Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin

Jerzy Osipiuk[1,2,*], Pawel M. Wydorski[3,4,*], Benjamin T. Lanham[5,*], Christine Tesar[1,2], Michael Endres[1,2], Elizabeth Engle[5], Robert Jedrzejczak[1,2], Vishruth Mullapudi[4], Karolina Michalska[1,2], Krzysztof Fidelis[6], David Fushman[5,#], Andrzej Joachimiak[1,2,7,#] and Lukasz A. Joachimiak[4,8#]

[1]*Center for Structural Genomics of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL, 60667 USA*

[2]*Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Lemont, IL, 60439 USA*

[3]*Molecular Biophysics Graduate Program, University of Texas Southwestern Medical Center, Dallas, TX 75390 USA*

[4]*Center for Alzheimer's and Neurodegenerative Diseases, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390 USA*

[5]*Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742 USA*

[6]*Protein Structure Prediction Center, Genome and Biomedical Sciences Facilities, University of California, Davis, CA, 95616 USA.*

[7]*Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60367 USA*

[8]*Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390 USA*

[*]These authors contributed equally

[#]Correspondence should be addressed to: David Fushman (fushman@umd.edu), Andrzej Joachimiak (andrzejj@anl.gov) and Lukasz A. Joachimiak (Lukasz.Joachimiak@utsouthwestern.edu)

Running Title: PLpro complexes

Keywords: PLpro, ubiquitin, ISG15, SARS-CoV-2, crystal structure, COVID-19

ABSTRACT

The Papain-like protease (PLpro) is a domain of a multi-functional, non-structural protein 3 of coronaviruses. PLpro cleaves viral polyproteins and posttranslational conjugates with poly-ubiquitin and protective ISG15, composed of two ubiquitin-like (UBL) domains. Across coronaviruses, PLpro showed divergent selectivity for recognition and cleavage of posttranslational conjugates despite sequence conservation. We show that SARS-CoV-2 PLpro binds human ISG15 and K48-linked di-ubiquitin (K48-Ub_2) with nanomolar affinity and detect alternate weaker-binding modes. Crystal structures of untethered PLpro complexes with ISG15 and K48-Ub_2 combined with solution NMR and cross-linking mass spectrometry revealed how the two domains of ISG15 or K48-Ub_2 are differently utilized in interactions with PLpro. Analysis of protein interface energetics uncovered differential binding stabilities of the two UBL/Ub domains. We emphasize how substrate recognition can be tuned to cleave specifically ISG15 or K48-Ub_2 modifications while retaining capacity to cleave mono-Ub conjugates. These results highlight alternative druggable surfaces that would inhibit PLpro function.

INTRODUCTION

The COVID-19 pandemic is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) belonging to the *Coronaviridae* family. SARS-CoV-2 is a spherical, enveloped, non-segmented, (+) sense RNA virion with a ~30 kbs genome[1,2]. The genome is used for translation of the replication-transcription complex (RTC) comprised of two polyproteins (Pp1a/Pp1ab), which are processed by two viral proteases: papain-like protease (PLpro) and 3C-like protease. PLpro is a domain of non-structural protein 3 (Nsp3). It cleaves three sites in SARS-CoV-2 polyproteins yielding Nsp1, Nsp2 and Nsp3. The "LXGG↓XX" motif found in the polyproteins is essential for protease recognition and cleavage. PLpro has been shown to have additional functions[3], including deubiquitinating[4-6] and deISGylating activities[7,8], occurring by the cleavage of the "LRGG" sequence motif found at the C-terminus of human ubiquitin (Ub) and ISG15. PLpro from SARS-CoV-1 ($PLpro^{CoV-1}$) preferentially removes K48-polyUb modification[9]. Removing these post-translational modifications (PTM) interferes with interferon (IFN) expression and blocking NF-kappaB signaling[10]. By cleaving off ISG15 from STAT3 it induces up-regulation of TGF-β1[11]. Some PLpro functions involve direct cleavage of host proteins influencing broad processes from blood coagulation to nuclear transport[12-14]. PLpro may also play roles beyond its proteolytic activity[6], illustrating its diverse and complex functions[3]. The SARS-CoV-2 PLpro ($PLpro^{CoV-2}$) sequence and structural fold are conserved among SARS-CoV-1 (83% identical), MERS-CoV (30% identical) and other coronaviruses. Despite low sequence identities (~10%)[7], PLpros also share common structural architecture and catalytic site with the human ubiquitin specific proteases (USPs), one of the five distinct deubiquitinating enzyme (DUB) families.

PLpro from SARS- and MERS-CoVs have deubiquitinating[4-6] and deISGylating activities[7,8], but they show somewhat different substrate selectivity, with $PLpro^{CoV-1}$ preferentially removing K48-polyUb modification[9] and MERS-CoV PLpro being the least specific. It is not fully understood how this activity influences interaction of the virus with the host immune system. Ubiquitination is an essential PTM engaged in multiple functions in humans, including signaling proteasome-dependent protein degradation. The modification is mediated by the Ub-conjugating system and could be reversed by DUBs[8]. ISG15 is an IFNα-stimulated gene that is a critical component of the antiviral response[15,16]. It is not precisely clear how ISG15 interferes with viral processes but it is believed that tagging newly translated viral proteins with ISG15 sterically prevents their folding, assembly or

interactions[17]. Removal of Ub and ISG15 conjugates from specific substrates in host cells may also have a diverse impact on numerous cellular processes and specifically may frustrate the host response to viral infection[15,16,18]. ISG15 is processed and subsequently activated in a manner similar to Ub using interferon-induced factors that follow the ubiquitination-like E1, E2 and E3 enzyme cascade to mediate co-translational ISGylation – an addition of ISG15, via its C-terminal LRGG motif, to substrate lysine residues[19]. PLpro can recognize and cleave off both appendages as they share PLpro recognition motif at their C-termini and have common other structural features: ISG15 comprises two Ub-like (UBL) domains and mimics a head-to-tail linked Ub_2. While K48-linked Ub_2 (K48-Ub_2) and ISG15 are homologous both in sequence and fold, the topologies of how the two domains are linked are distinct. It remains unknown how PLpro discriminates between the Ub_2 and ISG15 substrates.

Recently published work suggested that mutations in $PLpro^{CoV-1}$ to $PLpro^{CoV-2}$ changed its binding preference from K48-linked Ub_2 to human ISG15 (hISG15)[17,20]. Inspired and attracted by these results we investigated interaction of $PLpro^{CoV-2}$ with mono-, di-, and tri-ubiquitins (Ub_1, Ub_2, Ub_3) and hISG15. We employed complementary biochemical, structural x-ray, NMR and computational approaches, to understand how $PLpro^{CoV-2}$ can differentiate between Ub_1, K48-Ub_2 and hISG15 substrates. We find that $PLpro^{CoV-2}$ binds both hISG15 and K48-Ub_2 with high and similar affinity but shows weaker interactions with Ub_1. We also find lower affinity alternate binding modes for hISG15, K48-Ub_2 and Ub_1 which can be explained by non-stoichiometric binding model and sequence preference outside the conserved recognition motif. To reveal details of substrate recognition, we determined structures of non-covalent complexes of $PLpro^{CoV-2}$ (C111S proteolytically inactive mutant) with hISG15 and K48-Ub_2. These are the first structures of non-modified, complete complexes and they uncover that hISG15 binding is determined by recognition of both UBL domains while K48-Ub_2 is recognized mainly through the proximal Ub. We further examined this PLpro binding to K48-Ub_2, Ub_1 and hISG15 using NMR experiments. These data, together with cross-linking mass spectrometry (XL-MS) suggest that $PLpro^{CoV-2}$ interacts with both UBL domains of ISG15 whereas K48-Ub_2 is recognized largely through the proximal Ub, with distal Ub contributing less to binding through different interactions. We used modelling to predict alternative modes of PLpro binding to substrates that are consistent with cross-linking data. Finally, we tested our binding models by performing an *in silico* $\Delta\Delta G$ alanine scan on $PLpro^{CoV-2}$ in complex with K48-Ub_2/hISG15 substrates and show differential domain utilization by the PLpro for the two substrates. Our findings

uncover binding heterogeneity in PLpro interactions with hISG15 and ubiquitin substrates that decouples binding affinity from proteolytic activity.

RESULTS AND DISCUSSION

Sequence and topological differences between hISG15 and K48-Ub_2

Recent biochemical binding and cleavage assays have shown that $PLpro^{CoV-1}$ prefers K48-Ub_2 while the related $PLpro^{CoV-2}$ binds more tightly to both human and mouse ISG15 (hISG15 and mISG15). A nearly 20-fold stronger affinity compared to Ub_2[17,20] suggests that the sequence variation at the substrate binding interface between $PLpro^{CoV-1}$ and $PLpro^{CoV-2}$ may dictate substrate specificity[17,20]. Importantly, these previous studies on $PLpro^{CoV-1}$/$PLpro^{CoV-2}$ binding to Ub_2 used a non-hydrolysable synthetic triazole linker between the Ubs rather than a native isopeptide K48 linkage, raising questions how linker geometry and rigidity may influence binding to PLpro[17,20,21]. When considering both domains in K48-Ub_2 and hISG15, they are 33% identical in sequence (Supplementary Fig. 1a) while the distal (N-terminal) UBL domain of hISG15 is 29% identical to Ub, and the proximal (C-terminal) UBL domain of hISG15 has a slightly higher sequence identity of 37%. hISG15 and mISG15 are 63% sequence identical, and both have similar sequence identities to Ub (Supplementary Fig. 1a,b). Intriguingly, however, a recent study reported that hISG15 binding to $PLpro^{CoV-2}$ has an order of magnitude higher on- and off-rates than mISG15 binding[17]. Importantly, Ub and UBLs of hISG15 and mISG15 also vary in the active binding surfaces (Supplementary Fig. 1b,c). The protein domains in hISG15 and K48-Ub_2 have homologous folds but their sequences and topologies of how the two domains are linked are different (Fig. 1a and Supplementary Fig. 1d).

The earlier reported structure of the $PLpro^{CoV-1}$:K48-Ub_2 complex shows the proximal Ub bound to the Zn finger and palm domains via its surface hydrophobic patch[21] (comprising residues L8, I44 and V70) placing the C-terminal tail modified with allylamine in a groove that is covalently linked to active site C111 (Supplementary Fig. 1d)[22]. A recent structure of full-length mISG15 bound to $PLpro^{CoV-2}$ [20] revealed a distinct binding mode of the proximal and distal UBL domains of mISG15. The proximal UBL is shifted away from the finger domain compared to the proximal-Ub binding mode while still placing the C-terminal LRGG tail into the active site of the protease (Supplementary Fig. 1d). Comparison of mISG15 and Ub active surfaces reveals that the hydrophobic patch centered

on I44 in Ub (Supplementary Fig. 1b,c) is more polar in hISG15. Given that the prior studies used a triazole-linked Ub dimer we wanted to test the influence of the linker composition on binding to $PLpro^{CoV-2}$. We used microscale thermophoresis (MST) binding experiments to quantify affinity between $PLpro^{CoV-2}$ and three substrates: hISG15, K48-Ub_2 and Ub_1 (Fig. 1a). Additionally, we tested $PLpro^{CoV-2}$ binding to K48-Ub_2 and Ub_1 containing a C-terminal aspartic acid (D77) after the "LRGG" PLpro recognition site (Fig. 1a), which is typically used for controlled enzymatic synthesis of ubiquitin chains[23]. Fitting our binding data to a 1:1 binding model (Supplementary Fig. 2a) resulted in abnormally high χ^2 and systematic deviation in the residuals (Supplementary Fig. 2b). Improved fits were observed using a model that assumes two binding events with different K_ds (Supplementary Fig. 2a) yielding statistically significant reductions in the χ^2 compared to one binding event for all datasets except Ub_1-D77 (Supplementary Fig. 2b). We find that $PLpro^{CoV-2}$ binds both hISG15 and K48-Ub_2 with high affinity (90 ± 30 nM, 80 ± 30 nM, respectively) and more strongly than Ub_1 (apparent K_d, 160 ± 70 nM) (Fig. 1b) although the actual microscopic K_d of Ub_1 could be even higher as it may be able to bind to multiple sites on $PLpro^{CoV-2}$ (see below). Interestingly, K48-Ub_2 with a C-terminal aspartic acid (K48-Ub_2-D77) binds tenfold weaker (790 ± 540 nM) compared to Ub_2 (Fig. 1b and Supplementary Fig. 2c), and Ub_1-D77 exhibited weak binding (146.67 ± 90.74 μM) (Fig. 1b and Supplementary Fig. 2c). This is consistent with a lack of reported protease substrates with acidic residues at the C-terminus of the "LRGG(X)" motif[7]. Our analysis also suggests the presence of secondary binding events for hISG15 and K48-Ub_2 with μM affinities (Fig. 1b, gray insert). This is supported by cross-linking data where we observed covalent adducts with molecular weight corresponding to heterodimers (PLpro:substrate) but also heterotrimers ($(PLpro)_2$:substrate) (Fig. 1c and Supplementary Fig. 2d,e). The observed heterogeneity of the species formed suggests that ISG15 may bind in a more defined orientation to PLpro while Ub_2 appears to binds in several arrangements. These experiments indicate that there is a dominant binding mode between the substrate and PLpro but higher order complexes are also possible, thus justifying the need to fit our binding data with more complex binding models.

Derived from published structures of mISG15:$PLpro^{CoV-2}$ and Ub_2:$PLpro^{CoV-1}$ [17,20,21], we anticipated that hISG15 and Ub_2 bind PLpro utilizing both UBL/Ub domains. However, the difference in Ub_2 and Ub_1 affinities suggests that the second Ub contributes modestly to the binding. Additionally, Ub_1-D77 binds nearly 1000-fold more weakly compared to WT Ub_1 while Ub_2-D77 yields an affinity more similar to Ub_1, perhaps indicative of a change in binding mode primarily utilizing a single Ub.

To explore how affinity relates to PLpro proteolytic activity, we conducted $PLpro^{CoV-2}$ cleavage assays with hISG15 with modified C-terminal tails mimicking natural SARSC-CoV-2 substrates or K48-linked Ub_3/Ub_2. We found that $PLpro^{CoV-2}$ can efficiently cleave peptides containing the "LRGG motif", but the rate of cleavage varies. The hISG15-Nsp2 fusion is cut faster than fusions with Nsp3 and Nsp4 peptides (Fig. 1d and Supplementary Fig. 2f). We investigated how amino acid X (position 158) at the C-terminus of "LRGG(X)" motif can impact cleavage. When Ala of Nsp2 peptide is substituted with Glu or Asp, the fusion peptide is being cut the slowest (Fig. 1e and Supplementary Fig. 2g). We also found that $PLpro^{CoV-2}$ hydrolyzes K48-Ub_3 to Ub_2 and Ub_1 rapidly, but the subsequent cleavage of Ub_2 to Ub monomers is slow (Fig. 1f,g). Finally, we found that Ub_2-D77 is cleaved more rapidly compared to Ub_2, which suggests that despite Ub_2-D77 binding with a notably lower affinity to $PLpro^{CoV-2}$ it must be bound differently than Ub_2 to enable the hydrolysis (Supplementary Fig. 2h). If Ub_3 binds predominantly with two Ub units, there are two possible binding modes of Ub_3 on PLpro. To test which binding mode is dominant we used a K48-Ub_3 that contains three "distinct" Ub units: 1 (mutant Ub-K48R), 2 (U-^{15}N-labeled Ub) and 3 (Ub-D77) (Fig. 1g), allowing mass spectrometry-based identification of each cleavage product. Analysis by mass spectrometry of a cleavage time course of this Ub_3 construct reveals that the first cleavage occurs between Ub units 2 and 3, releasing the C-terminal Ub_1-D77 with only minor products for the other Ubs (Fig. 1g). This is consistent with units 1 and 2 bound to PLpro with the C-terminal tail of unit 2 fitting the active site for rapid hydrolysis (Fig. 1g). Three Ub binding sites (S2, S1, and S1') were proposed for PLpro[20]. In this model, unit 1 would bind to S2, unit 2 to S1 and unit 3 to S1' (Supplementary Fig. 1d). Interestingly, Ub_2 can bind to PLpro in two modes. In the high-affinity mode, as in Ub_3 binding, where Ub unit 1 (distal) binds to S2 and unit 2 to S1, results in no cleavage. Only the second, lower-affinity mode is productive, wherein Ub unit 1 is bound to S1 and unit 2 (proximal Ub) located at S1'. In this arrangement the C-terminal tail of unit 1 connecting the two Ubs is placed in the active site of PLpro, and Ub_2 is cut into monomers. These two modes of binding are competitive and because there is a difference in affinity, the rate of cleavage is reduced, but eventually resulting in complete disassembly of Ub_2. Thus, our experiments uncover more complex alternate binding modes and sequence dependence of PLpro for two related substrates that have not been described to date. Our data also provide alternative interpretation of recently published work[17,20] as the composition and flexibility of the Ub-Ub linker can significantly impact the binding or cleavage or both for Ub_2 and other protein substrates. This may explain the previously observed difference in

affinity of PLpro^{CoV-2} between Ub_2 and ISG15 which is likely attributed to changes in mode of binding and/or conformational flexibility of the substrate linkage rather than mutations in the PLpro enzyme.

Dual vs. single domain substrate recognition determines PLpro^{CoV-2} selectivity

We further investigated details of interaction between the PLpro^{CoV-2}, hISG15 and K48-Ub_2 to reveal similarities and differences. We determined crystal structures of PLpro^{CoV-2} with an active site C111S mutation, which inactivates PLpro, in complex with hISG15 at 2.98 Å resolution (Fig. 2a) and with K48-Ub_2 at 1.88 Å resolution (Fig. 2b). For the PLpro^{CoV-2}:hISG15 complex we observe well resolved electron density for the proximal and distal UBL domains bound to S1 and S2 sites respectively (Fig. 2a). By contrast, for the PLpro^{CoV-2}:K48-Ub_2 structures determined at much higher resolution we observe strong electron density for the proximal Ub bound to S1 site with only weak signal for the distal Ub (Fig. 2b and Supplementary Fig. 3b). Despite low electron density for the distal Ub, some regions of densities resemble α-helix and amino acid chains corresponding to portions of the distal Ub from superposed structure of PLpro^{CoV-1} bound to a K48-Ub_2 (PDB id: 5E6J) upon only slight adjustment (Supplementary Fig. 3c). Clearly, there is sufficient room in our crystals to fit the distal Ub at site S2 in a binding mode alike PLpro^{CoV-1} bound to a K48-Ub_2 (Supplementary Fig. 3d)[21]. An alternative explanation is that PLpro^{CoV-2} (C111S) can hydrolyze the K48 linkage slowly, leading to a mixture of Ub_1 and Ub_2. To test this directly, we ran SDS-PAGE gels of our crystals and observed predominantly Ub_2 with only a minor Ub_1 species (Supplementary Fig. 3e). Our structures revealed how the protease differentially recognizes hISG15 using both UBL domains while K48-Ub_2 is predominantly recognized using the proximal Ub. We additionally determined structures of the free hISG15 and K48-Ub_2 to 2.15 Å and 1.25 Å resolution, respectively. The bound and free hISG15 conformations are similar with a root-mean-square deviation (rmsd) for Cα atoms 1.34 Å (Fig. 2c) based on alignment of the C-terminal (proximal) UBL. An analogous comparison of our unbound K48-Ub_2 structure with the only known PLpro-bound conformation of K48-Ub_2[21] revealed an rmsd of 10.2 Å. The Ub units are oriented very differently relative to each other (Fig. 2d) with the PLpro-bound Ub_2 in an extended conformation (like hISG15) while the unbound Ub_2 in an "open" conformation with the functional hydrophobic surface patches exposed for binding. We also compared the "bound" conformation with two other published canonical "open" (Fig. 2e) and "closed" (Fig. 2f) conformations of K48-Ub_2 revealing large differences in rmsd of 10.4 Å (PDB id: 3NS8) and 12.64 Å (PDB id: 1AAR), respectively. In the "open" conformation the functional binding

surface on Ub is exposed while in the "closed" conformation the functional nonpolar surfaces are engaged in intramolecular Ub-Ub interactions (Fig. 2f, arrow). Interestingly, our new K48-Ub_2 structure is nearly identical to the previously reported "open" conformation (PDB id: 3SN8) with an rmsd of 0.23 Å[24]. These data reflect that hISG15 is more rigid while the K48-Ub_2 exists as ensemble of states which may influence binding and recognition by PLpro[25].

Functional surfaces of hISG15 and K48-Ub_2 are recognized differentially by $PLpro^{CoV-2}$

We first compared the binding modes of hISG15 and proximal Ub in our structures. The structures of PLpro in both complexes are similar with an rmsd of 0.85 Å, but the active surface contacts of the proximal Ub are shifted towards the palm domain of PLpro compared to the proximal UBL domain of ISG15 (Fig. 2g, h). This shift in binding mode is manifested by a rotation of 21.7° around the C-termini of hISG15 and Ub displacing the N-terminal residue by 10.3 Å (Fig. 2h). This is despite the structures of the hISG15 proximal UBL domain and Ub being similar, with an rmsd of 0.96 Å (Supplementary Fig. 4a); thus it is likely dictated by differences in PLpro binding to the Ub and UBL interacting surfaces (Supplementary Fig. 4b).

We also compared our structure to the previously published $PLpro^{CoV-2}$:mISG15[17] complex. Overlay of the two structures (PDB ids: 7RBS and 6YVA) reveals good structural similarity with the overall rmsd of 0.70 Å for PLpro and 1.40 Å for hISG15 and mISG15 (Supplementary Fig. 5). The proximal UBL domains of both ISG15s are well aligned and make several conserved interactions with the PLpro but interaction with the distal UBL domain shows important differences (see below) with the largest conformational deviation in the distal UBL domain (Supplementary Fig. 5a). We compared the contacts from the distal domain of mISG15 and hISG15 to previously determined hotspot residues (F69 and V66) on $PLpro^{CoV-2}$ [17]. We find that in the $PLpro^{CoV-2}$:mISG15 structure K30 and M23 of the distal UBL of mISG15 interact with F69 of $PLpro^{CoV-2}$, while V66 of $PLpro^{CoV-2}$ interacts with A2 of the substrate (Supplementary Fig. 5b). By contrast, in our new $PLpro^{CoV-2}$:hISG15 structure residue 30 of hISG15 is an alanine, thus leaving M23 alone to stabilize the interaction with F69 of $PLpro^{CoV-2}$, while the N-terminus of hISG15 interacts with V66 of $PLpro^{CoV-2}$ (Supplementary Fig. 5c). Residue 20 in ISG15 makes similar nonpolar contacts with V66 but it varies between the mouse (T20) and human (S20) protein (Supplementary Fig. 5b, c). We additionally compared the interactions between the proximal UBL domains of the mISG15 and hISG15 where the UBL binds in a similar binding mode (Supplementary Fig. 5d). We find that overall, the two ISG15 proteins

make similar, but not identical contacts determined by the sequence variation between mISG15 and hISG15. This suggests that the virus may have a different impact if infects distinctive species. The central interacting residues on PLpro^{CoV-2} are Y171, E167 and M208 which interact with conserved R153/R151, W123/W121 and P130/P128 on the proximal domains of hISG15 and mISG15, respectively (Supplementary Fig. 5e,f). The interaction is centered on a salt bridge between E167 of PLpro^{CoV-2} and R153 of hISG15 while the equivalent arginine (R151) in mISG15 is not oriented properly to form a salt bridge. Nonetheless, this core interaction is stabilized by nonpolar interactions of the surrounding residues from both sides of the interface including Y171 of PLpro^{CoV-2} and W123/W121 and P130/P128 from hISG15 and mISG15, respectively (Supplementary Fig. 5e, f). By contrast, interactions with R166 of PLpro^{CoV-2} vary more significantly between mISG15 and hISG15. In the mISG15 structure the side chain of M208 is not resolved while in the hISG15 structure M208 packs against R166 (Supplementary Fig. 5f). Interestingly, R166 forms a salt bridge with E87 of mISG15 which is changed to asparagine (N89) in hISG15 (Supplementary Fig. 5e, f). To compensate for this loss of interaction, N151 of hISG15 makes a hydrogen bond with R166 (Supplementary Fig. 5e,f). This highlights subtle sequence changes between mISG15 and hISG15 that allow interface rearrangements while preserving the binding mode. Our structures show that hISG15 binds PLpro^{CoV-2} utilizing both proximal and distal domains (Fig. 2a), while binding of K48-Ub_2 is primarily driven by interaction with the proximal domain with only weak density observed for the distal domain (Fig. 2b and Supplementary Fig. 3b). We also compared our PLpro^{CoV-2}:hISG15 structure to a recent structure of PLpro^{CoV-2} bound to only the proximal domain of hISG15[20] (Supplementary Fig. 6a; PDB id: 6XA9, 2.9 Å resolution). As in the PLpro^{CoV-2}:mISG15 complex, the structural similarity is high, with an overall Cα rmsd of 1.0 Å. A comparison of the interface contacts reveals nearly identical interactions, even preserving side-chain rotamers between the proximal hISG15 and PLpro in the two structures (Supplementary Fig. 6b). Finally, our new structure of PLpro^{CoV-2}:K48-Ub_2 is nearly identical in binding mode to the previously published structure of PLpro^{CoV-2}:Ub_1 (Supplementary Fig. 6c; PDB id: 6XAA, 2.7 Å resolution) with a Cα rmsd of 0.32 Å and nearly identical side chain rotamers at the interface (Supplementary Fig. 6d). Interestingly, our structure was determined to higher resolution and without the introduction of a covalent linkage of Ub_2 to PLpro^{CoV-2} suggesting that the covalent linkage does not alter the physiological binding of the substrate. In contrast, however, introduction of a synthetic linker between the Ubs in Ub_2 does influence binding of the substrate to PLpro^{CoV-2}.

Interactions of K48-Ub_2 and ISG15 with $PLpro^{CoV-2}$ in solution using NMR

We then used NMR to further characterize $PLpro^{CoV-2}$ binding to hISG15 and K48-Ub_2 and to examine if the contacts observed in crystals also occur in solution. The addition of unlabeled $PLpro^{CoV-2}$ (C111S) caused substantial perturbations in the NMR spectra of ^{15}N-labeled hISG15 (Fig. 3a). We observed disappearance of signals of free hISG15 and emergence of new ones; this indicates slow-exchange binding regime[26,27], consistent with the sub-μM K_d values measured by MST. The strongly attenuated signals of hISG15 residues, including the C-terminal G157, are consistent with our crystal structure of $PLpro^{CoV-2}$:hISG15 complex (Fig. 3a, h). A similar behavior was observed for K48-Ub_2 upon addition of $PLpro^{CoV-2}$ (Fig. 3b-c), where both the distal and proximal Ubs exhibited strong attenuation or disappearance of NMR signals and emergence of new signals, primarily for residues in and around the hydrophobic patch as well as the C-termini. The affected residues mapped to the binding interface in our $PLpro^{CoV-2}$:K48-Ub_2 crystal structure (Fig. 3i), and the slow-exchange behavior is also consistent with the sub-μM K_d values. The slow-exchange binding regime for both hISG15 and K48-Ub_2 is also generally consistent with the reported slow off-rates (0.2, 0.4 s^{-1}) [17]. $PLpro^{CoV-2}$ also caused noticeable perturbations in the NMR spectra of Ub_1, although these were weaker than in Ub_2, and several residues showed gradual signal shifts indicative of fast-exchange, consistent with our MST measurements (Fig. 3d).

We also performed reverse-titration NMR experiments where unlabeled ISG15, K48-Ub_2, or Ub_1 was added to ^{15}N-labeled $PLpro^{CoV-2}$. Both ISG15 and K48-Ub_2 caused substantial perturbations in the ^{15}N-$PLpro^{CoV-2}$ spectra (Fig. 3e-f). Particularly noticeable was the change in the indole NH signals of W93 and W106 located in close proximity to the active site of PLpro, as well as of imidazole NH signal that might be attributed to the active-site H272 (Fig. 3e-f, Supplementary Fig. 7), in agreement with the C-termini of ISG15 and Ub_2 entering the active site of PLpro in our crystal structures (Supplementary Fig. 7f). The addition of Ub_1 caused significantly lesser overall ^{15}N-$PLpro^{CoV-2}$ signal perturbations, although the W_ε and H_δ signal shifts were clearly visible when Ub_1 was in significant excess (Fig. 3g, Supplementary Fig. 7c,e). Even at 8-molar excess of Ub_1 both free and bound W_ε/H_δ signals were present, consistent with weaker binding. Taken together, the NMR data suggest that the apparent strength of $PLpro^{CoV-2}$ binding is: ISG15 ≈ K48-Ub_2 > Ub_1, consistent with our MST data.

These NMR data indicate that binding to PLpro involves both UBLs of hISG15 and both Ubs of K48-Ub_2. Interestingly, despite being identical and having very similar chemical shifts in the unbound

state (Supplementary Fig. 8a), the distal and proximal Ubs show markedly different signal perturbations indicative of distinct contacts with PLpro (Fig. 3b,c, and Supplementary Fig. 8b). While the perturbed residues in the proximal Ub agree well with our PLpro^{CoV-2}:K48-Ub$_2$ crystal structure, where this Ub occupies the S1 site, several perturbations observed in the distal Ub (most notably for N25, K27, K29, D32, K33 in the α-helix) are inconsistent with the PLpro^{CoV-1}:K48-Ub$_2$ crystal structure. Thus, we cannot exclude possible additional modes of interaction between the distal Ub and PLpro^{CoV-2}. This might explain the low electron density for the distal Ub in the PLpro^{CoV-2}:K48-Ub$_2$ crystal structure. Interestingly, for several residues in Ub$_1$ the shifted signals upon addition of PLpro^{CoV-2} appear at positions intermediate between those in the distal and proximal Ubs (Supplementary Fig. 8b), suggesting that Ub$_1$ might be sampling both S1 and S2 sites on PLpro.

In agreement with the MST results, our NMR data demonstrate that placement of an aspartate at the C-terminus of ISG15, K48-Ub$_2$, and Ub$_1$ reduced substantially their affinity for PLpro^{CoV-2}, as evident from noticeably weaker NMR signal perturbations observed in both the D-extended substrates and PLpro (Supplementary Figs. 8e and 9). It should be mentioned that in all the NMR studies presented here the addition of PLpro^{CoV-2} resulted in the overall NMR signal broadening/attenuation reflecting an increase in the size (hence slower molecular tumbling) upon complexation with a ~36 kDa protein. The finding that hISG15 and K48-Ub$_2$ bind to the same sites on PLpro^{CoV-2} enabled us to directly compare their affinities for PLpro^{CoV-2} in a competition assay where hISG15 was added to a preformed PLpro^{CoV-2}:K48-Ub$_2$ complex, and the bound state of Ub$_2$ was monitored by ^{1}H-^{15}N NMR signals of ^{15}N-labeled proximal Ub (Fig. 3j). Titration of unlabeled hISG15 into a 1:1.5 mixture of K48-Ub$_2$ and PLpro^{CoV-2} resulted in gradual disappearance of PLpro-bound signals of Ub and concomitant emergence of free K48-Ub$_2$ signals at their unbound positions in the spectra (Fig. 3j, Supplementary Fig.8c). The observed decrease in the intensity of the bound signals agrees with the prediction based on the K_{d1} values for K48-Ub$_2$ and ISG15 derived from our MST experiments (Fig. 3j) but not with the K_d values reported previously[17] (Supplementary Fig. 8d).

Since ISG15 binds to the same PLpro^{CoV-2} surface as Ub$_2$ and contains an uncleavable linkage, we then examined if ISG15 can inhibit polyUb cleavage by PLpro. When hISG15 was added to cleavage reaction of Ub$_3$ or Ub$_2$ by PLpro^{CoV-2} it did not interfere with Ub$_3$ cleavage to Ub$_2$ (Fig. 3k, left gel) but it blocked hydrolysis of Ub$_2$ to monomers (Fig. 3k, right gel, also Supplementary Fig. 8e). A similar effect was observed on cleavage of Ub$_2$ in the presence of Ub$_1$ (Supplementary Fig. 8e). This can be explained by different options for productive cleavage of Ub$_3$ and Ub$_2$. The productive

cleavage of Ub_3 is predominantly accomplished by binding of two Ub units (2 and 1) to S1 and S2 sites on $PLpro^{CoV-2}$, respectively and the unit 3 of Ub_3 occupying the S1' site, thus placing the isopeptide bond on its K48 in the active site of $PLpro^{CoV-2}$. This is a high affinity Ub_3 binding, and ISG15 and particularly Ub_1 cannot easily compete for binding. As discussed earlier, K48-Ub_2 can bind in two different modes, one with two Ub domains binding to S1 and S2 sites on PLpro, but this binding cannot not result in cleavage. In order to break the isopeptide bond between two Ubs the distal Ub must bind to the S1 site on PLpro such that the proximal Ub will then occupy the S1' site. The LRGG motif can then be recognized, and the isopeptide bond is cleaved. But Ub_2 binding through a single Ub unit to S1 site is of low affinity, thus both ISG15 and Ub_1 can compete with Ub_2 and inhibit its cleavage.

Specific contacts in PLpro:substrate complexes detected with XL-MS

Our structural experiments indicate differences in how ISG15 and Ub_2 are recognized by $PLpro^{CoV-2}$. To gain more insight into the proposed dynamics of the interactions we employed a XL-MS approach (Fig. 4a) using three different cross-linking chemistries. We found that the 4-(4,6 dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM) cross-linker produced robust heterodimers of $PLpro^{CoV-2}$ with hISG15, K48-Ub_2, Ub_1 or K48-Ub_2-D77 (Fig. 4b and Supplementary Fig. 10a). We identified two contacts between D61 and D62 on $PLpro^{CoV-2}$ to K35 on the distal UBL of hISG15 (Fig. 4c, 19 and 43 contacts) which map well onto our structure with the distances between carboxylates (D61 and D62) and N_ζ (K35) of 14.9 and 8.6 Å, respectively (Fig. 4c) with C_β-C_β distances below 30 Å consistent with the cross-linker geometry[28]. By contrast, we detected 19 cross-links between PLpro and K48-Ub_2, however, due to the sequence degeneracy between the two Ubs we interpreted the data based on shortest distance (Supplementary Fig. 10b). Using this strategy, 12 of the 19 observed contacts fall below a 30 Å threshold (Fig. 4d). Of these 12 contacts, 7 involve K6 from the distal Ub to the N-terminal thumb domain of $PLpro^{CoV-2}$ consistent with the distal Ub binding mode seen in the $PLpro^{CoV-1}$:Ub_2 structure[21]. Additionally, of the 12 contacts, two between K190 on the fingers domain of $PLpro^{CoV-2}$ to E64 and E18 of Ub have Cβ-Cβ 16 Å and 28.6 Å distances which are compatible with the proximal Ub geometry (Fig. 4d). These data suggest contacts compatible with placement of the Ubs in the S1 (proximal) and S2 (distal) sites. Similarly, for Ub_1 we found 23 cross-links that localize to both S1 and S2 sites (Supplementary Fig. 10c), consistent with our NMR data. For the $PLpro^{CoV-2}$:K48-Ub_2 cross-links we can explain 12 of 19

identified pairs based on Cβ-Cβ based on distances and geometry alone. To find alternate binding sites, we designed a modeling strategy in which the distal Ub is modeled by docking a Ub monomer to PLpro:Ub complex containing Ub in the S1 (proximal) binding site and utilizing constraints that place the docked (distal) Ub with its C-terminus in proximity to K48 of the proximal Ub (Fig. 4e, g, blue spheres). We produced over 5000 models and compared the energy of the assembly as a function of the sum of distances between 7 cross-linked atoms pairs that were unexplained in the first model (Supplementary Fig. 10d). We found a low energy model that explains an additional 4 of the 7 cross-links (Fig. 4f) that localize the Ub to the PLpro's UBL domain and the lower part of the N-terminal thumb domain (Fig. 4g). In a parallel docking approach, we assumed an alternative binding mode where the distal Ub was placed in the S1 (proximal) binding site (Supplementary Fig. 10e), and applied a constraint from the C-terminus of that distal Ub to K48 of the docked (proximal) Ub. Similarly, we compared the ensemble of structures as a function of total energy and the sum of distances between cross-link pairs (Supplementary Fig. 10e). We found a low energy model in which the docked Ub is placed near the PLpro's UBL that explains 4 of 7 contacts (Supplementary Fig. 10f). These analyses indicate that the 4 contacts can be explained with two different binding modes. To resolve this discrepancy, we performed XL-MS on $PLpro^{CoV-2}$:Ub_2 in which only the proximal Ub was uniformly ^{15}N labeled. Analysis of cross-links containing the unlabeled (distal) Ub uncovered 14 cross-links (Supplementary Fig. 10g) of which 8 are compatible with placement of this Ub in the S2 (distal) site and 7 of these again involve K6 interacting with the N-terminal thumb domain similar to the data collected with unlabeled Ub_2 (Fig. 4d). This allowed us to propose that K48-Ub_2 can bind to PLpro in two different binding modes (see Fig. 4e and Supplementary Fig. 10e).

Having more confidence that the distal Ub is bound in the S2 site, we again mapped the remaining 7 unexplained contacts onto the docked model in which we sampled movement of the distal Ub (Fig. 4e); this model can explain 4 additional contacts (Fig. 4f, g). Finally, we also interpreted XL-MS data on $PLpro^{CoV-2}$:Ub_2-D77 and using a model derived from the sampling of the S1 site (Supplementary Fig. 10e) we can explain the two contacts consistent with alternate binding modes of Ub_2-D77 (Supplementary Fig. 10i) with one contact comprising the distal Ub bound to the S1 site and the second involving the proximal Ub bound to the UBL domain of PLpro (S1' site). This change in binding mode may explain faster cleavage kinetics of Ub_2-D77 compared to Ub_2 (Supplementary Fig. 2h). Our combined experiments not only reaffirm the dominant binding modes between PLpro and

ISG15 and K48-Ub_2 (to S1 and S2 sites) but also begin to clarify alternate binding modes that explain the heterogeneity of binding to S1 and S1' sites.

In silico alanine scan of $PLpro^{CoV-2}$ in complex with substrates reveals binding hotspots for Ub_2 and ISG15

To better understand the energetic contribution of the residues at the PLpro:substrate interfaces, we applied an *in silico* alanine scan approach[29] to $PLpro^{CoV-2}$ in complexes with hISG15 and K48-Ub_2 (Fig. 4h). We first identified PLpro interface residues that contact the substrates and employed Rosetta[29] to calculate $\Delta\Delta G_{binding}$ comparing WT and alanine mutants. Analysis of $\Delta\Delta G_{binding}$ for $PLpro^{CoV-2}$ with the two substrates revealed interaction hotspots, most notably $PLpro^{CoV-2}$ residues E167/Y264 and F69, for stabilizing S1 and S2 Ub/UBL binding sites in K48-Ub_2 and hISG15 substrates (Fig. 4i). Additionally, we find that for $PLpro^{CoV-2}$, the distal site (S2) has a preference for ISG15 (Fig. 4i,j, colored in red) but the proximal site (S1) has an overall preference for Ub (Fig. 4i,j, colored in blue). For K48-Ub_2 the primary interaction is with the proximal Ub domain in S1 site but additional interaction comes also from the distal Ub interacting with S2 site contributing to stronger binding. In protein complexes, residues that surround protein interaction hotspots typically play important roles in determining specificity[30]. Indeed, in the S1 binding site, Y171 provides more stabilization for hISG15 compared to Ub_2 (Fig. 4i). These analyses highlight how prediction of binding energetics combined with structural data can help interpret dynamics of domain binding.

DISCUSSION

Dynamics of Ub/UBL binding – insight into coupling of linked domains

The literature highlights that $PLpro^{CoV-1}$ prefers binding to K48-linked polyubiquitin over ISG15[21]. Our work incited by recent studies[17,20] reveals that $PLpro^{CoV-2}$ recognizes ISG15 and K48-Ub_2 with very similar affinities. Sequence analysis suggests that the PLpro from these two viruses only vary at 8 amino acid positions at the substrate binding interface implicating only minor sequence changes responsible for improving binding of ISG15. Interestingly, both K48-Ub_2 and ISG15 utilize two Ub/UBL domains to recognize and bind $PLpro^{CoV-2}$ but our data show they do it differently. In ISG15 the two UBL domains are connected through a relatively short, likely more rigid peptide linker (DKCDEP in hISG15), and the C-terminal (proximal) UBL binds to S1 site while the N-terminal

(distal) UBL binds to S2 site on PLpro^{CoV-2}. The amino acid sequences of the proximal and distal UBLs are somewhat different and show distinct contacts that are required for productive binding. In K48-Ub_2 two identical Ub units are connected through a flexible linker (RLRGG-K48)[31,32] and contribute differently to binding. In the high affinity complex with PLpro the proximal Ub binds to the S1 site and the distal Ub binds to S2 site. The proximal Ub is very well ordered and shows multiple interactions with PLpro. The distal Ub interacts with PLpro differently. In the crystal structure this Ub domain is less ordered, suggesting multiple possible states. NMR data clearly show interactions of the distal Ub with PLpro, but presence of less occupied states cannot be excluded. K48-Ub_2 also seems to bind to PLpro with an altered register where the distal Ub binds to the S1 site while the proximal Ub occupies the S1' site. This is consistent with how PLpro disassembles Ub_3 by cleaving off the unit 3 and the fact that the binding curves of Ub_2 and Ub_1 to PLpro^{CoV-2} are explained better by presence of two binding sites. Both hISG15 and the K48-Ub_2 should be sensitive to mutations of S1 and S2 sites in PLpro, but because of different interaction modes, binding of the hISG15 and the K48-Ub_2 substrates is likely to have different sensitivity to such mutations. Furthermore, evolutionary analysis of SARS-CoV-2 variants (Supplementary Fig. 11) highlights sequence variation at key sites suggesting the occurrence of PLpro variants that may have altered substrate specificity with unknown disease outcomes.

The structure of PLpro^{CoV-1}:K48-Ub_2 complex[21] revealed both Ubs bound to the protease S1 and S2 binding sites, with proximal and distal Ubs connected via a non-hydrolysable triazole linker (in lieu of the native isopeptide linkage) and the C-terminal tail covalently attached to the protease. This may rigidify the K48-Ub_2 concealing the true interactions. There is no structure of the ISG15 bound to PLpro^{CoV-1} that could reveal their interactions. Our structure of PLpro^{CoV-2}:hISG15 as well as the previous structure of PLpro^{CoV-2}:mISG15 reveal a dual UBL domain recognition binding mode despite surprising species sequence variation at the UBL active surfaces (Supplementary Fig. 1a-c). Our *in silico* alanine scan of the interfaces uncovered residues that may play central roles in stabilizing the ISG15 binding mode for PLpro^{CoV-2} and implicated V66 and F69 as being important for stabilizing the distal UBL domain of ISG15.

Conformational changes in PLpro substrates responsible for observed "single domain" and "dual domain" recognition mode

Much of the focus on understanding how sequence variation impacts pathogenicity, infectivity and virulence of SARS-CoV-2 has been centered on sequence changes in surface proteins such as the receptor-binding domain (RBD) of the spike protein which are essential for recognition of ACE2, virus entry into host cell and thus infectivity. Furthermore, there is concern that mutations in the viral receptors may overcome vaccines which were designed against an engineered prefusion "stabilized" conformation of RBD of the spike protein, particularly worrisome with emerging new variants like the Omicron BA.2 and Ontario WTD clade[33]. Therefore, additional SARS-CoV-2 life cycle steps must be explored, and appropriate key drug targets identified to expand treatment options.

Viral interference with host innate immune response is one of these steps of which ISG15 is integral. Several coronavirus Nsps have been shown to contribute to diminishing this complex response mechanism. Modeling of the protein interfaces suggests that the sequence variation between PLpro from SARS-CoV-1 and SARS-CoV-2 plays a role in recognition specificity of host factors. Furthermore, we also show that sequence variation within PLpro from 2.3 million SARS-CoV-2 isolates is overall distributed with some hotspots that mimic sequence variation observed between SARS-CoV-1 and SARS-CoV-2. While we do not understand how differential recognition of Ub compared to ISG15 impacts pathogenicity and virulence of SARS coronaviruses, balance between dysregulation of the protective interferon response and ubiquitin-proteasome systems likely influences virus interaction with the host defense mechanisms. Future work must be focused on understanding how protease specificity impacts pathogenicity. Furthermore, it remains unknown whether PLpro encodes additional specificity for the substrates that are linked to Ub or ISG15 modifications.

$PLpro^{CoV-2}$ must recognize and process multiple substrates: polyproteins 1a and 1ab, polyUb and ISG15. It is also known to cleave several human host proteins. All these substrates have common sequence recognition motif "LRGG", however they differ in several ways. In coronavirus polyproteins and several host proteins PLpro cleaves a regular peptide bond. In K48-polyUb and ISG15-modified protein substrates the cleaved isopeptide bond is between the C-terminal carboxylate of Ub/UBL and a lysine side chain of Ub or other protein, but these latter substrates differ in how the Ubs or UBLs are linked. It is interesting that conformation of PLpro in complexes with Ub, hISG15 and mISG15 is very similar (0.7 Å rmsd). However, the substrates conformations differ.

What may be the biological implication of single vs dual domain recognition of Ub/UBL for hydrolysis of polyUb/ISG15 modifications? ISG15 is gene-coded fused dimer, it functions as a di-

UBL and is attached covalently to proteins as such. Its specific removal by viral PLpro is also hard-wired to its dimer structure. Ubiquitin is different as it exists as a monomer and is added to a polyUb chain or other proteins in units of monomer. However, PLpro shows the highest affinity for Ub_2 (or longer chains) vs Ub_1 and it most efficiently removes the proximal Ub (unit 3) from Ub_3. Because PLpro binds single Ub less strongly, this suggests that in the cell proteins tagged with polyUb chains containing an odd number of Ubs may accumulate Ub_1-substrate adducts, as shown in our cleavage studies.

In summary, we find that $PLpro^{CoV-2}$ binds both hISG15 and K48-Ub_2 with high and similar affinity but shows weaker interactions with Ub_1. Our data also suggest the presence of lower affinity secondary binding sites for hISG15 and K48-Ub_2 which can be explained by alternate binding modes. We determined the first structures of untethered $PLpro^{CoV-2}$ complexes with hISG15 and K48-Ub_2. We combine these data with solution NMR, cross-linking mass spectrometry, binding and computational modeling. Our collective data showed that high affinity binding to hISG15 is determined by dual domain recognition while K48-Ub_2 is recognized mainly through proximal Ub, with distal Ub contributing to binding using different binding modes. Our findings pave the way to understand the interaction of PLpro with hISG15 and (poly)ubiquitin substrates and uncover binding heterogeneity that appears to decouple binding affinity from protease activity. Future experiments will focus on how sequence changes in PLpro influence the distribution of primary and secondary binding sites and how this modifies proteolytic activity. Finally, it remains unknown how changes in the protease sequence dictate substrate binding modes and viral pathogenesis outcomes.

Materials and Methods Summary

WT and mutant $PLpro^{CoV-2}$, hISG15, ISG15-D158, Ub, and Ub-D77 were expressed and purified as previously described[17,23,34]. K48-linked Ub_2, Ub_2-D77 and Ub_3 were enzymatically synthesized as previously described. For MST experiments, substrates were labeled with NHS Cy5 dye. Labeled substrate was incubated across a range of PLpro concentrations in triplicate and changes in fluorescence were measured on a NanoTemper Monolith NT.115 available at University of Texas Southwestern Medical Center. Obtained data were fit and analyzed in PALMIST 1.5.8 using 1:1 and 1:2 binding models and visualized in GUSSI 1.4.2[35,36]. WT $PLpro^{CoV-2}$ was incubated with hISG15-Nsp2, hISG15-Nsp3, hISG15-Nsp4, Ub_3, Ub_2, Ub_2-D77 and the reactions resolved by SDS-PAGE.

Ub_3 ((Ub-K48R)-(^{15}N-Ub)-(Ub-D77)) was proteolyzed with WT PLpro^{CoV-2} and cleavage products were quantified in parallel by SDS-PAGE and mass spectrometry on a Bruker Maxis-II ultra-high resolution Q-TOF mass spectrometer available at University of Maryland (UMD). For glutaraldehyde (GA) cross-linking experiments, C111S PLpro^{CoV-2} was incubated with hISG15, Ub_2, Ub_1 at 1:4, 1:1 and 1:0.25 ratios (using 6 μM protease) and reacted with 0.05% GA, quenched and resolved on an SDS-PAGE gel. Crystals for C111S PLpro^{CoV-2}:Ub_2 (K48R/D77), C111S PLpro^{CoV-2}:hISG15, Ub_2 alone and hISG15 alone were obtained using sitting drop vapor diffusion using screen conditions in MCSG-2 F11, Index E11, Pi-PEG D1 and MCSG1 G2 and frozen using 25% glycerol. Diffraction data was collected at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source at Argonne National Laboratory using the program SBCcollect. The data were processed with HKL3000[37], refined using REFMAC[38] from CCP4 suite[39] and modeled into the density maps using Coot[40]. ^{15}N-Ub_1, ^{15}N-Ub_2, ^{15}N-ISG15, and ^{15}N-PLpro^{CoV-2} (C111S and C111S/Y171H) were expressed and purified as previously described. NMR measurements were performed at 25 °C on Avance III 600 MHz and 800 MHz Bruker NMR spectrometers equipped with cryoprobes at University of Maryland. The data were processed using Topspin (Bruker) and analyzed using Sparky 3.190[41]. NMR signal assignments for hISG15 were obtained from Biological Magnetic Resonance Data Bank (BMRB Entry ID 5658) and adjusted to the temperature and buffer conditions used in our studies. For XL-MS studies, C111S PLpro^{CoV-2} was incubated with hISG15, Ub_2, Ub_1 and reacted with DMTMM, DSS and SuFEx[42-44]. Cross-linked reactions were resolved by SDS-PAGE and the heterodimer bands were extracted and the peptides processed through the xQuest XL-MS pipeline[45]. Samples were analyzed on a Thermo Orbitrap Fusion Tribrid system at the UTSW and the data processed by xQuest to identify high confidence contacts. Cross-links were visualized on the structures using PyMOL and the distances between cross-linked atoms were calculated using an in-house script. Ub was docked onto PLpro:$Ub_{proximal}$ complexes to identify alternate binding sites. Over 5000 models were produced using the docking module in Rosetta[46] using two different geometric restraints: 1) between K48 of proximal Ub and C-term of docked Ub to identify alternate S2 sites or 2) between C-term of distal Ub and K48 of docked Ub to identify alternate S1' sites. The sum of distances between all cross-link pairs were evaluated as a function of energy of the complex. Low energy docked models that explained the cross-links were further evaluated. The Flex ddG protocol[29] was combined with an *in silico* alanine scan to evaluate the per residue energetics of PLpro^{CoV-}

^{2}:hISG15 and PLpro^{CoV-2}:Ub$_2$ complexes. A full description of the materials and methods is described in supplementary information.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Statistics and Reproducibility

The results included in this manuscript can be reproduced by following protocols and using materials described in Materials and Methods.

Data availability

The structural datasets generated during the current study are available in the Protein Data Bank repository (https://www.rcsb.org/) under PDB id: 7RBR for PLpro^{CoV-2}, C111S mutant, in complex with K48-Ub$_2$, PDB id: 7RBS for PLpro^{CoV-2}, C111S mutant, in complex with human ISG15, PDB id: 7S6P for human ISG15 alone and PDB id: 7S6O for K48-Ub$_2$ alone. Diffraction images are available on server in Dr. W. Minor laboratory https://proteindiffraction.org. Plasmid for expression PLpro^{CoV-2} C111S mutant (NR-52897 Vector pMCSG53 Containing the SARS-Related Coronavirus 2, Wuhan-Hu-1 Papain-Like Protease) is available in the NIH the BEI Resources Repository (https://www.niaid.nih.gov/research/bei-resources-repository). All MST, raw cross-linking mass spectrometry and $\Delta\Delta G_{calc}$ data are available in Source Data 1. All other data generated during the current study including the raw kinetic and biophysical data are available upon request.

ACKNOWLEDGEMENTS

We thank the members of the SBC at Argonne National Laboratory, especially Darren Sherrell, Krzysztof Lazarski and Alex Lavens for their help with setting beamline and data collection at beamline 19-ID. We thank Chad Brautigam in the UTSW Macromolecular Biophysics Resource core for help with MST data analysis. We also appreciate the help of the Proteomics Core Facility at UTSW. Funding for this project was provided in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract HHSN272201700060C (AJ). The use of SBC beamlines at the Advanced

Photon Source is supported by the U.S. Department of Energy (DOE) Office of Science and operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This research was supported by NIH grant GM065334 to D.F. NMR experiments were performed at UMD on instruments supported in part by NSF grant DBI1040158, and the instrument for mass spectrometry measurements was supported by NSF grant CHE-2018860. We acknowledge Dr. Yue Li at UMD for help with Maxis-II Q-TOF measurements. EE was supported by Undergraduate Maryland Summer Scholarship. LAJ is supported by an Effie Marie Cain Scholarship in Medical Research. PMW is supported by an O'Donnell Brain Institute Pilot grant.

AUTHOR CONTRIBUTIONS

AJ, LAJ and DF initiated the project. RJ cloned, expressed and purified the first batch of protein, ME continued with different constructs, mutations and complexes. CT purified and crystallized protein and complexes with ligands and with JO collected diffraction data. JO determined, refined and together with KM analyzed structures. BTL and EE synthesized all Ub chains, expressed and purified all proteins for NMR studies and carried out NMR studies. BTL and CT performed Ub_3 cleavage reactions and their MS analysis. PMW carried out the MST binding and XL-MS experiments and analysis. ΔΔG calculations and analysis were carried out by PMW and VM. Finally, AJ, JO, DF, KM and LJ conceived of and directed the research as well as wrote the manuscript.

COMPETING INTEREST

The authors declare that they have no competing interests.

ORCID

Pawel M. Wydorski https://orcid.org/0000-0002-6983-421X
Andrzej Joachimiak https://orcid.org/0000-0003-2535-6209
David Fushman: https://orcid.org/0000-0002-6634-8056
Lukasz Joachimiak https://orcid.org/0000-0003-3061-5850

FIGURES and LEGENDS

Figure 1

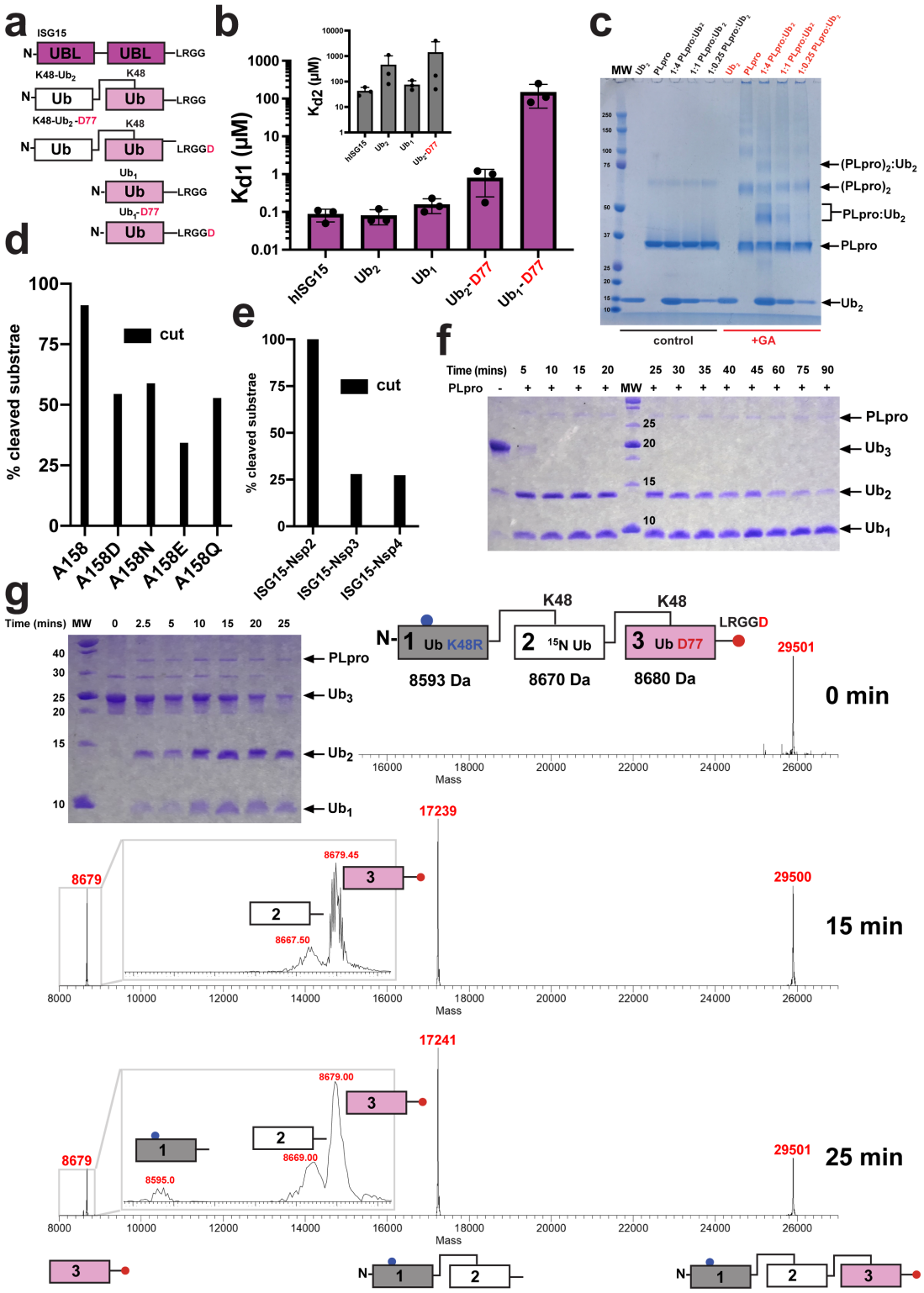

Figure 1. PLpro architecture and substrate recognition. (**a**) Topology of utilized substrates: ISG15, K48-linked Ub_2 ending with G76 on the C-terminus (K48-Ub_2), K48-linked Ub_2 ending with D77 on the C-terminus (K48-Ub_2-D77), and corresponding monomeric ubiquitins. (**b**) MST binding analysis of substrates to $PLpro^{CoV-2}$. Comparison of dissociation constants for the primary binding event is shown in magenta (K_{d1}) and comparison of lower affinity alternate binding modes (K_{d2}) is shown as an insert in grey. Points represent individual replicates, lines represent mean and standard deviation across all replicates. (**c**) Inter- and intramolecular non-specific cross-linking of $PLpro^{CoV-2}$ complexes. Cross-linked samples with different molar ratios reveal formation of heterogeneous covalent $PLpro^{CoV-2}$:Ub_2 heterodimer complex bands (red, for glutaraldehyde cross-linker) compared to untreated reactions (control) by SDS-PAGE. (**d**) Quantification of the cleavage efficiency of hISG15 C-terminal fusions with peptides from Nsp2 (AYTRYVDNNF), Nsp3 (APTKVTFGDD), and Nsp4 (KIVNNWLKQL) from SARS-CoV-2 that mimic natural substrates of $PLpro^{CoV-2}$, as revealed by SDS-PAGE gel. Shown is the percentage of the input population that has been cut by $PLpro^{CoV-2}$. (**e**) Quantification of the cleavage efficiency of Nsp2 peptides as dependent on amino acid located at the C-terminus of "LRGG(X)" motif. Shown is the percentage of the input population that has been cut by $PLpro^{CoV-2}$. (**f**) $PLpro^{CoV-2}$ cleavage of K48-Ub_3. SDS-PAGE gel reveals that PLpro cleaves Ub_3 into Ub_2 and Ub_1 efficiently with various rates. (**g**) Mass spectrometry detection of cleavage patterns for $PLpro^{CoV-2}$ hydrolyzing Ub_3, in which the distal Ub (1) carries K48R mutation, the middle (2) Ub is ^{15}N-labeled, and the proximal (3) Ub contains C-terminal D77 extension. Analysis of the time course reveals that this Ub_3 is primarily hydrolyzed between Ubs 2 and 3. Masses of individual Ub units are shown on the top, and the identified products are shown at the bottom.

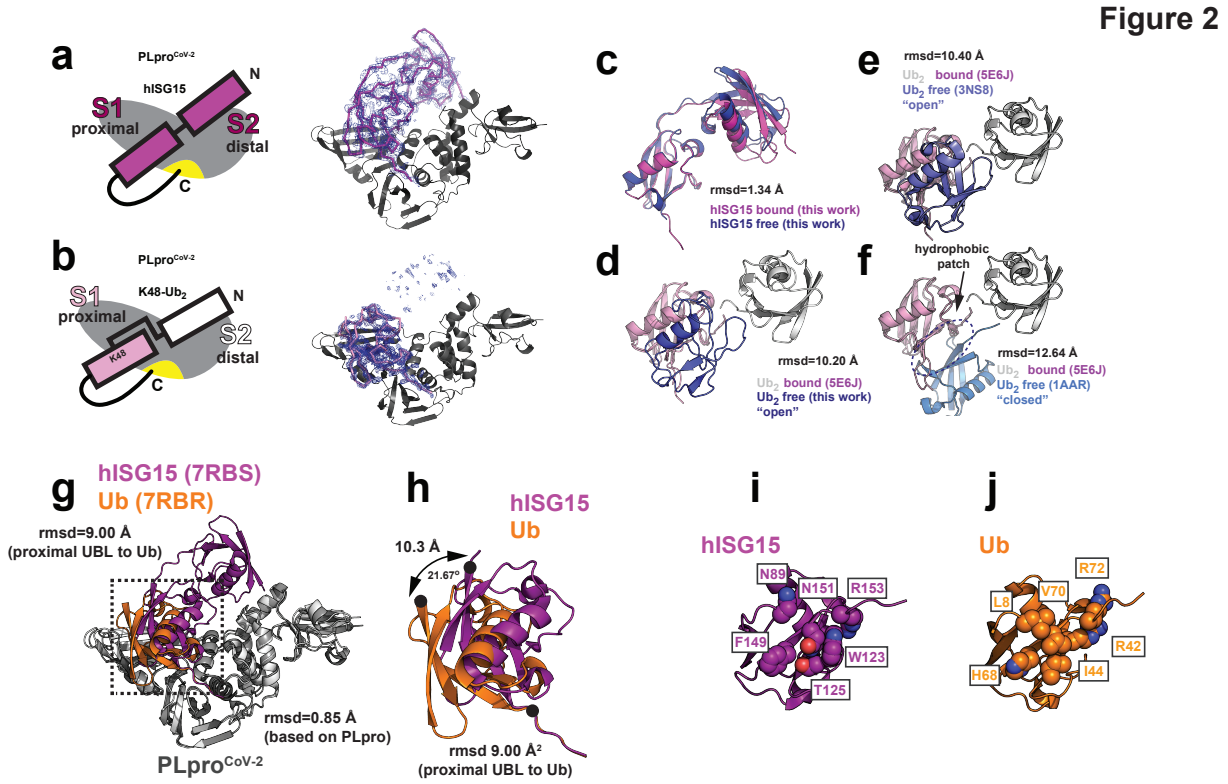

Figure 2. MX structure of C111S PLpro^{CoV-2} bound to human ISG15 and K48-Ub$_2$ reveals differential usage of distal domains. Schematic of PLpro bound to (**a**, left) hISG15 and (**b**, left) K48-linked Ub$_2$. PLpro is shown in grey with an active site indicated in yellow. hISG15 is shown in magenta. Proximal and distal Ub domains are pink and white, respectively. Crystal structures of PLpro bound to (a, right) hISG15 and (b, right) K48-Ub$_2$. PLpro is shown in cartoon representation, hISG15 and K48-Ub$_2$ are shown as a backbone trace. PLpro, hISG15 and K48-Ub$_2$ are colored as in (a, left) and (b, left). The electron density map for PLpro substrates is contoured at 2.0σ. (**c**) Overlay of crystal structures of bound hISG15 (magenta, PDB id: 7RBS, this study) and unbound hISG15 (blue, PDB id: 7S6P, this study). (**d**) Overlay of bound conformation of K48-linked Ub$_2$ observed in complex with PLpro^{CoV-1} (PDB id: 5E6J) with unbound conformation of K48-linked Ub$_2$ (PDB id: 7S6O, this study). Proximal Ub of both bound and unbound conformations is shown in pink, distal Ub of bound conformation is shown in white, and distal Ub of unbound conformation is shown in blue. (**e**) Overlay of bound conformation of K48-linked Ub$_2$ observed in complex with PLpro^{CoV-1}

(PDB id: 5E6J) with unbound "open" conformation of Ub_2 (PDB id: 3NS8). Ub units are shown as in (d). (**f**) Overlay of bound conformation of K48-linked Ub_2 observed in complex with $PLpro^{CoV-1}$ (PDB id: 5E6J) with unbound "closed" conformation of Ub_2 (PDB id: 1AAR). Ub units are shown as in (d). Intramolecular Ub-Ub interface is indicated with an arrow. (**g**) Structural overlay of $PLpro^{CoV-2}$:hISG15 and $PLpro^{CoV-2}$:Ub_2 (PDB id: 7RBR, this study). Proteins are shown in cartoon and colored grey (PLpro), magenta (hISG15) and orange (Ub_2). (**h**) Zoom in of the boxed area in (g) representing overlay of the proximal domain of hISG15 and Ub. Proximal domain of hISG15 and Ub are represented as in (g). Rotation of the active surface indicated with arrows. (**i,j**) Comparison of the active surfaces of hISG15 (**i**) and Ub_1 (**j**). Proximal domain of hISG15 and Ub are represented as in (g). Key residues are shown in spacefill representation.

Figure 3

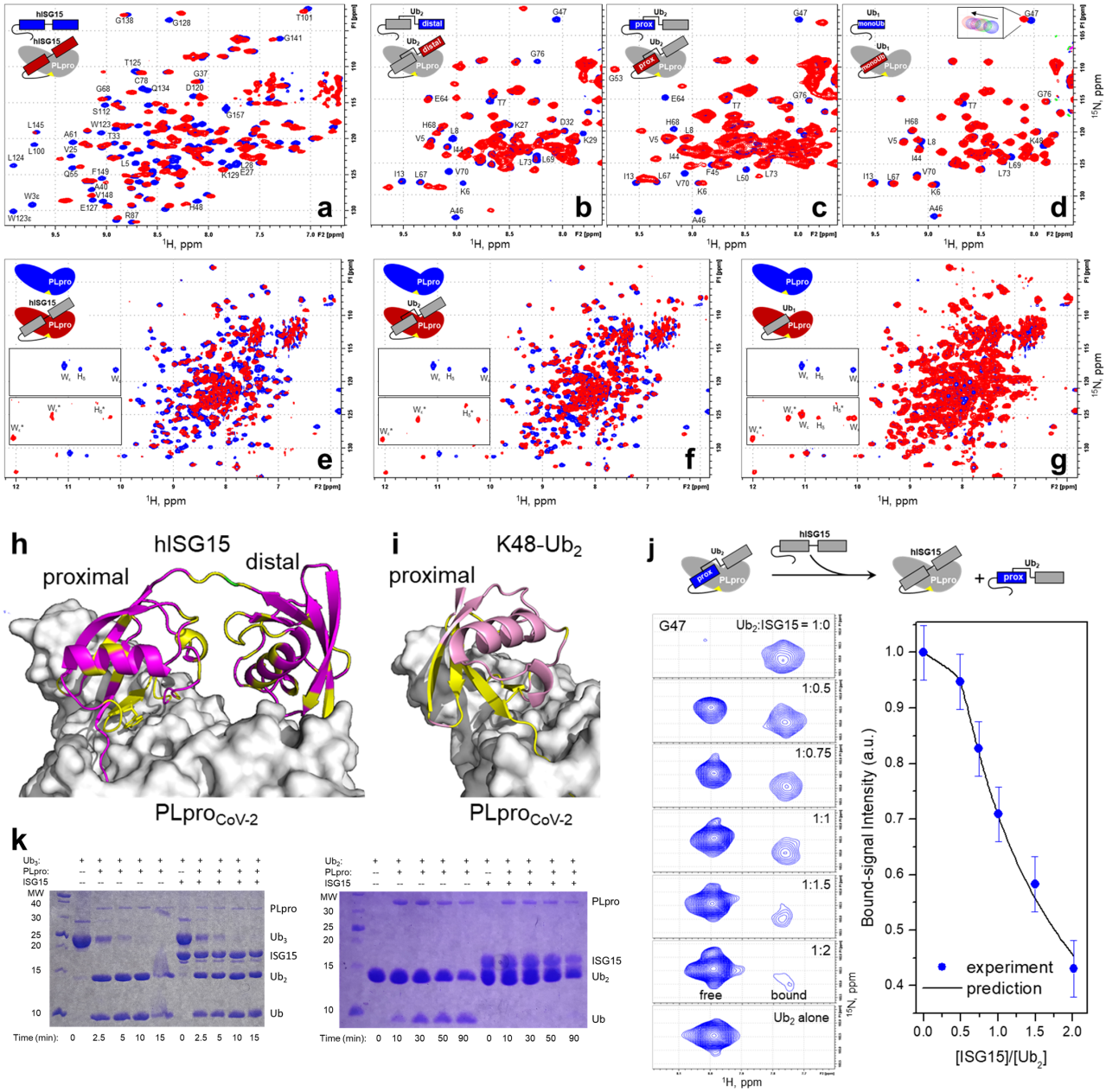

Figure 3. NMR data showing PLpro^{CoV-2} interactions with ISG15 and K48-Ub$_2$. (**a-d**) Overlay of ^{1}H-^{15}N spectra of ^{15}N-labeled (a) hISG15 (TROSY), (b) distal Ub in Ub$_2$, (c) proximal Ub in Ub$_2$, and (d) monomeric Ub (all SOFAST-HMQC), alone (blue) and with 1 (for ISG15) or 2 molar equivalents of unlabeled PLpro^{CoV-2} (C111S) (red). Signals of select residues are indicated. (**e-g**) Overlay of ^{1}H-^{15}N TROSY spectra of ^{15}N-PLpro^{CoV-2} (C111S, Y171H) alone (blue) and (in red) with 1.5 molar equivalents of unlabeled (e) hISG15, (f) Ub$_2$, and (g) 8x excess of Ub$_1$ (red). Insets zoom

on the region containing indole NH signals of tryptophans (W93 and W106) and of a histidine imidazole NH (aliased, likely H272). It should be noted that in the absence of NMR signal assignments for PLpro^{CoV-2} we cannot pinpoint the observed changes to individual residues, however the indole NH (tryptophan) and imidazole NH (histidine) have unique chemical shifts, and there are only two tryptophans in PLpro. (See also Supplementary Fig. 7) (**h**) hISG15 residues with strong signal perturbations mapped (yellow) onto our structure of PLpro^{CoV-2}:hISG15 complex. (**i**) Residues in the proximal Ub with strong signal perturbations mapped (yellow) onto our structure of PLpro^{CoV-2}: K48-Ub_2 complex. (**j**) Competition for PLpro^{CoV-2} binding between Ub_2 and ISG15 (also Supplementary Fig. 8). Shown on the left are representative ^{1}H-^{15}N NMR signals of G47 in ^{15}N-labeled proximal Ub of K48-Ub_2 mixed with 1.5x unlabeled PLpro^{CoV-2} (C111S, Y171H) upon addition of unlabeled ISG15, for indicated values of ISG15:Ub_2 molar ratio. The bottom spectrum corresponds to Ub_2 alone. Shown on the right is the intensity of the PLpro-bound signal of G47 as a function of [ISG15]:[Ub_2] (dots, normalized to the initial intensity) and the predicted molar fraction of bound Ub_2 (black line) using the K_{d1} values obtained in this work (see Supplementary Figs. 2c and 8d) and the equations from[47] (**k**) SDS-PAGE gels showing the inhibitory effect of ISG15 on disassembly of K48-linked Ub_2 by PLpro^{CoV-2} (right gel) and minimal effect (if any) of ISG15 on disassembly of K48-linked Ub_3 (left gel), see also Supplementary Fig. 8e.

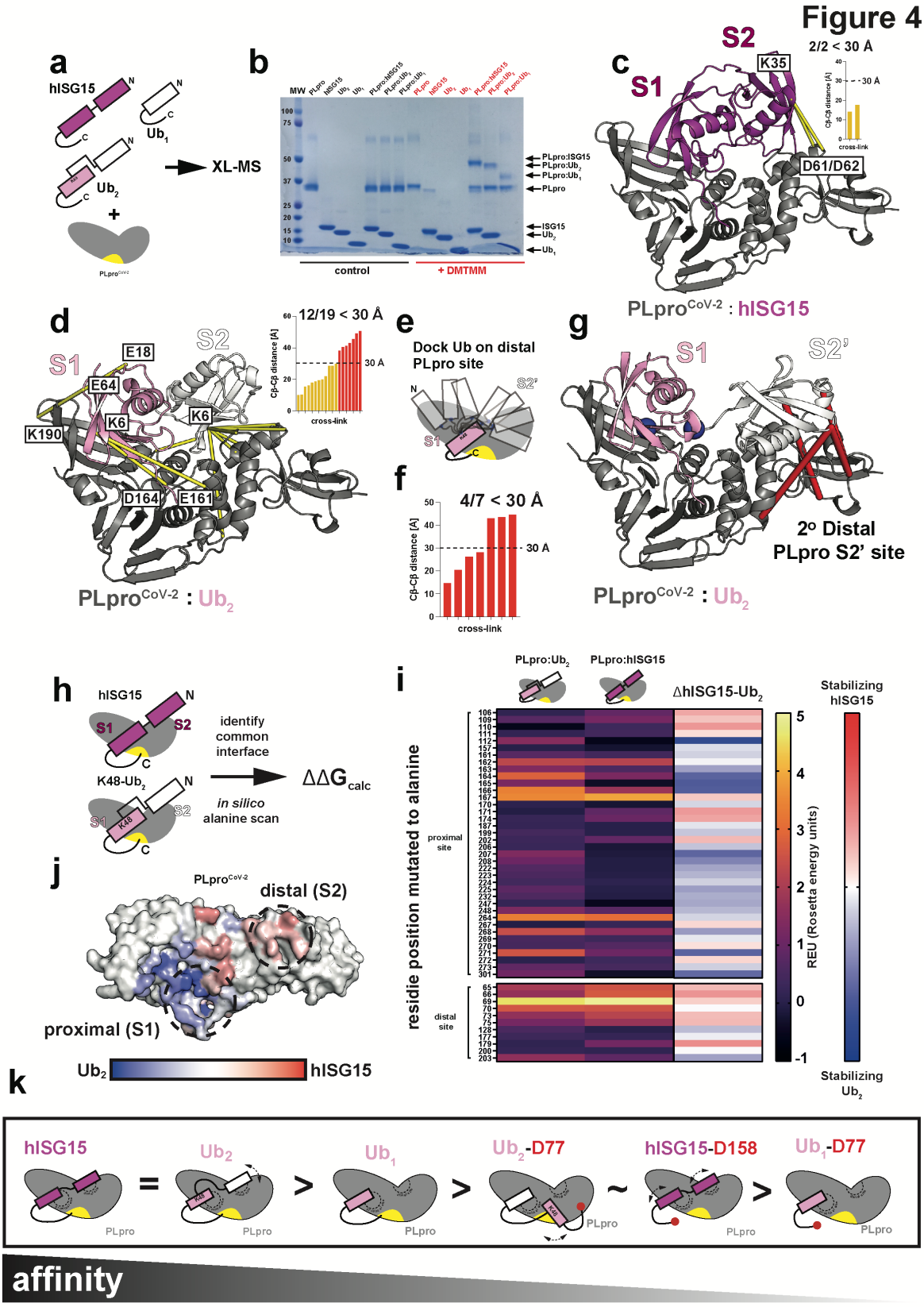

Figure 4. Cross-linking mass spectrometry and energetic analysis of PLpro in complex with

hISG15 and K48-Ub_2. (**a**) Schematic illustration of cross-linking mass spectrometry experiments for heterodimer complexes of $PLpro^{CoV-2}$ with hISG15, K48-Ub_2, and Ub_1. Active site of $PLpro^{CoV-2}$ is indicated in yellow. (**b**) Cross-linked samples reveal formation of covalent heterodimer complex bands (red, DMTMM cross-linker) compared to untreated reactions (control) by SDS-PAGE. (**c**) Interactions between $PLpro^{CoV-2}$ (D61/D62) and the distal hISG15 domain (K35) identified by XL-MS mapped on the structure of heterocomplex. Both identified contacts are shorter than 30 Å. PLpro (grey) and hISG15 (magenta) are shown in cartoon representation. Cross-links are colored yellow. (**d**) Interactions between $PLpro^{CoV-2}$ and K48-Ub_2 identified by XL-MS mapped on the structure of the heterocomplex. Twelve contacts found to be shorter than 30 Å show interaction of K48-Ub_2 with finger, palm and thumb domains of $PLpro^{CoV-2}$. $PLpro^{CoV-2}$ is shown in cartoon representation and colored grey. K48-Ub_2 is shown in cartoon representation and colored white and pink for the proximal and distal Ubs, respectively. Cross-links are colored yellow. (**e**) Modeling strategy showing docked Ub monomer to $PLpro^{CoV-2}$:Ub complex with Ub placed into the S1 binding site. Utilized constraint maintaining proximity between K48 of the proximal Ub and C-terminus of the docked Ub is shown in blue spheres. (**f**) A low energy model generated with strategy presented in (e) explains 4 of the remaining 7 contacts from the XL-MS in (d). (**g**) Contacts shorter than 30 Å between low energy model of $PLpro^{CoV-2}$ bound to Ub_2 in the S1 and S2' sites as shown in (e, f). Cross-links between distal Ub and thumb and UBL domains of $PLpro^{CoV-2}$ are colored red. Constraint used in docking (e) is indicated as blue spheres. $PLpro^{CoV-2}$ and K48-Ub_2 are shown as in (d). (**h**) Schematic illustration of identification of substrate binding surfaces on $PLpro^{CoV-2}$ and *in silico* mutagenesis for heterodimer complexes with hISG15 and K48-Ub_2. (**i**) Heat-map results of $\Delta\Delta G$ calculations of *in silico* alanine scan for $PLpro^{CoV-2}$ in complex with hISG15 or K48-Ub_2. Row titles represent positions of residues on the interaction surface. The last column represents results of calculations of difference between REU (Rosetta energy units) for $PLpro^{CoV-2}$:hISG15 and $PLpro^{CoV-2}$:K48-Ub_2. (**j**) Results of *in silico* mutagenesis for complexes of $PLpro^{CoV-2}$ with hISG15 and K48-Ub_2 mapped on surface representation of $PLpro^{CoV-2}$. Hotspot sites identified as those driving stability towards hISG15 are colored red, and those driving stability towards K48-Ub_2 are colored blue. (**k**) Schematic representation of differences in binding of $PLpro^{CoV-2}$ with hISG15 and ubiquitin species.

References

1. Kim, D. *et al*. A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome. *Nature Communications* **12,** 5120–16 (2021).
2. Kim, D. *et al*. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181,** 914–921.e10 (2020).
3. Mahmoudvand, S. & Shokri, S. Interactions between SARS coronavirus 2 papain-like protease and immune system: A potential drug target for the treatment of COVID-19. *Scand J Immunol* **94,** e13044 (2021).
4. Ziebuhr, J., Thiel, V. & Gorbalenya, A. E. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. *Journal of Biological Chemistry* **276,** 33220–33232 (2001).
5. Mielech, A. M., Chen, Y., Mesecar, A. D. & Baker, S. C. Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res* **194,** 184–190 (2014).
6. Clasman, J. R., Everett, R. K., Srinivasan, K. & Mesecar, A. D. Decoupling deISGylating and deubiquitinating activities of the MERS virus papain-like protease. *Antiviral Res* **174,** 104661 (2020).
7. Sulea, T., Lindner, H. A., Purisima, E. O. & Ménard, R. Binding site-based classification of coronaviral papain-like proteases. *Proteins* **62,** 760–775 (2006).
8. Snyder, N. A. & Silva, G. M. Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response. *J. Biol. Chem.* **297,** 101077 (2021).
9. Ratia, K., Kilianski, A., Báez-Santos, Y. M., Baker, S. C. & Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathog* **10,** e1004113 (2014).
10. Chen, X. *et al*. SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3-TBK1 complex. *Protein Cell* **5,** 369–381 (2014).
11. Li, S.-W. *et al*. SARS coronavirus papain-like protease induces Egr-1-dependent up-regulation of TGF-β1 via ROS/p38 MAPK/STAT3 pathway. *Scientific Reports* **6,** 25754–13 (2016).
12. Moustaqil, M. *et al*. SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg Microbes Infect* **10,** 178–195 (2021).
13. Reynolds, N. D. *et al*. The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect Dis* **7,** 1483–1502 (2021).
14. Mohamud, Y. *et al*. The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochemical and Biophysical Research Communications* **540,** 75–82 (2021).
15. Swaim, C. D. *et al*. Modulation of Extracellular ISG15 Signaling by Pathogens and Viral Effector Proteins. *Cell Reports* **31,** 107772 (2020).
16. Liu, G. *et al*. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat Microbiol* **6,** 467–478 (2021).
17. Shin, D. *et al*. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587,** 657–662 (2020).

18. Perez-Bermejo, J. A. *et al*. SARS-CoV-2 infection of human iPSC-derived cardiac cells reflects cytopathic features in hearts of patients with COVID-19. *Science Translational Medicine* **13,** (2021).
19. Daczkowski, C. M. *et al*. Structural Insights into the Interaction of Coronavirus Papain-Like Proteases and Interferon-Stimulated Gene Product 15 from Different Species. *Journal of Molecular Biology* **429,** 1661–1683 (2017).
20. Klemm, T. *et al*. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J* **39,** e106275 (2020).
21. Békés, M. *et al*. Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. *Mol Cell* **62,** 572–585 (2016).
22. Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M. & Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc Natl Acad Sci USA* **93,** 861–866 (1996).
23. Pickart, C. M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Meth. Enzymol*. **399,** 21–36 (2005).
24. Lai, M.-Y., Zhang, D., Laronde-Leblanc, N. & Fushman, D. Structural and biochemical studies of the open state of Lys48-linked diubiquitin. *Biochim. Biophys. Acta* **1823,** 2046–2056 (2012).
25. Varadan, R., Walker, O., Pickart, C. & Fushman, D. Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology* **324,** 637–647 (2002).
26. Boughton, A. J., Zhang, D., Singh, R. K. & Fushman, D. Polyubiquitin and ubiquitin-like signals share common recognition sites on proteasomal subunit Rpn1. *J. Biol. Chem*. **296,** 100450 (2021).
27. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, W. & Skelton, A. J. *Protein NMR Spectroscopy: Principles and Practice*. (2006).
28. Leitner, A. *et al*. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci USA* **111,** 9455–9460 (2014).
29. Barlow, K. A. *et al*. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J Phys Chem B* **122,** 5389–5399 (2018).
30. Karanicolas, J. *et al*. A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Molecular Cell* **42,** 250–260 (2011).
31. Castañeda, C. A. *et al*. Linkage-specific conformational ensembles of non-canonical polyubiquitin chains. *Phys. Chem. Chem. Phys*. **18,** 5771–5788 (2016).
32. Berlin, K. *et al*. Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc*. **135,** 16595–16609 (2013).
33. Pickering, B. *et al*. Highly divergent white-tailed deer SARS-CoV-2 with potential deer-to-human transmission. *bioRxiv* 2022.02.22.481551 (2022). doi:10.1101/2022.02.22.481551
34. Osipiuk, J. *et al*. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nature Communications* **12,** 743–9 (2021).
35. Scheuermann, T. H., Padrick, S. B., Gardner, K. H. & Brautigam, C. A. On the acquisition and analysis of microscale thermophoresis data. *Analytical Biochemistry* **496,** 79–93 (2016).
36. Tso, S.-C. *et al*. Using two-site binding models to analyze microscale thermophoresis data. *Analytical Biochemistry* **540-541,** 64–75 (2018).
37. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62,** 859–866 (2006).

38. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1,** 213–220 (2014).
39. Winn, M. D. *et al*. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67,** 235–242 (2011).
40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).
41. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327 (2015).
42. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* **9,** 120–137 (2014).
43. Leitner, A. *et al*. Chemical cross-linking/mass spectometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 9455–9460 (2014).
44. Yang, B. *et al*. Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 11162–11167 (2018).
45. Rinner, O. *et al*. Identification of cross-linked peptides from large sequence databases. *Nat Meth* **5,** 315–318 (2008).
46. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *Journal of Molecular Biology* **373,** 503–519 (2007).
47. Wang, Z. X. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Letters* **360,** 111–114 (1995).