parallel β-sheets and four α-helices. The C1034, C1037, C1040, and H1088 coordinate a Zn^{2+} ion near β3. Next, the palm domain consists of four α-helices, with the catalytic C1152 found on α5. The residues C1175, C1178, C1225, and C1227 coordinate a Zn^{2+} ion near α6. Finally, the thumb domain consists of five β-sheets with the other catalytic residue, H1273, on β7.

Our structure also confirmed the findings of Liu *et al.* (30) that C1175, C1178, C1225, and C1227 are Zn^{2+}-binding cysteines. A previously unknown quartet of Zn^{2+}-binding residues, C1034, C1037, C1040, and H1088, were also uncovered at the N-terminal region of this novel structure. These two Zn^{2+}-coordinating sites were well defined by the characteristic tetrahedral coordination geometry of sulfur and nitrogen atoms from Cys and His, respectively. At the first Zn^{2+}-binding site, the residues C1034, C1037, C1040, and H1088 are located 2.4, 2.3, 2.4, and 2.1 Å away from Zn^{2+}. While at the second Zn^{2+}-binding site, all cysteine residues are 2.3 Å away from Zn^{2+} (Fig. 2*C*).

The catalytic dyad C1152 and H1273 are located on the palm and the thumb, respectively. The positions of these catalytic residues are similar to that of SARS-CoV-2 PLpro, whereby the catalytic triad is also found at the interface between the palm and the thumb regions (49). The sulfur atom of C1152 is 4.2 Å apart from the delta nitrogen of H1273 (Fig. 2*C*). In contrast to other cysteine proteases, no stabilizing Asn/Asp residue was found near H1273.

Sequence alignment of the RubPro with the protease domain of RUHV and RUSV was performed to analyze for the presence of conserved motifs among the *Rubivirus* genus. The protease domain of RUHV and RUSV shows 53% and 40% amino acid similarity, respectively, to RubPro, with conserved catalytic dyad and zinc-coordinating residues (Fig. 2*E*).

Structural alignment with other cysteine proteases

The Dali server was utilized to search for structural homologs in the PDB database (Table S2) (50). We selected the crystal structures of SARS-CoV-2 PLpro (PDB ID: 6WZU) (49), human Ub-specific protease (USP) 30 (PDB ID: 5OHN) (51), and FMDV PCP (PDB ID: 4QBB) (52) and for further analysis against that of RubPro (Fig. 3).

For all three cysteine proteases, there was structural similarity to RubPro around the catalytic sites at secondary elements, alpha helices α5–α8, and beta sheets β5, β8, and β6 (Fig. 3). The catalytic residues were in similar positions and orientations with less than 5 Å deviations (Fig. 3). While RubPro has a Cys-His catalytic dyad, the other proteases had catalytic triads; Cys-His-Asp for PLpro and FMDV PCP, and Cys-His-Ser for USP30. The N-terminal fingers domain of RubPro was distinct and had no structural homologs with any of the other three proteases.

RubPro cleaves RubProHel C1152A

RubProHel C1152A, with the natural cleavage junction between RUBV protease and helicase, was used to assess the transcleavage activity of the RubPro. A catalytic C1152A mutation was introduced to RubProHel to prevent any self-cleavage in *cis*. The RubProHel C1152A was incubated with doubling concentrations of RubPro. The *in vitro* transcleavage activity of RubPro was demonstrated by the decreased RubProHel C1152A, 64.9 kDa, and the formation of the helicase domain, 34.7 kDa (Fig. 4*A*).

RubProHel is cleaved during expression

Expression of RubProHel in *Escherichia coli* cells was used to ascertain cleavage of the p150–p90 junction. RubProHel is a construct with RUBV helicase at the C-terminal region of RubPro. RubProHel WT and a C1152A mutant, both with an N-terminal His-Sumo tag, were expressed and purified. From Figure 4*B*, the WT expression of His-Sumo-RubProHel produces the cleaved products, the N-terminal 42.8 kDa His-Sumo-RubPro and the 34.2 kDa helicase at the C-terminal region. The identity of the N-terminal cleavage product was validated by Sumo protease cleavage, which cleaved the 42.8 kDa His-Sumo-RubPro into 30.3 kDa RubPro. This shows that there is protease activity in RubProHel, whether in *cis* or *trans*, in *E. coli* expression. A mutant of the catalytic dyad, C1152A, showed no cleavage of RubProHel (Fig. S2).

Enzymatic activity of RubPro on two peptide substrates

Enzyme concentration and buffer pH optimization were first conducted to identify the suitable conditions for conducting protease assay. Enzyme concentration greater than 3.75 μM was chosen for easier quantification and calculation of the enzymatic rate (Fig. S3*A*). Buffer pH 6.4 to 6.6 displayed the highest enzymatic rate (Fig. S3*B*). Finally, the protease activity of RubPro was measured at a protein concentration of 5 μM at 50 mM potassium phosphate buffer (pH 6.6).

The protease assay of RubPro and RubPro C1152A mutants was performed using benzyloxycarbonyl-Arg-Leu-Arg-Gly-Gly-7-amino-4-methylcoumarin (Z-RLRGG-AMC) substrate and acryl-LSRGG-AMC substrate (GenScript). The RubPro has a higher affinity toward Z-RLRGG-AMC substrate than

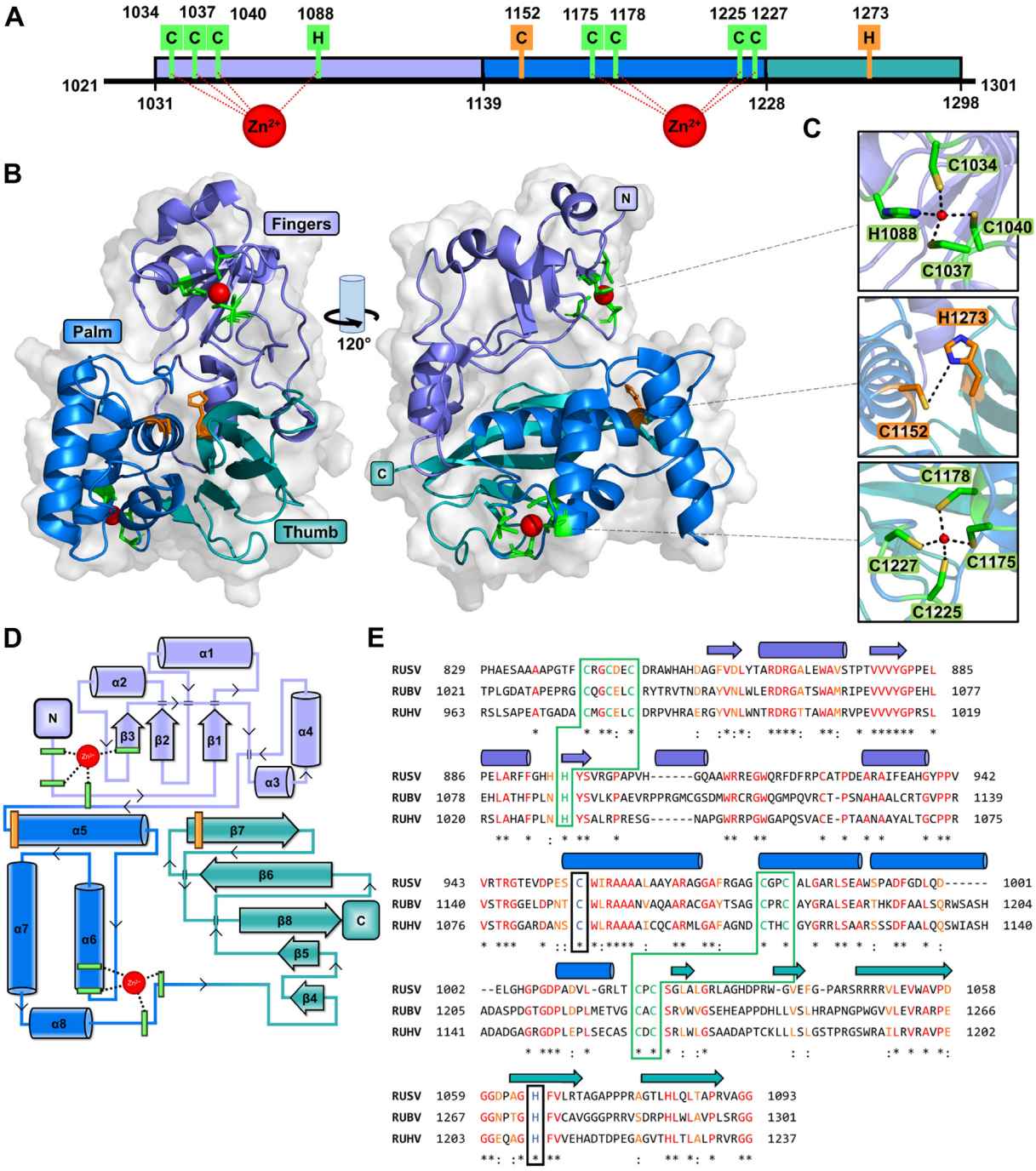

Figure 2. Crystal structure of RubPro. *A*, schematic representation of RubPro with residues of interest being highlighted. The RubPro is divided into three sections, "fingers" in *lavender*, "palm" in *marine blue*, and "thumb" in *teal*. The catalytic residues are colored in *orange*, Zn^{2+}-coordinating residues in *green*, and Zn^{2+} as *red spheres*. *B*, model of RubPro with a right-hand architecture, from the front-facing catalytic site. A view of RubPro with a 120° lateral rotation is also shown, with the N terminals and C terminals labeled. *C*, atomic view of Zn^{2+}-coordination sites and the catalytic site. The distances are shown as *black dashed lines* in Å. *D*, two-dimensional topology map of the RubPro model, showing α-helices as *cylinders* and β-sheets as *arrows*. α-helices and β-sheets are labeled and numbered starting from the N-terminal region. *E*, sequence alignment of RubPro with RUHV and RUSV proteases. A perfect alignment of residues is indicated in *red* with a (*), whereas alignment of residues with very strong similarities is indicated in *orange* with a (:). The catalytic dyad is boxed in *black*, and the Zn^{2+}-coordinating sites are boxed in *green*. RubPro, RUBV papain-like protease; RUBV, Rubella virus; RUHV, Ruhugu virus; RUSV, Rustrela virus

acryl-LSRGG-AMC substrate with K_m values of 580 ± 176 μM and 1490 ± 919 μM, respectively (Fig. 4, *C* and *D*). The RubPro also exhibits a catalytic rate of 7.09E-5 ± 0.919E-5 s^{-1} against Z-RLRGG-AMC substrate while having a 10-fold slower catalytic rate of 7.98E-6 ± 2.79E-6 s^{-1} against acryl-LSRGG-AMC substrate (Fig. 4, *C* and *D*).

Binding of the two substrate peptides to the RubPro active site

Two hexameric peptides, Ace-Leu-Ser-Arg-Gly-Gly-Gly-Nme (LSRGGG) and Ace-Arg-Leu-Arg-Gly-Gly-Gly-Nme (RLRGGG), were simulated together with the RUBV protease. Several steps that combined peptide docking, molecular dynamics (MD) simulations, and manually structure

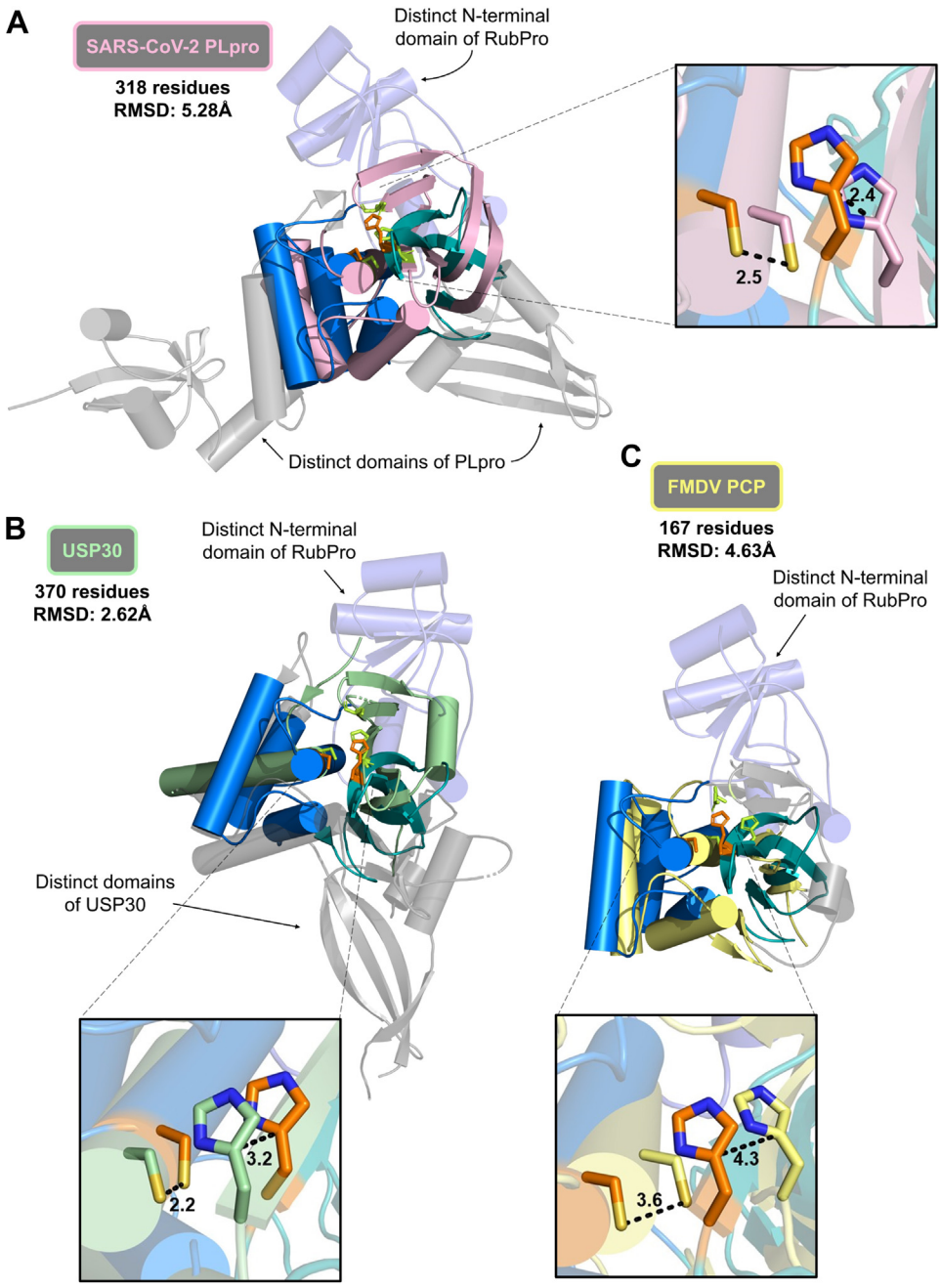

Figure 3. Structural alignment of RubPro with human USP30, FMDV PCP, and SARS-CoV-2 PLpro. RubPro follows the same coloring scheme as for Figure 2, with the fingers domain in *lavender*, palm domain in *marine blue*, and thumb domain in *teal*. The structurally similar core catalytic domains are shown in *solid color*. The distinct N-terminal fingers domain is shown as transparent as it does not have any alignment with the other protease. The distinct domains of the other proteases that are not aligned with RubPro are depicted in *transparent gray cartoons*. The α-helices are represented as *cylinders* and β-sheets as *arrows*. Close-up views of the catalytic sites are shown in the *boxes*, with the catalytic residues shown in *sticks*. The distance differences between the two structures are shown as *black dashed lines* in Å. *A*, residues 103 to 176 and 241 to 306 of SARS-CoV-2 PLpro, which were aligned with RubPro, depicted in *pink*. *B*, residues 58 to 97, 107 to 175, and 422 to 495 of USP30, which were aligned with RubPro, depicted in *pale green*. *C*, residues 43 to 126, 137 to 153, and 177 to 187 of FMDV PCP, which were aligned with RubPro, depicted in *light yellow*. FMDV, foot-and-mouth disease; PCP, papain-like cysteine protease; PLpro, papain-like protease; RubPro, RUBV papain-like protease; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; USP30, ubiquitin-specific protease.

manipulations, were needed to find the optimized poses of the peptides. During the final 100 ns MD simulations, the RMSD of the peptides after superimposing the protease showed stable binding (Fig. S4). The RMSD values of LSRGGG have larger fluctuations than those of RLRGGG, indicating a bit less stable binding of LSRGGG than RLRGGG. The binding modes of the two peptides are quite similar: the last three glycine residues sit in the narrow groove formed mainly by residues: G1213, W1153, and C1152C on one side and T1271, G1272, and H1273 on the other side. The common third residue, arginine,

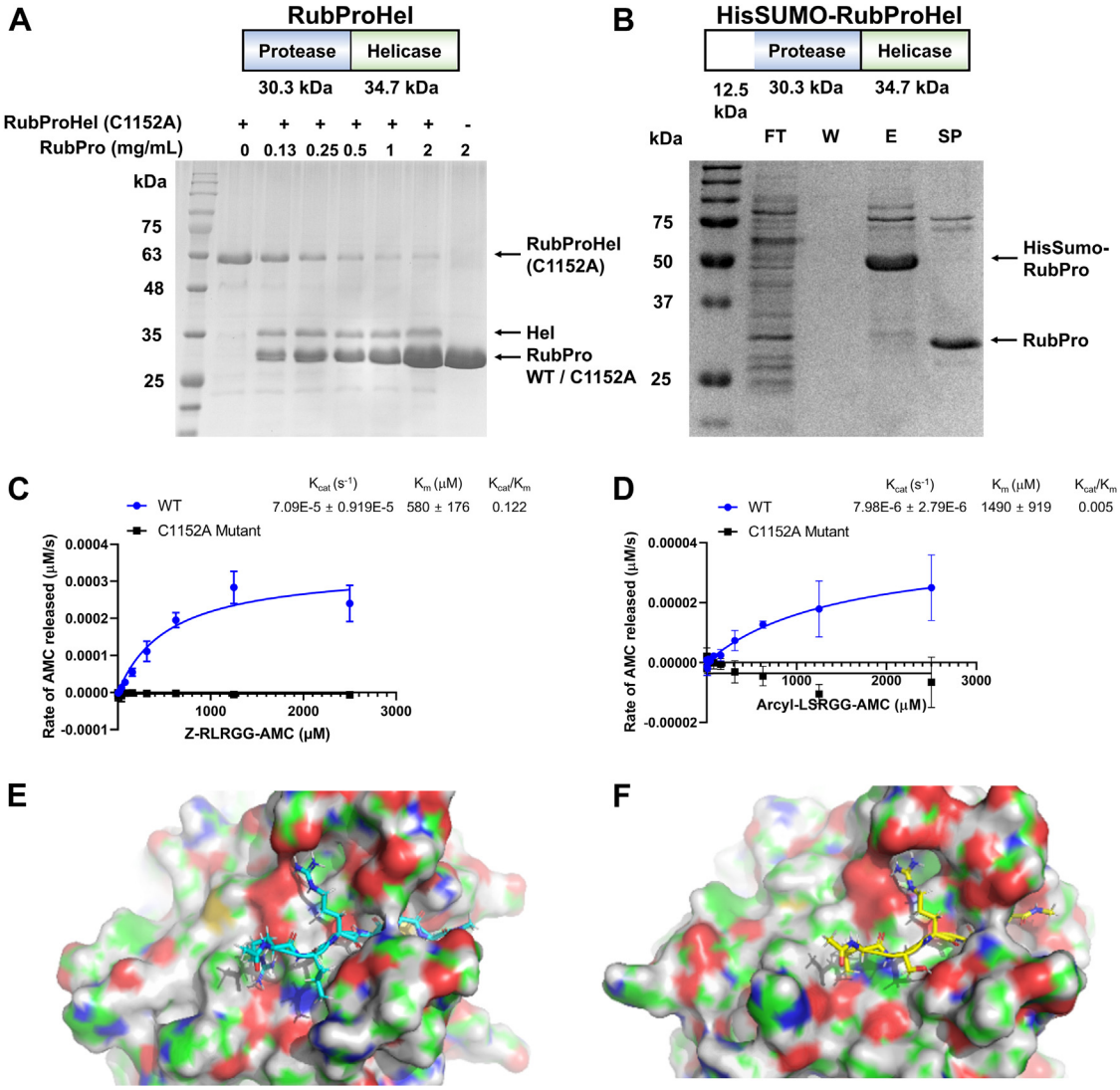

Figure 4. Enzymatic activity of RubPro. *A*, SDS-PAGE profile of *in vitro* transcleavage activity of RubPro using RubProHel C1152A as a substrate. About 0.5 mg/ml of RubProHel C1152A was incubated with 0.125 to 2 mg/ml of RubPro, at room temperature for 18 h in buffer (25 mM Hepes buffer at pH 7.5, 150 mM NaCl, 5% w/v glycerol, and 2 mM DTT). The molecular weight of the RubPro is 30.3 kDa and RubProHel is 64.9 kDa. While the molecular weight of the expected cleavage products helicase and RubPro is 34.7 and 30.3 kDa, respectively. *B*, expression of HisSUMO-RubProHel WT (77.5 kDa) in *Escherichia coli*. It is cleaved into HisSumo-RubPro (42.8 kDa, in the elution E) and helicase (34.7 kDa), which is not apparent in the flowthrough. RubPro (30.3 kDa) is obtained after Sumo protease cleavage (lane SP), validating the identity of the 42.8 kDa protein. The protease activity of RubPro was measured using the (*C*) Z-RLRGG-AMC and (*D*) acryl-LSRGG-AMC substrates. Assays were carried out as triplicates at 37 °C at 5 μM enzyme concentration with varying substrate concentrations ranging from 0 to 2.5 mM. *E*, the complex of the protein and the peptide RLRGGG. *F*, the complex of the protein and the peptide LSRGGG. LSRGGG, Leu-Ser-Arg-Gly-Gly-Gly-Nme; RLRGGG, Arg-Leu-Arg-Gly-Gly-Gly-Nme; RubPro, RUBV papain-like protease; Z-RLRGG-AMC, benzyloxycarbonyl-Arg-Leu-Arg-Gly-Gly-7-amino-4-methylcoumarin.

forms salt bridges with D1206 and D1217, showing the arginine residue is preferred here to anchor the peptide around the active side.

The side chain of serine residue of LSRGGG forms hydrogen bond with the side chain of T1271. The first residue leucine has hydrophobic contacts with L1216, L1247, and L1291, which form a hydrophobic patch. However, such interaction is not very stable, during the simulation, the side chain of this leucine was found to flip out and the methyl group of Ace makes close contacts with this hydrophobic patch.

The first residue, arginine, of RLRGGG, seems more fit in this position. Its side chain not only makes contacts with the hydrophobic patch (L1216, L1247, and L1291) on one side; on the other side stacks with the ring of H1290, also makes hydrogen bond with the carbonyl group of H1290 and a salt bridge with D1214. The second residue, leucine, however, seems no good interaction partners, loosely interacts with the methyl group of T1271.

The molecular mechanics generalized born (GB) surface area method was used to calculate the binding energy between the protein and the peptides. The results are consistent with the simulation RMSD values: the peptide RLRGGG has a binding energy of −49.06 ± 4.68 kcal/mol, and the peptide LSRGGG has a binding energy of −42.76 ± 7.13 kcal/mol. The slightly stronger binding of RLRGGG is also in line with the experimental measured K_m values.

RubPro is a K48-linkage–specific deubiquitinase

Given the structural similarities to numerous USPs, FMDV protease, and SARS-CoV-2 PLpro (Fig. 3 and Table S2), we seek to investigate if RubPro also demonstrates deubiquitylation and deISGylation activity. RubPro was tested on Ub and the Ub-like modifier ISG15. The K63-triubiquitin (K63-Ub_3), K48-triubiquitin (K48-UB_3), and M1-triubiquitin (linear) (M1-Ub_3), and ISG15-HIS were purified as per our previous work (53). Starting with Ub, robust RubPro activity and high specificity was observed toward Lys48-linked polyubiquitin (Fig. 5*B*), where triubiquitin (K48-Ub_3) was cleaved to mono-ubiquitin products (K48-Ub_1) but not for Lys63-linked triubiquitin (K63-Ub_3) (Fig. 5*A*) and linear-linked triubiquitin (M1-Ub_3) (Fig. 5*D*). DeISGylating activity was absent in RubPro where it is unable to remove the HIS tag from the C terminus of ISG15-HIS (Fig. 5*C*). A mutant of the catalytic dyad, C1152A, showed no cleavage of K63-triubiquitin, K48-triubiquitin, and M1-triubiquitin (linear) chains and ISG15-HIS (Fig. 5, *A*–*D*). Time-course analysis also suggest that RubPro is a K48-linkage–specific deubiquitinase, fully processed K48-Ub_3 within 5 min incubation at room temperature (Fig. 5*E*).

Discussion

The novel structure of RubPro presented here sheds much insight into the poorly understood RUBV (Fig. 2). The novelty of the structure is further exemplified by the unsuccessful prediction using the Alphafold2 server (Fig. S5). This 1.64 Å structure represents the first structure of the RUBV nonstructural protein. All 10 of the other PDB entries of RUBV proteins are the structural E1 glycoprotein and capsid proteins. While RubPro is similar to other PCPs, especially at the domains surrounding the catalytic dyad, it has a very distinct and unique N-terminal "fingers" domain. This domain could be important for coordinating the Zn^{2+} cation for structural support. It may also serve as a docking site for other cleavage targets or responsible for other functions as part of the p150

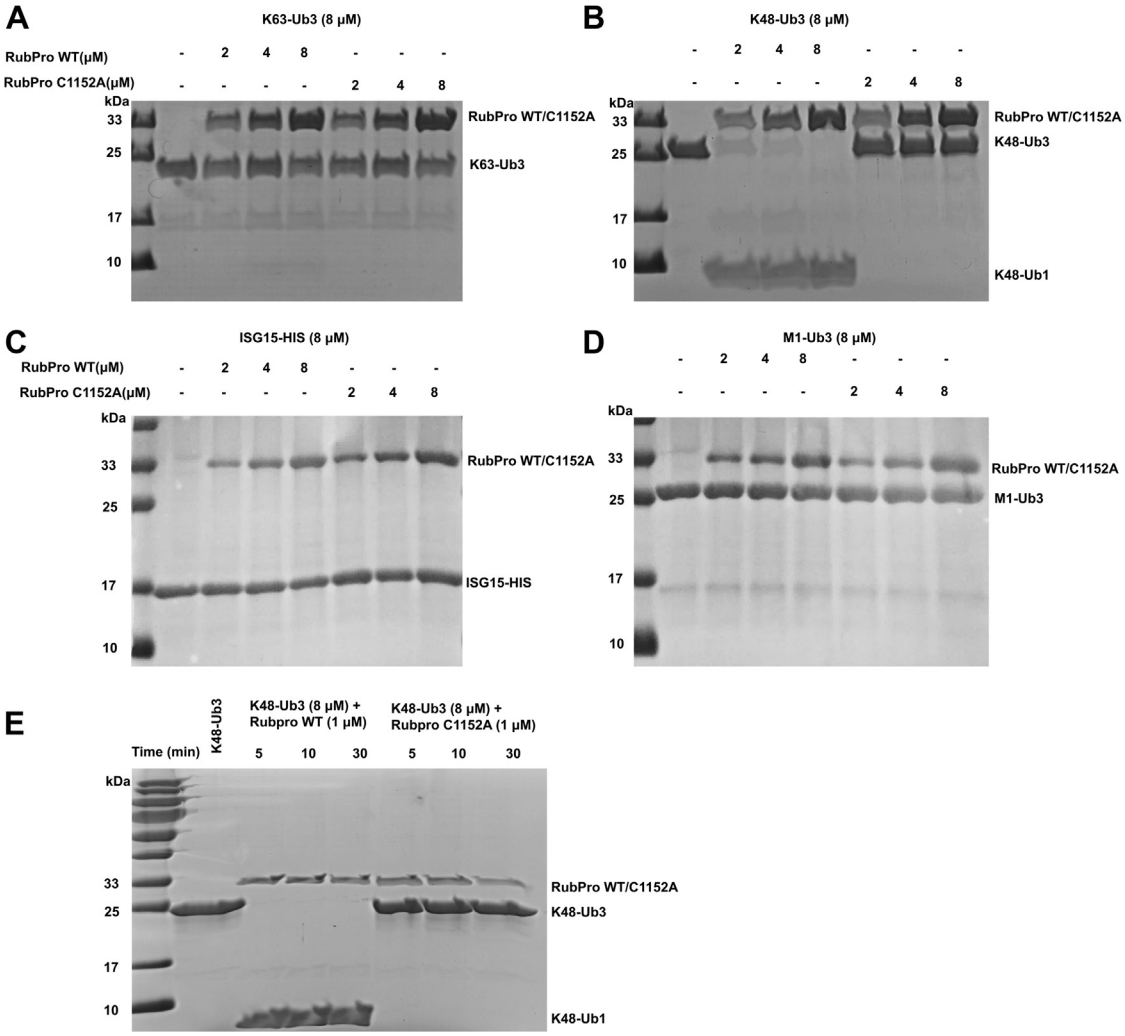

Figure 5. RubPro is a K48-ubiquitin–specific deubiquitinase. WT and C1152A mutants of RubPro were tested for their ability to process triubiquitin and ISG15 *in vitro*. Purified RubPros at indicated concentration were incubated at 25 °C with (*A*) K63-Ub3, (*B*) K48-Ub3, (*C*) ISG15-HIS, and (*D*) M1-Ub3 in a reaction buffer for 10 min. Cleavage products were analyzed by SDS-PAGE and stained with Coomassie blue. *E*, time-course analysis of triubiquitin (8 μM) hydrolysis using 1 μM RubPro in a reaction buffer, resolved on a Coomassie-stained SDS-PAGE gel. ISG15, interferon-stimulated gene 15; RubPro, RUBV papain-like protease.

protein. More work needs to be done to better understand and characterize this unique N-terminal domain and the multifunctional p150 protein.

Another unique observation is that the structure could represent the postcleavage configuration of the protein, with its' C-terminal ends in a β-sheet β8 that is pointing away from the catalytic site (Fig. 2*B*). A similar observation was also observed for Chikungunya virus capsid protease domain, which is only active for one proteolytic reaction, after which the active site is inaccessible (54). To gain a further understanding of the precleavage configuration of the protein, structural work may be carried out on RubProHel, to shed light on how the cleavage junction binds to the active site. Cocrystallization can also be carried out with a bound inhibitor or a substrate with both WT and catalytic mutant, C1152A or H1273A, to further characterize the active site.

Unlike what was reported in a previous study (34), we did identify the presence of neither the Ca^{2+}-binding sites nor a EF-hand domain in our crystal structure. We tried supplementing $CaCl_2$ into the crystallization buffer and soaking the crystals in a solution containing $CaCl_2$. One possible future work would be to conduct a denaturation-refolding assay of RubPro in the presence of Ca^{2+}.

The newly discovered *Rubivirus* members, RUHV and RUSV, are the closest relatives of RUBV. From sequence alignment of RubPro with RUHV and RUSV proteases (Fig. 2*E*), there are conserved motifs, especially around the catalytic dyads of C1152 and H1273. Unlike RUBV, RUHV and RUSV are pathogens in nonhuman mammals, capable of crossing host species barriers. It is not shown that humans can be infected with RUHV or RUSV, but there are some concerns about future zoonoses arising from RUBV-like viruses like RUHV and RUSV. Knowledge about RubPro may prove useful for other *Rubivirus* proteases given that conservation of the important motifs is observed.

Despite limited sequence similarity with other cysteine proteases, RubPro is structurally similar to the other cysteine proteases, namely, SARS-CoV-2 PLpro, human USP30, and FMDV PCP at the catalytic core (Fig. 3 and Table S2). The structure of RubPro was also compared with the protease domain of the Chikungunya virus, as RUBV was previously classified under the Togaviridae family. However, both structures are distinctively different (Fig. S6), and no members of the Togaviridae family were identified as hits in the Dali structural homology database search (Table S2).

Expression of the WT RubProHel showed significant cleavage into the 42.8 and 34.7 kDa products (Fig. 4*B*), demonstrating *in vivo* protease activity in *E. coli* cells. As such, the RubProHel C1152A mutant is likely to be a good candidate for structural studies, as it can bring structural insights into the self-cleavage mechanism of p200 into p150 and p90. The orientation of the cleavage junction SRGG/GTCA relative to the catalytic site can inform on the mechanism of RubPro activity on p200, whether it works in *cis* or *trans* dominantly. RubProHel C1152A was cleaved by RubPro in *trans* (Fig. 4*A*), and this differs from the conclusion by Liang *et al.* (20) that transcleavage requires additional residues 920 to 974. The additional residues might contribute to improved protease activity but are not essential for protease activity. In our study, RubPro at a high concentration of 2 mg/ml or 66 μM was unable to fully process all the RubProHel C1152A substrates, which is at a lower concentration of 0.5 mg/ml or 7.7 μM, at room temperature after 18 h. However, RubPro at 1 μM concentration can fully process K48-Ub_3 at concentration of 8 μM within 5 min incubation at room temperature (Fig. 5*E*). Similar results were observed for SARS-CoV-2 PLpro, SARS-CoV-1 PLpro whereby they cleave Ub much more efficiently than its own viral polyprotein substrate (55). Furthermore, the RubPro also displayed weak enzymatic activity having K_{cat} of 0.000071 s^{-1} and affinity for the benzyloxycarbonyl-Arg-Leu-Arg-Gly-Gly-4-methylcoumaryl-7-amide (Z-RLRGG-MCA) substrate with K_m of 580 μM. SARS-CoV-2 PLpro, SARS-CoV-1 PLpro, and Middle East respiratory syndrome coronavirus PLpro also have a weak enzymatic efficiency, approximate K_{cat}/K_m at 0.0051 μM min^{-1}, 0.3 μM min^{-1}, and 0.003 μM min^{-1}, respectively, toward RLRGG-AMC (55, 56). A RubPro construct containing the upstream residues from 920 onward could perhaps display more active transcleavage activity (20).

Given the structural similarities to numerous USPs, FMDV protease, and SARS-CoV-2 PLpro (Fig. 3 and Table S2), we seek to investigate if RubPro also demonstrates deubiquitylation activity. RubPro was shown to be a K48-linkage–specific deubiquitinase (Fig. 5). Preference toward K48-linkage deubiquitylation was observed for numerous viruses such as SARS-CoV-2 PLpro and SARS-CoV-1 PLpro, whereby the removal of K48-Ub from IκBα prevents the degradation of IκBα and the activation of both NF-κB signaling pathway and type I interferon response (55, 56). Future works seek to investigate the cellular targets of RubPro and the possible role of RubPro in modulating innate immune responses during RUBV infection.

Conclusion

Given the lack of fundamental understanding of RUBV replication processes, we set out to characterize the functional protease domain of the nonstructure p200 and p150 polypeptides, both structurally and functionally. RubPro might be a drug target candidate as an antiviral therapy for RUBV infections. Here, we managed to express and purify a pure RubPro construct and capture high-quality crystals for X-ray diffraction. This led to the structure of RubPro, which we solved at a resolution of 1.64 Å. This novel structure provides much insight into RubPro, but by the same token, it poses many new questions that are yet to be answered. We have validated the Zn^{2+}-binding sites atomically and structurally. We found a strong structural alignment of RubPro with SARS-CoV-2 PLpro and FMDV PCP at the core regions around the catalytic sites. Sequence alignment of RubPro with RUHV and RUSV proteases shows strong conservation of certain motifs. Overall, we have successfully characterized the structure of RubPro, but our construct has relatively low protease enzymatic activity. We were able to demonstrate cleavage of RubProHel in *E. coli*, and *RubPro* showed transprotease

activity against the natural substrate RubProHel C1152A mutant. Host factors that are targeted by similar viral PCPs might be deubiquitinated by RubPro as well. The work presented here may prove to be merely the cusp of deeper research into RUBV replication.

Experimental procedures

Designs of protein constructs

Cloning and expression tests of various proteins were carried out by the Protein Production Platform at the Nanyang Technological University. The rubella strain sequence used was National Center for Biotechnology Information reference sequence ID, NC_001545.2. RubPro, RubPro C1152A mutant, and RUBV protease–helicase (RubProHel) C1152A mutant were cloned in pSUMO-LIC vector, which encodes an N-terminal His-Sumo tag that is cleavable by Sumo protease. RubPro consisted of RUBV residues 1021 to 1301, and RubProHel consisted of RUBV residues 1021 to 1616. C1152A mutants were generated by site-directed mutagenesis of the respective WT constructs.

Mutagenesis of RubPro and RubProHel

C1152A mutagenesis was performed, for both RubPro and RubProHel using Q5 Site-Directed Mutagenesis Kit (New England BioLabs). The forward and reverse primers used were TCCCAACACTGCGTGGCTGAGAGCCGCCG and TCGA GTTCCCCGCCCCTT, respectively. Mutagenesis for both constructs was conducted as per the manufacturer's protocol.

Expression and purification of protein constructs

The protein expression plasmids were transformed into *E. coli* Rosetta 2 DE3 strain. The bacteria were cultured at 37 °C in LB broth supplemented with kanamycin and chloramphenicol. They were induced with 1 mM IPTG and cultured overnight at 18 °C. Bacteria cells were harvested by spinning in a Sorvall LYNC 4000 centrifuge (Thermo Fisher Scientific) at 6000 rpm for 15 min at 4 °C.

The pellet was resuspended in lysis buffer containing 1× PBS, supplemented with 160 mM NaCl, 5 mM β-mercaptoethanol, and 10% v/v glycerol. The sample was sonicated at 70% amplitude for 5 min at a 5 s on/5 s off interval using a Vibra-Cell sonicator (Sonics). The cells were further lysed using PandaPLUS 2000 homogenizer (GEA). The sample was centrifuged in a Sorvall LYNC 4000 centrifuge at 20,000 rpm for 1 h at 4 °C. The supernatant was collected and filtered with a 0.45 μm filter. The supernatant was subjected to 1.5 h of incubation with HisPur Ni-NTA Resin (Thermo Fisher Scientific) at 4 °C for immobilized metal affinity chromatography. Next, the resin was washed with lysis buffer supplemented with 20 mM imidazole and eluted with lysis buffer with 300 mM imidazole. Fractions were analyzed with SDS-PAGE. Chosen fractions were incubated with Sumo protease at 4 °C, in dialysis with lysis buffer as the dialysate. Cleaved proteins were collected in reverse immobilized metal affinity chromatography using nickel–nitrilotriacetic acid resin, where the flowthrough was collected and concentrated in concentrators (Merck Millipore) with appropriate molecular weight cutoff filter sizes. Samples then underwent size-exclusion chromatography in an ÄKTA Pure FPLC using HiLoad 16/600 Superdex 200 or Superdex 75 column (GE Healthcare Life Sciences) in GF-300 buffer (25 mM Hepes buffer at pH 7.5, 300 mM NaCl, 5% w/v glycerol, and 2 mM DTT). Fractions were analyzed with SDS-PAGE to determine purity. Pure fractions were then concentrated in concentrators of appropriate molecular weight cutoff filter sizes. The purified proteins were then aliquoted, flash-frozen with liquid nitrogen, and stored at −80 °C.

Protein crystallization

One microliter of RubPro at a final concentration of 16 mg/ml in a GF-300 buffer supplement with 2 mM Tris(2-carboxyethyl) phosphine was mixed with 1 μl of reservoir solution containing 0.1 M Hepes (pH 7.5) and 15% w/v polyethylene glycol 3350, the mixture was subjected to crystallization *via* hanging drop vapor diffusion method at 20 °C. Crystals appeared after 2 days of incubation. The crystals were mounted onto a cryoloop and cryoprotected with reservoir solution supplemented with 20% v/v glycerol and flash-frozen in liquid nitrogen.

Data collection, structure determination, and model building

Diffraction intensities of the native crystals were recorded on Dectris EIGER 16M detector at MX2 beamline at wavelength 1.284 Å at the Australian Synchrotron. Data processing was carried out using the XDS data processing package (Max Planck Institute for Medical Research).

The structure of the RubPro was determined using the single-wavelength anomalous diffraction method. Measurements of anomalous signals from Zn^{2+} were used to derive the positions of the heavy atom substructure. Crank2 (57) and BUCCANEER (58) software (Collaborative Computational Project No. 4) were used to calculate the initial phases and for automatic model building, respectively. The structure was subjected to iterative rounds of refinement using the Phenix_refine program (59–61) and manual rebuilding using Coot (62) (Medical Research Council Laboratory of Molecular Biology). Figures were generated using PyMOL (Schrödinger). Data collection and refinement statistics can be found in Table S1.

Sequence alignment with RUHV and RUSV proteases

The amino acid residue sequence of our RubPro construct was aligned with the sequences of the nonstructural p200 polypeptides of RUHV and RUSV. The sequences were obtained from UniProtKB with the identifiers A0A7L5KV54 for RUHV and A0A7L5KV68 for RUSV. Alignments were done using Clustal Omega (63) [European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI)].

Testing of RubProHel C1152A cleavage by RubPro

About 0.5 mg/ml of RubProHel C1152A was incubated with doubling concentrations of RubPro, from 0.125 to 2 mg/ml, at room temperature for 18 h in cleavage buffer (25 mM Hepes buffer at pH 7.5, 150 mM NaCl, 5% w/v glycerol, and 2 mM DTT). Samples were analyzed by electrophoresis on a 15% SDS-PAGE gel and visualized by Coomassie blue staining.

Enzyme concentration optimization for protease assay

The assay was performed in a 96-well half-area black clear-bottom microplate (Corning). The reaction consists of varying concentrations of RubPro from 0.9 to 15 μM in assay buffer containing 25 mM Hepes (pH 6.8) and 150 mM NaCl. The measurements were started by adding the Z-RLRGG-MCA peptide substrate (Peptide Institute, Inc) at 50 μM. The relative fluorescence readings were measured using Synergy H1 Microplate Reader (BioTek) at 3-min intervals over 1 h with the excitation wavelength (λ_{ex}) at 380 nm and emission wavelength (λ_{em}) at 460 nm. The assays were conducted as triplicates at 37 °C.

Buffer pH optimization for protease assay

The assay was performed in a 96-well half-area black clear-bottom microplate. The reaction consists of 5 μM of RubPro in assay buffer containing 50 mM potassium phosphate buffer with varying pH from 5.8 to 7.8. The measurements were started by adding the Z-RLRGG-MCA peptide substrate at 50 μM. The relative fluorescence readings were measured using Synergy H1 Microplate Reader at 3-min intervals over 1 h with λ_{ex} at 380 nm and λ_{em} at 460 nm. The assays were conducted as triplicates at 37 °C.

Protease assay

Z-RLRGG-MCA substrate with starting concentration of 2.5 mM was serially diluted two times in assay buffer (50 mM potassium phosphate buffer [pH 6.6]) and added to Corning 96-well black plates containing 5 μM RubPro or RubPro C1152A mutant diluted in the same buffer. The relative fluorescence readings were measured using Synergy H1 Microplate Reader at 3-min intervals over 1 h with λ_{ex} at 380 nm and λ_{em} at 460 nm. The assays were carried out as triplicates at 37 °C. To determine the amount of AMC released, a standard AMC curve was plotted with various concentrations of AMC. Initial velocities were calculated using the linear regression function in the GraphPad Prism software (GraphPad Software, Inc). Data were analyzed and plotted using the Michaelis–Menten equation with GraphPad Prism 8 for Windows. The calculated K_{cat} and K_m values are rounded to three significant figures. The assay was repeated using another chemically synthesized peptide substrate, acryl-LSRGG-AMC.

Structural prediction of RubPro using Alphafold2 web server

Structural prediction of RubPro was performed using the Alphafold2 Colab web server (64).

Docking and MD simulations

Isolated peptides, LSRGGG and Ace-Arg-Leu-Arg-Gly-Gly-Gly-Nme (RLRGGG), were simulated in a water box with size of 3.82 nm × 3.82 nm × 3.82 nm for 500 ns. From the second 250 ns trajectories, one structure of every 100 ps was chosen (in total, 2500 structures for each peptide) to dock onto the protein structure near the active side (C1152 and H1273). The quick autodock vina tool (65) (The Scripps Research Institute) was used for docking by freezing all the peptide conformations. These docked structures were ranked based on the distance between the amide nitrogen atom of the third glycine and the center of C1152 and H1273, the catalytic dyad. Five poses of each peptide were chosen with the distance (aforementioned) less than 0.5 nm. MD simulations were performed in the explicit water box for the chosen protein–peptide complexes. In most of the simulations, the peptide moves away from the active site. Only two systems of LSRGGG show stable binding of the peptide. The last two glycine residues were found forming hydrogen bonds with the residue near the catalytic dyad; however, the distance between the amide nitrogen atom of the third glycine and the center of C1152 and H1273 is still far, around 0.6 nm. Here, we made a structure manipulation: based on the binding pose found in the simulation, we shifted the peptide backbone toward the active side by two residues by superimposing the first glycine (G4) to the last glycine (G6). The new complex structure was used to perform MD simulation in the explicit water box to equilibrate. In this way, the protein–LSRGGG complex structure was obtained. The initial protein–RLRGGG structure was got based on the protein–LSRGGG structure and mutating LSRGGG to RLRGGG.

Parameters of the proteins and peptides were based on the AMBER99SB-ILDN force field (66), and the system was solvated with TIP3P (67) water molecules, and counterions were added to neutralize the system. The MD simulations were performed using GROMACS (68) 5.1.2 software (University of Groningen). The LINCS (69) algorithm was used to constrain bonds between heavy atoms and hydrogen to enable a timestep of 2 fs. A 1.2 nm cutoff was used for van der Waals interaction and short-range electrostatic interaction calculations, and the particle mesh Ewald method was implemented for long-range electrostatic calculations. Simulation temperature was maintained at 300 K using a V-rescale thermostat (70) and 1 bar pressure using Parrinello–Rahman (71) barostat.

Binding energy calculations

To analyze the behavior of interactions between the proteins and the peptides, we calculated the binding energies between them using the molecular mechanics GB surface area method (72). The entropy was not calculated:

$$\Delta G_{\text{Bind}} = G_{\text{Complex}} - \left(G_{\text{pro}} + G_{\text{pep}}\right) \quad (1)$$

where, G_{Complex}, G_{pro}, and G_{pep} is the free energy of complex, the protein, and the peptide, respectively. Free energy (ΔG) of each state was calculated as follows:

$$G = E_{\text{MM}} + G_{\text{GB}} + G_{\text{SA}} - \text{TS} \quad (2)$$

$$E_{\text{MM}} = E_{\text{vdW}} + E_{\text{ele}} + E_{\text{int}} \quad (3)$$

where E_{MM} is the molecular mechanical energy and G_{GB} is the polar contribution toward solvation energy calculated by GB

method, respectively. G_{SA} is the contribution from the nonpolar terms toward solvation energy, and TS is the entropic contribution of the system. E_{MM} was obtained by summing contributions from the electrostatic energy (E_{ele}), the van der Waal energy (E_{vdw}), and the internal energy including bond, angle, and torsional angle energy terms (E_{int}) using the same force field as that of MD simulations. G_{GB} was calculated with Onufriev's method (73). G_{SA} in Equation 2 is proportional to the solvent-accessible surface area and was computed by molsurf module:

$$G_{SA} = \gamma^* SASA + b \quad (4)$$

where, the surface tension proportionality constant (γ) was set to 0.0072 kcal/mol/Å^2, whereas the free energy of nonpolar solvation for a point solute (b) was set to a default value.

In vitro *assay for deubiquitylating and delSGylating*

The protein substrate ubiquitination was prepared as previously described (53). Purified RubPro WT protein or mutant (C1152A) protein was incubated with 8 μM of the substrate (K63-triubiquitin, K48-triubiquitin, and M1-triubiquitin [linear] chains and ISG15-HIS) at 25 °C for 5 to 30 min in 10 μl of reaction buffer containing 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, and 5% glycerol. The reaction was quenched by the addition of SDS sample buffer, analyzed by electrophoresis on a 15% SDS-PAGE gel, and visualized by Coomassie blue staining.

Data availability

Atomic coordinates and associated structure factors have been deposited in the PDB (PDB ID: 7FAV).

Supporting information—This article contains supporting information.

Acknowledgments—This research is supported by the Ministry of Education, Singapore, under its MOE AcRF tier 1 Award 2021-T1-002-021. We acknowledge the funding support for this project from Nanyang Technological University under the URECA Undergraduate Research Program. We gratefully acknowledge the beamline staff at MXII beamline in Australian Light Source, Melbourne, Australia for providing us with outstanding support during the data collection.

Author contributions—E. Z. K. C., J. P. Q., L. X., C. L., J. Y. C., C. W. L., Y. M., J. Z., and D. L. formal analysis; E. Z. K. C., J. P. Q., L. X., C. L., J. Y. C., C. W. L., Y. M., J. Z., and D. L. data curation; E. Z. K. C., Y. M., J. Z., and D. L. writing–review & editing; D. L. supervision.

Funding and additional information—J. P. Q. is supported by the Nanyang Presidential Graduate Scholarship and the Lee Kong Chian School of Medicine.

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

Abbreviations—The abbreviations used are: CRS, congenital rubella syndrome; FMDV, foot-and-mouth disease; GB, generalized born; ISG15, interferon-stimulated gene 15; LSRGGG, Leu-Ser-Arg-Gly-Gly-Gly-Nme; MD, molecular dynamics; PCP, papain-like cysteine protease; PDB, Protein Data Bank; PLpro, papain-like protease; RLRGGG, Arg-Leu-Arg-Gly-Gly-Gly-Nme; RUBV, Rubella virus; RubPro, RUBV papain-like protease; RUHV, Ruhugu virus; RUSV, Rustrela virus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; Ub, ubiquitin; UB_3, triubiquitin; USP, ubiquitin-specific protease; Z-RLRGG-AMC, benzyloxycarbonyl-Arg-Leu-Arg-Gly-Gly-7-amino-4-methylcoumarin; Z-RLRGG-MCA, benzyloxycarbonyl-Arg-Leu-Arg-Gly-Gly-4-methylcoumaryl-7-amide.

References

1. Atkinson, W. (2006) *Epidemiology and Prevention of Vaccine-Preventable Diseases*, Department of Health & Human Services, Centers for Disease Control and Prevention, Atlanta, GA
2. Frey, T. K. (1994) Molecular biology of rubella virus. *Adv. Virus. Res.* **44**, 69–160
3. Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Alfenas-Zerbini, P., *et al.* (2021) Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2021). *Arch. Virol.* **166**, 2633–2648
4. Lambert, N., Strebel, P., Orenstein, W., Icenogle, J., and Poland, G. A. (2015) Rubella. *Lancet* **385**, 2297–2307
5. Givens, K. T., Lee, D. A., Jones, T., and Ilstrup, D. M. (1993) Congenital rubella syndrome: ophthalmic manifestations and associated systemic disorders. *Br. J. Ophthalmol.* **77**, 358–363
6. Shukla, S., and Maraqa, N. F. (2020) *Congenital Rubella*, StatPearls, Treasure Island (FL)
7. Lanzieri, T., Redd, S., Abernathy, E., and Icenogle, J. (2017) Congenital rubella syndrome. In *Manual for the Surveillance of Vaccine-Preventable Diseases*. Centers for Disease Control and Prevention, Atlanta, GA
8. World Health Organization. (2014) *Observed Rate of Vaccine Reactions - MMR Vaccines*, World Health Organization, Geneva
9. Wadman, M. (2017) *The Vaccine Race: Science, Politics, and the Human Costs of Defeating Disease*, Random House, NY
10. Centers for Disease Control and Prevention. (2020) *Rubella Vaccination*, Centers for Disease Control and Prevention, Atlanta, GA
11. World Health Organization. (2009) *WHO Position on Measles Vaccines*, World Health Organization, Geneva
12. Paredes, A. M., Brown, D. T., Rothnagel, R., Chiu, W., Schoepp, R. J., Johnston, R. E., *et al.* (1993) Three-dimensional structure of a membrane-containing virus. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9095–9099
13. Zhou, Y., Chen, X., Ushijima, H., and Frey, T. K. (2012) Analysis of base and codon usage by rubella virus. *Arch. Virol.* **157**, 889–899
14. Kujala, P., Ahola, T., Ehsani, N., Auvinen, P., Vihinen, H., and Kaariainen, L. (1999) Intracellular distribution of rubella virus nonstructural protein P150. *J. Virol.* **73**, 7805–7811
15. Dominguez, G., Wang, C. Y., and Frey, T. K. (1990) Sequence of the genome RNA of rubella virus: Evidence for genetic rearrangement during togavirus evolution. *Virology* **177**, 225–238
16. Marr, L. D., Wang, C. Y., and Frey, T. K. (1994) Expression of the rubella virus nonstructural protein ORF and demonstration of proteolytic processing. *Virology* **198**, 586–592
17. Atreya, C., Mohan, K., and Kulkarni, S. (2004) Rubella virus and birth defects: molecular insights into the viral teratogenesis at the cellular level. *Birth Defects Res. A: Clin. Mol. Teratol.* **70**, 431–437
18. Lee, J. Y., and Bowden, D. S. (2000) Rubella virus replication and links to teratogenicity. *Clin. Microbiol. Rev.* **13**, 571–587
19. Krausslich, H. G., and Wimmer, E. (1988) Viral proteinases. *Annu. Rev. Biochem.* **57**, 701–754
20. Liang, Y., Yao, J., and Gillam, S. (2000) Rubella virus nonstructural protein protease domains involved in trans- and cis-cleavage activities. *J. Virol.* **74**, 5412–5423
21. Bowden, D. S., and Westaway, E. G. (1984) Rubella virus: structural and non-structural proteins. *J. Gen. Virol.* **65**, 933–943

22. Bromme, D. (2001) Papain-like cysteine proteases. *Curr. Protoc. Protein Sci.* https://doi.org/10.1002/0471140864.ps2102s21
23. Fehr, A. R., and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses* **1282**, 1–23
24. Byler, K. G., Collins, J. T., Ogungbe, I. V., and Setzer, W. N. (2016) Alphavirus protease inhibitors from natural sources: a homology modeling and molecular docking investigation. *Comput. Biol. Chem.* **64**, 163–184
25. Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., *et al.* (2020) Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662
26. Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., *et al.* (2006) Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5717–5722
27. Wang, D., Fang, L., Li, P., Sun, L., Fan, J., Zhang, Q., *et al.* (2011) The leader proteinase of foot-and-mouth disease virus negatively regulates the type I interferon pathway by acting as a viral deubiquitinase. *J. Virol.* **85**, 3758–3766
28. Li, J., Chai, Q. Y., and Liu, C. H. (2016) The ubiquitin system: a critical regulator of innate immunity and pathogen-host interactions. *Cell Mol Immunol* **13**, 560–576
29. Ghosh, A. K., Brindisi, M., Shahabi, D., Chapman, M. E., and Mesecar, A. D. (2020) Drug development and medicinal chemistry efforts toward SARS-coronavirus and Covid-19 therapeutics. *ChemMedChem* **15**, 907–932
30. Liu, X., Yang, J., Ghazi, A. M., and Frey, T. K. (2000) Characterization of the zinc binding activity of the rubella virus nonstructural protease. *J. Virol.* **74**, 5949–5956
31. Chen, J. P., Strauss, J. H., Strauss, E. G., and Frey, T. K. (1996) Characterization of the rubella virus nonstructural protease domain and its cleavage site. *J. Virol.* **70**, 4707–4713
32. Liu, X., Ropp, S. L., Jackson, R. J., and Frey, T. K. (1998) The rubella virus nonstructural protease requires divalent cations for activity and functions in trans. *J. Virol.* **72**, 4463–4466
33. Zhou, Y., Tzeng, W. P., Wong, H. C., Ye, Y., Jiang, J., Chen, Y., *et al.* (2010) Calcium-dependent association of calmodulin with the rubella virus nonstructural protease domain. *J. Biol. Chem.* **285**, 8855–8868
34. Zhou, Y., Tzeng, W. P., Yang, W., Zhou, Y., Ye, Y., Lee, H. W., *et al.* (2007) Identification of a Ca2+-binding domain in the rubella virus nonstructural protease. *J. Virol.* **81**, 7517–7528
35. Zhou, Y., Tzeng, W. P., Ye, Y., Huang, Y., Li, S., Chen, Y., *et al.* (2009) A cysteine-rich metal-binding domain from rubella virus non-structural protein is essential for viral protease activity and virus replication. *Biochem. J.* **417**, 477–483
36. DuBois, R. M., Vaney, M. C., Tortorici, M. A., Kurdi, R. A., Barba-Spaeth, G., Krey, T., *et al.* (2013) Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* **493**, 552–556
37. Prasad, V. M., Klose, T., and Rossmann, M. G. (2017) Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. *PLoS Pathog.* **13**, e1006377
38. Frey, T. K. (1997) Neurological aspects of rubella virus infection. *Intervirology* **40**, 167–175
39. Bennett, A. J., Paskey, A. C., Ebinger, A., Pfaff, F., Priemer, G., Hoper, D., *et al.* (2020) Relatives of rubella virus in diverse mammals. *Nature* **586**, 424–428
40. Ujiie, M., Nabae, K., and Shobayashi, T. (2014) Rubella outbreak in Japan. *Lancet* **383**, 1460–1461
41. Sugishita, Y., Shimatani, N., Katow, S., Takahashi, T., and Hori, N. (2015) Epidemiological characteristics of rubella and congenital rubella syndrome in the 2012-2013 epidemics in Tokyo, Japan. *Jpn. J. Infect. Dis.* **68**, 159–165
42. Paradowska-Stankiewicz, I., Czarkowski, M. P., Derrough, T., and Stefanoff, P. (2013) Ongoing outbreak of rubella among young male adults in Poland: increased risk of congenital rubella infections. *Euro Surveill.* **18**, 20485
43. Janta, D., Stanescu, A., Lupulescu, E., Molnar, G., and Pistol, A. (2012) Ongoing rubella outbreak among adolescents in Salaj, Romania, September 2011-January 2012. *Euro Surveill.* **17**, 20089
44. Chary, A., and Holodniy, M. (2010) Recent advances in hepatitis C virus treatment: review of HCV protease inhibitor clinical trials. *Rev. Recent Clin. Trials* **5**, 158–173
45. Lv, Z., Chu, Y., and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* **7**, 95–104
46. Pardo, J., Shukla, A. M., Chamarthi, G., and Gupte, A. (2020) The journey of remdesivir: from ebola to COVID-19. *Drugs Context* **9**
47. Udwadia, Z. F., Singh, P., Barkate, H., Patil, S., Rangwala, S., Pendse, A., *et al.* (2021) Efficacy and safety of favipiravir, an oral RNA-dependent RNA polymerase inhibitor, in mild-to-moderate COVID-19: a randomized, comparative, open-label, multicenter, phase 3 clinical trial. *Int. J. Infect. Dis.* **103**, 62–71
48. Koff, R. S. (2014) Review article: the efficacy and safety of sofosbuvir, a novel, oral nucleotide NS5B polymerase inhibitor, in the treatment of chronic hepatitis C virus infection. *Aliment. Pharmacol. Ther.* **39**, 478–487
49. Osipiuk, J., Azizi, S. A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K. A., *et al.* (2021) Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **12**, 743
50. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
51. Gersch, M., Gladkova, C., Schubert, A. F., Michel, M. A., Maslen, S., and Komander, D. (2017) Mechanism and regulation of the Lys6-selective deubiquitinase USP30. *Nat. Struct. Mol. Biol.* **24**, 920–930
52. Steinberger, J., Grishkovskaya, I., Cencic, R., Juliano, L., Juliano, M. A., and Skern, T. (2014) Foot-and-mouth disease virus leader proteinase: structural insights into the mechanism of intermolecular cleavage. *Virology* **468**, 397–408
53. Song, B., Chen, Y., Liu, X., Yuan, F., Tan, E. Y. J., Lei, Y., *et al.* (2021) Ordered assembly of the cytosolic RNA-sensing MDA5-MAVS signaling complex *via* binding to unanchored K63-linked poly-ubiquitin chains. *Immunity* **54**, 2218–2230.e5
54. Aggarwal, M., Sharma, R., Kumar, P., Parida, M., and Tomar, S. (2015) Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. *Sci. Rep.* **5**, 14753
55. Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., *et al.* (2020) Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infect. Dis.* **6**, 2099–2109
56. Baez-Santos, Y. M., Mielech, A. M., Deng, X., Baker, S., and Mesecar, A. D. (2014) Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J. Virol.* **88**, 12511–12527
57. Skubak, P., and Pannu, N. S. (2013) Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777
58. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011
59. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367
60. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., *et al.* (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
61. Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., *et al.* (2012) Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 381–390
62. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
63. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucl. Acids Res.* **47**, W636–W641
64. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682

65. Alhossary, A., Handoko, S. D., Mu, Y., and Kwoh, C.-K. (2015) Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **31**, 2214–2216
66. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., *et al*. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958
67. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
68. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) Gromacs: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718
69. Hess, B. (2008) P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Comput.* **4**, 116–122
70. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101
71. Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190
72. Genheden, S., and Ryde, U. (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461
73. Onufriev, A., Bashford, D., and Case, D. A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: struct. Funct. Bioinformatics* **55**, 383–394