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 Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) genome is evolving as the viral pandemic continues its active phase around the world. The Papain-like protease (PLpro) is a domain of Nsp3 – a large multidomain protein that is an essential component of the replication-transcription complex, making it a good therapeutic target. PLpro is a multi-functional protein encoded in coronaviruses that can cleave viral polyproteins, poly-ubiquitin and protective Interferon Stimulated Gene 15 product, ISG15, which mimics a head-to-tail linked ubiquitin (Ub) dimer. PLpro across coronavirus families showed divergent selectivity for recognition and cleavage of these protein substrates despite sequence conservation. However, it is not clear how sequence changes in SARS-CoV-2 PLpro alter its selectivity for substrates and what outcome this has on the pathogenesis of the virus. We show that SARS-CoV-2 PLpro preferentially binds ISG15 over Ub and K48-linked Ub_2. We determined crystal structures of PLpro in complex with human K48-Ub_2 and ISG15 revealing that dual domain recognition of ISG15 drives substrate selectivity over Ub and Ub_2. We also characterized the PLpro substrate interactions using solution NMR, cross-linking mass spectrometry to support that ISG15 is recognized via two domains while Ub_2 binds primarily through one Ub domain. Finally, energetic analysis of the binding interfaces between PLpro from SARS-CoV-1 and SARS-CoV-2 with ISG15 and Ub_2 define the sequence determinants for how PLpros from different coronaviruses recognize two topologically distinct substrates and how evolution of the protease altered its substrate selectivity. Our work reveals how PLpro substrate selectivity may evolve in PLpro coronaviruses variants enabling design of more effective therapeutics.

INTRODUCTION

The current pandemic of COVID-19 is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). This virus is a typical member of the *Coronaviridae* family and has spherical, enveloped, non-segmented, (+) sense RNA virion with a large ~30 kbs genome[1,2]. The genome is used as mRNA for translation of a replication-transcription complex (RTC) constituents made of two large polyproteins, Pp1a and Pp1ab, and as a template for replication of its own (-) sense copy. The Pp1a and Pp1ab polypeptides are processed by two viral proteases: papain-like protease (PLpro, a domain within non-structural protein 3 (Nsp3)), and 3C-like protease (Nsp5 or 3CLpro or Mpro). The PLpro enzyme is essential for several viral replication processes, including cleavage and maturation of viral polyproteins, assembly of the RTC, and disruption of host viral response machinery to facilitate viral proliferation and replication. Moreover, this enzyme is conserved and found in all coronaviruses, often in two copies, denoted as PL1pro and PL2pro. It was shown that PL1pro is dispensable but PL2pro is essential suggesting that their functions are non-overlapping[3]. These enzymes are found in coronaviruses (alpha, beta, gamma and delta) and in arteriviruses[4]. PLpro is a domain of non-structural protein 3 (Nsp3) – a large multidomain protein (~212 kDa) that is a component of the RTC and has other functions. The enzyme is located between the SARS unique domain (SUD/HVR) and a nucleic acid-binding domain. PLpro cleaves 3 sites in SARS-CoV-2 polyproteins yielding Nsp1, Nsp2 and Nsp3; and the "LXGG↓XX" motif found in Pp1a/Pp1ab corresponds to the P4–P1 substrate positions of cysteine proteases and is essential for recognition and cleavage by PLpro. The "LXGG" motif is also found in human ubiquitin and ISG15. The PLpro active site contains a canonical cysteine protease catalytic triad (Cys111, His272 and Asp286). PLpro may have catalytic properties more common with other cysteine proteases, with the generally accepted thiolate form of Cys111 acting as a nucleophile and Asp286 promoting deprotonation of His272, which serves as a base. Interestingly, the PL1pro and PL2pro enzymes from different coronaviruses exhibits diverse specificities for substrates suggesting that discrete sequence changes play a role in modulating binding affinity. PLpro has been shown to inactivate TBK1, block NF-kappaB signaling, prevent translocation of IRF3 to the nucleus, inhibit the TLR7 signaling pathway, and induce Egr-1-dependent up-regulation of TGF-β1[1,2]. Further illustrating the complex and diverse functions of the protein, in some reports, some PLpro roles are decoupled from its proteolytic activity[5].

The SARS-CoV-2 PL2pro sequence is conserved between SARS-CoV-1 (83% identical, 90% similar), MERS-CoV (30% identical, 49% similar) and other coronaviruses, and their structures are very similar. Fortuitously, it has low sequence similarity to human enzymes. PLpros share structural architecture and catalytic site with the human ubiquitin specific proteases (USPs), one of the five distinct deubiquitinating enzyme (DUB) families, despite low sequence identities (~10%)[6]. In addition, due to the centrality of PLpro to viral replication, it is an excellent and attractive candidate for therapeutic targeting.

In addition to processing virus polyproteins, it was shown in SARS- and MERS-CoVs that their PLpro have deubiquitinating activity, efficiently disassembling Ub and polyUb conjugates[3-5]. Ubiquitination is essential post-translational modification mediated by the Ub-conjugating system that plays complex regulatory roles in humans, including proteasome-dependent protein degradation. This process could be reversed by DUBs enzymes[7]. PLpro also has deISG15ylating activities[6,7]. ISG15 is interferon alpha stimulated gene that is a critical component of antiviral response[8,9]. ISG15 has two Ub-like (UBL) domains and mimics a head-to-tail linked Ub_2. Both Ub and ISG15 proteins carry the PLpro recognition motif at their C-termini. Removal of these modifications from specific substrates in host cells may have diverse impact on many cellular processes and specifically may frustrate the host response to viral infection[8-11]. 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[12]. While it is not precisely clear how ISG15 interferes with viral processes it is believed that it can play many different roles by tagging newly translated viral proteins to sterically prevent their folding, assembly or recognition[13]. While K48-linked Ub_2 and ISG15 are homologous both in sequence and fold, the topologies of how the two domains are linked are distinct. In addition to cleavage of Ub and ISG15 there is also evidence that there are other proteins in human genome that contain PLpro cleavage sites and have been shown to be cleaved, these proteins include MYH6, MYH7, FOXP3, PROS1, ErbB4 ULK1 and IRF-3 [10]. Cleavage of these host proteins may be linked to cardiomyopathy and coagulopathy[11]. It thus remains unknown how PLpro discriminates between substrates and how sequence changes in the protease alter this selectivity.

To understand how SARS-CoV-2 PLpro can differentiate between Ub_1, K48-Ub_2 and human ISG15 (hISG15) substrates, we employed complimentary biochemical, structural and computational approaches. Consistent with published reports[12,13], we find that SARS-CoV-2 PLpro binds with high

affinity to hISG15 but binds more weakly to both K48-Ub_2 and Ub_1. To test how this differential substrate recognition is manifested structurally, we determined structures of SARS-CoV-2 PLpro in complex with human ISG15 and K48-Ub_2 to uncover that high affinity binding to hISG15 is determined by dual domain recognition while K48-linked Ub_2 is recognized through only one domain. We further tested this PLpro substrate binding model leveraging domain specific isotope labeling of K48-Ub_2, Ub_1 and ISG15 in chemical shift perturbation (CSPs) experiments. The NMR experiments, together with cross-linking mass spectrometry suggest that SARS-CoV-2 PLpro recognizes a single Ub domain of K48-Ub_2 but both domains in ISG15 are recognized. Finally, we interpret our findings in the context of PLpro sequence variation performing an *in silico* $\Delta\Delta G$ alanine scan on PLpro from SARS-CoV-2 and SARS-CoV-1 in complex with K48-Ub_2/ISG15 substrates to uncover recognition rules how sequence evolution impacts substrate selectivity. Our findings pave the way to understand the biological impact of sequence changes in PLpro on substrate binding and how evolution of the virus impacts this process.

RESULTS AND DISCUSSION

Sequence and topological differences between hISG15 and K48-linked Ub_2

The two substrates, K48-linked Ub_2 and ISG15 are homologous both in sequence and fold of domains but the topologies of how the two domains are linked are distinct (Fig. 1a). The structure of the $PLpro_{CoV-1}$:K48-Ub_2 complex shows the proximal Ub bound to the Zn finger and palm domains via its nonpolar patch[14] (including residues L8, I44 and V70) placing the C-terminal tail in a groove with the carboxylate near the active site (Fig. 1a)[14]. A recent structure of full-length mouse ISG15 (mISG15) bound to $PLpro_{CoV-2}$[15] revealed a distinct binding mode of the proximal and distal UBL domains of mISG15 (Fig. 1a). 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 (Fig. 1a). Comparison of mISG15 and Ub active surfaces reveals that the nonpolar patch on Ub (Fig. 1b, pink) centered on I44 is more polar in human ISG15 (Fig. 1b, magenta) suggesting a deviation in the polar/nonpolar patterning of this important recognition surface in Ub. 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 mISG15 and cleaves it with near 20-fold enhancement compared to Ub_2[16,17] suggesting that the sequence variation at the substrate binding interface between $PLpro_{CoV-1}$ and

$PLpro_{CoV-2}$ may dictate substrate specificity (Supplementary Fig. 1a). Based on these sequence differences a specificity switch between Ub and ISG15 for PLpro was proposed[13,15]. Furthermore, the human and mouse ISG15 sequence are 65% identical (Supplementary Fig. 1b) but also vary quite significantly in the active binding surfaces (Fig. 1b, purple and Supplementary Fig. 1c). Furthermore, sequence comparison of the distal and proximal UBL domains of human ISG15 to Ub reveals additional variation beyond the topology. To gain insight whether the proposed hISG15 and K48-linked Ub_2 specificity switch is maintained, we used microscale thermophoresis binding experiments to quantify binding between $PLpro_{CoV-2}$ and three substrates: hISG15, K48-linked Ub_2 and Ub_1. (Fig. 1d). Our data indicate that $PLpro_{CoV-2}$ binds to hISG15 with 778 ± 104 nM affinity (Fig. 1f, purple). Binding of $PLpro_{CoV-2}$ to K48 Ub_2 (Fig. 1d, grey) is much weaker than for hISG15, but somewhat stronger than Ub_1 (Fig. 1d, pink). Indeed, while the sequence conservation between mouse and hISG15 deviates even in the "conserved" binding surface, there remains a clear preference of $PLpro_{CoV-2}$ to bind ISG15 over K48-linked Ub_2 and Ub_1. We also show that $PLpro_{CoV-2}$ is active and can efficiently cleave ISG15 with cleavage sequences derived from the Nsp2, Nsp3 and Nsp4 cleavage sites appended following the LRGG motif (Fig. 1e). Finally, we show that the $PLpro_{CoV-2}$ can efficiently hydrolyze K48-linked Ub_3 but the protease surprisingly has no preference for cleavage of the first or second K48-link (Fig. 1f) consistent with no strict requirement to bind Ub in both the proximal and distal domains as indicated by similar affinities between Ub_1 and Ub_2 (Fig. 1d).

Dual vs single domain substrate recognition determines $PLpro_{CoV-2}$ selectivity

To explain differences in affinity between the hISG15 and K48-linked Ub_2 we determined two crystal structures of PLpro from SARS-CoV-2, including a catalytically inactive C111S mutant of $PLpro_{CoV-2}$ in complex with hISG15 at 2.97 Å resolution (Fig. 2a) and K48-Ub_2 at 1.88 Å resolution (Fig. 2b). These structures were solved by molecular replacement and refined as described in Methods; data and refinement statistics are shown in Supplementary Table 1. Structures were refined as described in Methods; data and refinement statistics are shown in Supplementary Table 1. For the $PLpro_{CoV-2}$:hISG15 we observed well resolved electron density for the proximal and distal domains (Fig. 2a). By contrast, for the $PLpro_{CoV-2}$: K48-Ub_2 (Supplementary Fig. 2a), we observed only strong electron density in the proximal domain with only weak signal for the distal domain (Fig. 2b and Supplementary Fig. 2b). Despite of questionable electron density for Ub_2 distal domain, we are convinced of its presence in our crystals because portions of densities resemble amino acid chains

and only slight tuning of Ub_2 distal domain from superposed structure of SARS PLpro bound to a K48-linked Ub_2 (PDB id: 5E6J)[16] is necessary to fit densities (Supplementary Fig. 2c). Also, there is enough room in our crystals to accommodate distal domain in conformation almost identical to the structure of $PLpro_{CoV-1}$ bound to a K48-linked Ub_2 (Supplementary Fig. 2d). The structures of PLpro complexes with two topologically distinct substrates, Ub_2 and ISG15, revealed how the protease differentially recognizes hISG15 using both domains while K48-Ub_2 is predominantly recognized using the proximal domain. These data are consistent with our binding experiments showing that ISG15 binds tightly while Ub_2 and Ub_1 bind similarly but Ub_1 binds more weakly. We additionally determined structures of the hISG15 and K48-Ub_2 in the absence of the protease (Fig. 2c). The hISG15 and the K48-Ub_2 structures were determined to 2.15 Å and 1.25 Å resolution, respectively, using molecular replacement as described in the Methods. We initially compared the bound hISG15 to the free hISG15 conformation and surprisingly, they are quite similar with a Cα rmsd 1.1 $Å^2$ (Fig. 2d). We performed a similar comparison of our unbound K48-Ub_2 structure with the only bound conformation of K48-Ub_2 observed in the previously reported $PLpro_{CoV-1}$ structure[16] revealing a Cα rmsd of 11.9 $Å^2$. We observe that the domains are oriented very differently relative to each other (Fig. 2e) with the bound conformation in an extended conformation (similar to hISG15) while the unbound conformation in an "open" conformation with the functional surfaces (i.e. I44 etc) exposed for binding. We also compared the "bound" conformation with two other canonical "open" (Fig. 2f) and "closed" (Fig. 2g) conformations of K48-Ub_2 revealing large differences by Cα rmsd of 10.4 $Å^2$ and 11.3 $Å^2$, respectively. In the "open" conformation we again observe that the functional binding surface on Ub is exposed while in the "closed" conformation the functional surfaces are engaged in inter-ubiquitin self-self interactions (Fig. 2g, arrow). These data highlight that the hISG15 and the K48-Ub_2 substrates are recognized differently by PLpro and that the substrates appear to have different conformational ensembles in the unbound state that may influence recognition of the substrates.

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

We first compared the binding modes of ISG15 and proximal Ub from our new structures. Similar to previous analyses we find that the structures have a Cα rmsd of 0.85 $Å^2$. This deviation is largely dictated by the conformation of the proximal domain where the active surface of Ub is shifted towards the palm domain compared to the proximal UBL domain of ISG15 (Fig. 3a). This shift in binding

mode (Fig. 3b) between the two domains is largely manifested by a rotation of the functional surface of Ub (Fig. 3c and Supplementary Fig. 3a) compared to ISG15 (Fig. 3d and Supplementary Fig. 3a). This is despite the structure of the two domains being nearly identical with a Cα rmsd of 0.95 $Å^2$ (Supplementary Fig. 3b) thus this largely must be dictated by changes in the binding mode due to the varying amino acid properties of the binding surfaces between Ub and ISG15 (Supplementary Fig. 3cd, Fig. 1b and Supplementary Fig. 1b).

We also compared our new structure to the previously published $PLpro_{CoV-2}$:mISG15[18]. Overlay of the two structures reveals high structural similarity with the overall Cα rmsd of 1.4 $Å^2$ (both chains) and only a 0.7 $Å^2$ rmsd between the PLpro molecules alone (Fig. 3e). The proximal UBL domains of the two ISG15's are also quite well aligned and make several conserved interactions with the PLpro. The largest conformational deviation is found in the distal UBL domain (Fig. 3e). We compared the contacts from the distal domain of mISG15 and hISG15 to hotspot residues on $PLpro_{CoV-2}$. Prior work described that F69 and V66 are important for recognition of Ub substrates in the distal domain[19]. We find that in the $PLpro_{CoV-2}$:mISG15 structure K30 makes a second shell contact to F69 and V66 interacts with A2 (Fig. 3g). By contrast, in our new $PLpro_{CoV-2}$:hISG15 structure this K30 is replaced with an alanine residue and cannot make this contact (Fig. 3f). We also find that M21 of hISG15 makes contact to F69 on $PLpro_{CoV-2}$. This comparison highlights subtle sequence changes between mouse and human ISG15 that allow correct interpretation of the interactions in sites important for recognition of the distal ISG15 domain. Our structures show that the hISG15 binds to $PLpro_{CoV-2,}$ 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. 2b). Additionally, we compared our $PLpro_{CoV-2}$:hISG15 structure to a recent structure of $PLpro_{CoV-2}$ bound to only the proximal domain of hISG15[15] (Supplementary Fig. 3e; PDB id 6xa9 2.9 Å resolution). Similar to the mISG15 structure the overall structural similarity is high with an overall Cα rmsd of 1.0$Å^2$. Additionally, our new structure of $PLpro_{CoV-2}$:K48-Ub_2 is similar in conformation to the previously published structure of $PLpro_{CoV-2}$:Ub_1 (Supplementary Fig. 3f; PDB id 2.7 Å resolution) but it has been determined to higher resolution and without the introduction of covalent linkage of Ub_2 to $PLpro_{CoV-2}$, therefore it should correspond to a functional complex.

NMR studies of K48 Ub_2 and ISG15 binding to $PLpro_{CoV-2}$

We then used NMR to further characterize $PLpro_{CoV-2}$ interactions with 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. 4A). We observed disappearance of signals present in free hISG15 and emergence of new ones; this indicates slow-exchange binding regime[20,21] consistent with the sub-μM K_d values measured by MST. The strongly attenuated hISG15 residues are consistent with our crystal structure of $PLpro_{CoV-2}$:hISG15 complex (Fig. 4b). By contrast, spectral perturbations in K48-Ub_2 and Ub in the presence of $PLpro_{CoV-2}$ were significantly weaker (Fig. 4c-e). Proximal Ub in K48-Ub_2 exhibited gradual signal shifts and residue-specific attenuations (Fig. 4c), primarily for residues in and around the I44-centered nonpolar surface patch and the C-terminus of Ub. The affected residues mapped to the binding interface in our $PLpro_{CoV-2}$:K48-Ub_2 crystal structure (Fig. 4f). Similar signal shifts and attenuations were also detected in the distal Ub of K48-Ub_2, albeit to a somewhat lesser extent (Fig. 4d). $PLpro_{CoV-2}$ also caused perturbations in the NMR spectra of monomeric Ub, although these were weaker than in Ub_2, consistent with our binding data. It should be mentioned that the addition of $PLpro_{CoV-2}$ resulted in the overall NMR signal broadening/attenuation in agreement with the increase in the size upon complexation with a ~36 kDa protein. Taken together, the NMR data suggest that the apparent strength of $PLpro_{CoV-2}$ binding is: ISG15 >> proximal Ub > distal Ub > Ub_1, consistent with our MST data and crystal structures.

XL-MS-based characterization of PLpro:substrate complexes.

Combining cross-linking with mass spectrometry has helped define contact points in proteins and protein complexes[17] and has complemented protein structure prediction methods for establishing relative domains orientation in proteins and assemblies[18-20]. To gain insights into the $PLpro_{CoV-2}$:substrate contacts, we employed a combination of three cross-linkers that can link different moieties on protein surface. Disuccinimidyl suberate (DSS) and 4-(4,6 dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM) allow detection of lysine-lysine and acid-lysine contacts, respectively[21,22]. We additionally used a newly developed hetero-bifunctional cross-linker that combines succinimide reactive group and a sulfonyl fluoride (SuFEx) which link lysines with amino acids containing hydroxyls but also lysines and histidines[23]. We performed cross-linking experiments using individual components and heterodimer complexes formed between PLpro with Ub_2 or hISG15 using all three chemistries. We proceeded with the DMTMM condition which yields high-resolution

zero-length cross-links and where we observed robust production of a heterodimer complex for $PLpro_{CoV-2}$:hISG15 but weaker signal for the $PLpro_{CoV-2}$:Ub_2 (Supplementary Fig. 4a). XL-MS analysis of replicate datasets for the heterodimer species isolated from the DMTMM reactions reveals robust and high frequency detection of 11 and 17 contacts between D61 and D62 on $PLpro_{CoV-2}$ to K35 on the distal domain of ISG15, respectively (Fig. 5a). By contrast, we only detected 4 cross-links between K45 of $PLpro_{CoV-2}$ to D58 on Ub and 4 cross-links between D164 on $PLpro_{CoV-2}$ to K6 on Ub (Fig. 5a). Mapping the D61 and D62 from $PLpro_{CoV-2}$ to K35 to hISG15 cross-link on the new x-ray structure reveals overall good consistency with cross-link geometries between the carboxylate (D61/62) and Nε (K35) with distances of 9.38 Å and 7.7-11.7 Å for the contacts across heterodimer molecules in the asymmetric unit, respectively. These data suggest that this binding loop located between the UBL and thumb domain in $PLpro_{CoV-2}$ may be mobile in solution. By contrast, because of the sequence degeneracy between the proximal and distal Ub domains we cannot unambiguously assign the positions but can interpret the data based on best distance match (Supplementary Fig. 4b). We find that the Ub K6 to $PLpro_{CoV-2}$ D164 cross-link is most consistent with the interaction with the proximal Ub domain with a distance of 21.1Å (Supplementary Fig 4c). The other cross-link between K45 in the UBL domain of $PLpro_{CoV-2}$ and D58 in Ub is long for both proximal and distal Ub placement. We wondered whether the UBL domain on PLpro might be flexible and function as substrate thus placing the UBL onto the distal binding site. Analysis of $PLpro_{CoV-2}$ intramolecular contacts from the $PLpro_{CoV-2}$:hISG15 and the $PLpro_{CoV-2}$:Ub_2 heterodimer datasets identified 17 and 2 intramolecular contacts in the protease, respectively (Supplementary Fig. 4d). Mapping the contacts on the structure of the protease revealed good distance agreement with cross-linker geometries with the exception of the contacts involving the UBL domain and the terminal helix from the thumb domain that reach to the palm domain (Fig. 5c,d and Supplementary Fig. 4d). These intramolecular contacts that cannot be explained by the structure may suggest that the UBL domain in PLpro can potentially sample "substrate bound" conformations. Our cross-link experiments reaffirm that the hISG15 binds tightly to PLpro utilizing a stable bound distal domain. By contrast, in the PLpro:K48-linked Ub_2 complex we only detect a contact between the proximal Ub domain and $PLpro_{CoV-2}$. Finally, our cross-link data may indicate possible structural rearrangements involving the N-terminal UBL domain from PLpro. Additionally, Nsp3 encodes a second UBL domain (Ubl1, PDB id 7KAG) but the role of these UBLs in PLpro function remains unknown.

***In silico* alanine scan of PLpros from SARS-CoV-1 and CoV-2 in complex with substrates reveals binding hotspots and specificity determining residues.**

To better understand the energetic contribution of the residues at the PLpro:substrate interfaces, we applied an *in silico* alanine scan approach on a crystal structure of $PLpro_{CoV-2}$:hISG15 and a model of $PLpro_{CoV-2}$: K48-linked Ub_2 (Fig. 6a). Additionally, we performed similar calculations on a crystal structure of $PLpro_{CoV-1}$:K48-linked Ub_2[28] and a model of $PLpro_{CoV-1}$:hISG15 (Fig. 6b). For each complex, we first identified amino acids on the PLpros that contact both substrates (Fig. 6a,b) and employed a $\Delta\Delta G$ method in Rosetta[24] to mutate each PLpro interface residue to alanine and compare the energy of WT to mutant to calculate a $\Delta\Delta G_{binding}$. Interestingly, this analysis identified interaction hotspots and showed how they vary between the four complexes. Pairwise comparison of the calculated $\Delta\Delta G_{binding}$ for the two substrates bound to $PLpro_{CoV-2}$ revealed positions that provide significant binding contribution in each complex, most notably PLpro residues E167 and F69/Y264 were identified as hotspot residues that partition to the proximal and distal UBL binding sites in both substrates (Fig. 6c). However, comparing the difference in the $\Delta\Delta G_{binding}$ between the two substrates reveals a substrate preference at each site. For $PLpro_{CoV-2}$, the distal site has a preference for ISG15 (Fig. 6c,e, colored in red) but the proximal site has a preference for Ub (Fig. 6c,e, in blue). A similar analysis for $PLpro_{CoV-1}$:hISG15 bound to the two substrates identified similar hotspots sites utilizing E167 and F69/Y264 as core binding sites for each substrate (Fig. 6d). Unexpectedly, the $\Delta\Delta G_{binding}$ difference highlights the opposite trend observed in $PLpro_{CoV-2}$ where $PLpro_{CoV-1}$ prefers Ub binding in the distal site (Figs. 6d,f, colored in blue). Consistent with our structural and biochemical findings $PLpro_{CoV-2}$ stabilizes interactions to the distal domain of ISG15 (Fig. 6). In protein complexes, residues that surround protein interaction hotspots typically play important roles in determining specificity[25]. Indeed, residues that surround the hotspots contribute differently between the two complexes. For example, in the proximal-binding site, residue Y171 provides significant binding energy in the complex with hISG15 but does not contribute much in the complex with Ub_2 (Fig. 6c). Consistent with the $PLpro_{CoV-1}$ preference for Ub_2 this residue is modified to histidine (Fig. 6d). More importantly this site also appears to be mutating in the SARS-CoV-1 population (Supplementary Fig. 5), suggesting that this position may play a central role in switching substrate specificity and probing how these changes impact virus fitness in human populations. Additional residues that vary between the two coronaviruses are V66, T75 and P77 which surround the second hotspot F69 position and thus also likely switch substrate specificity in the distal UBL binding site. Mutations of these

individual positions that vary between the two predict modest changes with only V66S in $PLpro_{CoV-2}$ appears to shift binding of the distal domain to Ub (Supplementary Fig. 5a) and conversely mutations of S66V in $PLpro_{CoV-1}$ have a predicted opposite effect (Supplementary Fig. 5b). These analyses highlight how prediction of binding energetics combined with structural data can help guide the interpretation in changes in specificity for substrate binding in similar complexes against a panel of substrates.

Evolution of $PLpro_{CoV-2}$ sequence suggests alterations in substrate specificity.

The evolutionary origins of SARS-CoV-2 have been proposed to be derived from bats, and indeed comparison of PLpro from BAT_RaTG13[32] to the $PLpro_{CoV-2}$ reveals that overall the two sequences are conserved with only 7 substitutions of which two (S170 and K232) fall in the substrate binding sites that could impact specificity (Fig. 7a). We also wondered whether the proposed hosts for RaTG13 coronaviruses have conserved ISG15 sequences. Different species belonging to the genus of Pteropodidae (large flying bats) have been identified as a major source of lineage D β coronaviruses[26]. Sequence alignment of ISG15 from 4 species of large flying bats shows overall high sequence conservation for this protein (Supplementary Fig. 6a). Furthermore, it appears that while this sequence is not absolutely conserved in comparison with human ISG15 the active binding surface on the proximal domain is more similar to human than mouse. Furthermore, sequence comparison of human ISG15 to ISG15 derived from the black flying fox bat (*Pteropus alecto*), a putative host of the BAT_RaTG13 has only 62% identity but the active surface on the proximal UB domain is more conserved than mouse ISG15 (Supplementary Fig. 6b and Supplementary Fig. 1a). Together with high PLpro sequence similarity between SARS-CoV-2 and RATG13 it may be possible that PLpro evolved to interact with ISG15 in bats which is also compatible with human ISG15. Mapping the SARS-CoV-1 and SARS-CoV-2 variation in PLpro in the context of Ub_2 and ISG15 binding reveals 8 sites of sequence variation that localize to substrate binding surfaces in the fingers, the lower thumb and upper thumb that make contacts with the proximal and distal Ub/UBL substrate binding sites (Fig. 7b). The two mutated sites observed in the BAT_RATG13 sequence are also present in the SARS-CoV-1. Additionally, there are sites (i.e. Y171) that make stabilizing contacts to ISG15 (Fig. 7b). Specifically, residue Y171 in $PLpro_{CoV-2}$ makes key interactions to P130 in a loop on ISG15 while in the SARS-CoV-1 sequence the tyrosine is mutated to histidine which cannot form stabilizing contacts with ISG15. Another important PLpro residue in this region is E167. It coordinates a network

of interactions including two direct hydrogen bonds to R153 of ISG15 and its aliphatic portion of the side chain makes hydrophobic interaction with W123. In the complex of PLpro with Ub E167 makes one hydrogen bond to R42. Because of W123 mutation to R in proximal Ub this interaction is no longer possible. In the distal Ub/UBL-binding sites, we also observe several mutations that include changes like V66S, T75L and P77E (Fig. 7b). Notably, these changes alter the nonpolar *vs*. polar patterning of the protease surface which might play a role in recruiting the distal UBL domain and might indicate adaptation of PLpro to hISG15. Finally, we performed a similar analysis on 1.1 million available SARS-CoV-2 sequences derived from the Coronavirus3D database[33] to begin to understand the scale of sequence variation in PLpro across all sequenced isolates. Surprisingly, many positions that vary between SARS-CoV-1 and SARS-CoV-2 sample mutational sites observed in the SARS-CoV-2 isolates, including contacts to the proximal and distal Ub/UBL domains (Fig. 7c). This analysis suggests that the $PLpro_{CoV-2}$ protease may be evolving new substrate specificities, perhaps towards higher selectivity to hydrolyze ubiquitinated substrates. It is not clear how changes in substrate specificity alter disease outcomes, but current data suggest that ISG15 is a preferred target of $PLpro_{CoV-2}$ potentially allowing SARS-CoV-2 to narrow its response and more efficiently evade the interferon induced mechanism by inactivating ISGylation[34,35]. The MERS-CoV infections had the highest mortality rate from all coronaviruses[36] and its PLpro activity is described as a "general" broad-specificity DUB[37]. It has been shown to disassemble Ub chains of almost all linkages by engaging proximal and preproximal (C-terminal to proximal) sites for Ub chain recognition. In contrast to MERS-CoV, $PLpro_{CoV-1}$ engages the proximal and distal sites to recognize the "distal" Ub to bind Ub_2 (ISG15) on the substrate and developed specificity for K48-linked Ub chains[38].

DISCUSSION

PLpro sequence changes impact recognition of ubiquitin and ISG15 substrates.

The literature highlights that $PLpro_{CoV-1}$ encodes substrate specificity with a preference for binding to K48-linked polyubiquitin over ISG15[16]. By contrast, our work coupled with recent studies from Dikic et al. and Komander et al. highlight that $PLpro_{CoV-2}$ preferentially recognizes ISG15 over K48-linked polyubiquitin. 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 this specificity switch. Thus far it has remained unclear how this specificity switch is

distributed across the sequence and binding interfaces. The original structure of $PLpro_{CoV-1}$ bound to K48-linked Ub_2 revealed both Ubs bound to the protease through binding sites present in both proximal and distal Ubs with minor caveats of the linkage not being hydrolysable and the C-terminal tail covalently linked to the protease. There remains no structure of the ISG15 bound to $PLpro_{CoV-1}$ suggesting that the dramatically reduced affinity must impact binding both proximal and distal binding UBL domains of ISG15. Our new structure of $PLpro_{CoV-2}$ bound to human ISG15 as well as the previously determined structure of $PLpro_{CoV-2}$ bound to mISG15 reveal a dual UBL domain recognition binding mode despite surprising sequence variation between the mouse and human forms of ISG15, in particular the substrate binding interface (Fig. 1b). Our *in silico* alanine scan of the interfaces uncovered which residues may play central roles in stabilizing the ISG15 binding mode for $PLpro_{CoV-2}$ and implicated V66 as being important for stabilizing the distal UBL domain of ISG15. Furthermore, our evolutionary analysis of SARS-CoV-2 variants highlights sequence variation at key sites suggesting the PLpro variants exist that may have altered substrate specificity with unknown disease outcomes.

Impact of PLpro sequence drift on substrate specificity

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. 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 suggest that the sequence variation between PLpro from SARS-CoV-1 and SARS-CoV-2 plays are role in recognition specificity of host factors. Furthermore, we also show that sequence variation within PLpro from 1.1 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, balance between dysregulation of

the protective interferon response and UPS systems likely has important impacts in the virus evading host defense systems. 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.

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

$PLpro_{CoV-2}$ must recognize and process multiple substrates: polyproteins 1a and 1ab, polyUb and ISG15. All these substrates have common sequence recognition motif "LXGG", however they differ in several ways. In polyproteins PLpro cleaves regular peptide bond. In K48-linked poly-Ub and ISG15 the cleaved peptide bond is between lysine side chain (Ub or other protein substrate) and C-terminal carboxylate, but these substrates have two distinct topologies. We and others have showed that $PLpro_{CoV-2}$ switched specificity from K48-Ub_2 to ISG15. The switch is associated with sequence variation but also with conformational changes that are required to utilize these mutations. It is interesting that conformation of PLpro in complexes with Ub, hISG15 and mISG15 is very similar (0.7 $Å^2$ rmsd on Cα atoms). However, conformations of substrates differ.

CONCLUSIONS

Our studies revealed in detail the sequence determinants for how PLpro from SARS-CoV-2 recognizes two substrates and how sequence evolution of the protease may have remarkable effects on substrate specificity but the role of these sequence changes in disease remain unknown. Analysis of mutations on SARS-CoV-2 shows that many different factors beyond interaction with host cell receptors, receptor activation or viral replication rates, etc. may contribute to viral infectivity, pathogenicity and virulence. PLpro encodes new enzymatic activity that interferes with innate immune responses of host and thus it impacts how viral components interact with host factors. The fact that PLpro from SARS-CoV-2 prefers ISG15 over Ub as a substrate suggests that the virus response is simplified and now focuses on ISG15-mediated pathways. Insight into how the protease balances binding different substrates as the virus continues to evolve will enable design of more effective therapies.

MATERIALS AND METHODS

Gene cloning, protein expression and purification of WT and C111S mutant of PLpro

The gene cloning, protein expression and purification were performed using protocols published previously[34]. Briefly, the Nsp3 DNA sequence corresponding to PLpro protease SARS-CoV-2 was optimized for *E. coli* expression using the OptimumGene codon optimization algorithm followed by manual editing and then cloned directly into pMCSG53 vector (Twist Bioscience). Human ISG15 gene was also synthesised and cloned directly into pMCSG53 vector. The plasmids were transformed into the *E. coli* BL21(DE3)-Gold strain (Stratagene). *E. coli* cells harboring plasmids for SARS CoV-2 PLpro WT and C111S mutant and ISG15 expression were cultured in LB medium supplemented with ampicillin (150 μg/mL).

For large-scale purification of PLpro-C111S, 4 L cultures of LB Lennox medium were grown at 37 °C (200 rpm) in the presence of ampicillin 150 μg/ml. Once the cultures reached OD600 ~1.0, the temperature setting was changed to 4 °C. When the bacterial suspensions cooled down to 18 °C they were supplemented with 0.5 mM IPTG and 40 mM K_2HPO_4 (final concentration). The temperature was set to 18 °C for 20 hours incubation. Bacterial cells were harvested by centrifugation at 7,000g and cell pellets were resuspended in a 12.5 ml lysis buffer (500 mM NaCl, 5% (v/v) glycerol, 50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 10 mM β-mercaptoethanol, 1 μM $ZnCl_2$) per liter culture and sonicated at 120W for 5 minutes (4 sec ON, 20 sec OFF). The cellular debris was removed by centrifugation at 30,000g for 90 minutes at 4 °C. The supernatant was mixed with 3 ml of Ni^{2+} Sepharose (GE Healthcare Life Sciences) which had been equilibrated with lysis buffer supplemented to 50 mM imidazole pH 8.0, and the suspension was applied on Flex-Column (420400-2510) connected to Vac-Man vacuum manifold (Promega). Unbound proteins were washed out via controlled suction with 160 ml of lysis buffer (with 50 mM imidazole pH 8.0). Bound proteins were eluted with 15 ml of lysis buffer supplemented to 500 mM imidazole pH 8.0, followed by Tobacco Etch Virus (TEV) protease treatment at 1:25 protease:protein ratio. The solutions were left at 4 °C overnight. Size exclusion chromatography was performed on a Superdex 75 column equilibrated in lysis buffer. Fractions containing cut protein were collected and applied on a Flex-Column with 3 ml of Ni^{2+} Sepharose which had been equilibrated with lysis buffer. The flow through and a 7 ml lysis buffer rinse were collected. Lysis buffer was replaced using a 30 kDa MWCO filter (Amicon-Millipore) via 10X concentration/dilution repeated 3 times to crystallization buffer (20 mM HEPES

pH 7.5, 150 mM NaCl, 1 μM $ZnCl_2$, 10 mM DTT). The final concentrations of WT PLpro was 25 mg/mL and C111S mutant was 30 mg/mL.

Expression and purification of isotope labeled ISG15 and Ub

ISG15 was purified following the same protocol as for PLpro except that the buffers did not contain $ZnCl_2$ and a 10 kDa MWCO filter was used for buffer exchange and concentration. The final concentrations of ISG15 was 40 mg/mL.

Unlabeled Ub variants were expressed in BL12(DE3) *E.coli* cells containing a helper pJY2 plasmid and purified as described elsewhere[27]. For expression of uniformly ^{15}N-labeled Ubs and ISG15 the cells were grown in auto-inducing minimal medium containing $^{15}NH_4Cl$ as the sole source of nitrogen[28].

Synthesis of K48-linked Ub chains

Ub_2 and Ub_3 chains were assembled from the respective recombinant Ub monomers using controlled chain synthesis catalyzed by Ub-activating E1 enzyme UBA1 and K48-specific Ub-conjugating E2 enzyme UBE2K (aka E2-25K) as detailed elsewhere[27,29,30]. Specifically, Ub variants bearing chain-terminating mutations, Ub K48R (or Ub K48C) and Ub D77, were used to ensure that only Ub dimers are produced and to enable incorporation of ^{15}N labeled Ub units at the desired distal (^{15}N-Ub K48R) or proximal (^{15}N-Ub D77) position in the resulting chain for NMR and MS studies. The Ub_3 chain was made in a stepwise manner, by first removing D77 from the proximal Ub in Ub_2 by Ub C-terminal hydrolase YUH1 and subsequently conjugating this Ub_2 to Ub D77 using E1 and E2-25K to produce Ub_3. The resulting Ub_3 chain for cleavage assays had Ub K48R at the distal unit, ^{15}N-labeled Ub at the endo unit, and Ub D77 at the proximal unit, in order to allow unambiguous identification of the possible cleavage products by mass spectrometry. The correct mass of the synthesized Ub_2 and Ub_3 chains was confirmed using ESI-MS and SDS PAGE.

Microscale Thermophoresis binding assay

MST experiments were performed using NanoTemper Monolith NT.115 available in the Macromolecular Biophysics Resource core at UTSW and standard protocol was employed during analysis[31]. Ub_1, K48-Ub_2 or hISG15 were labeled with Cyanine5 NHS ester dye (Cy5) and titrated by a 1:1 serial dilution of wild-type or mutant PLpro-SARS-CoV-2. Obtained data were fit in PALMIST

1.5.8 in a 1:1 binding model and analyzed in GUSSI 1.4.2 [31]. All measurements were done in technical triplicates. The binding stoichiometry of the complexes was verified using cross-linking of complexes and visualized by SDS-PAGE. PALMIST and GUSSI software is freely available for academic users on UTSW MBR core's website (https://www.utsouthwestern.edu/labs/mbr/software/).

Crystallization of PLpro complexes

Crystallizations were performed with the protein-to-matrix ratio of 1:1 using the sitting drop vapor-diffusion method with the help of the Mosquito liquid dispenser (TTP LabTech) in 96-well CrystalQuick plates (Greiner Bio-One). MCSG1, MCSG2, MCSG3, and MCSG4 (Anatrace), Index (Hampton Research) and Wizard 1&2 (Jena Bioscience) screens were used at 16 °C. The PLpro-C111S/ISG15 complex (13 mg/ml) crystallized in Index E11 (0.02 M $MgCl_2$, 0.1 M HEPES pH 7.5, 22% (w/v) polyacrylic acid sodium salt 5,100). For the PLpro-C111S/Ub_2-K48R,D77 complex (11 mg/ml), crystals appeared in MCSG-2 F11 and were improved in hanging drops with a protein-to-matrix ratio of 2:1 in 0.2 M di-sodium tartrate, 15 % (w/v) PEG3350 after seeding with 1/10 volume of PLpro-C111S/Ub_2-K48R,D77 microcrystals. Crystals selected for data collection were washed in the crystallization buffer supplemented with either 25% glycerol (apo-protein crystals) or 25% ethylene glycol (protein-inhibitor crystals) and flash-cooled in liquid nitrogen.

PLpro cleavage assay

K48-linked Ub_3 and ISG15 cleavage reactions were performed at 20 °C in 20 mM Tris buffer (pH 7.52) containing 100 mM NaCl, 10 mM DTT, 1 μM $ZnCl_2$. The initial volume was 350 μL and contained 20 μM of Ub_3. Upon addition of 0.5 μM PLpro, equal amounts (10 μL) of reaction samples were aliquoted out at given time points, mixed with equal volume of SDS load buffer and immediately placed in a water bath at 70 °C for 5 minutes to stop the reaction. The samples were then loaded onto 15% urea polyacrylamide gel and resolved using SDS-PAGE.

Data collection, structure determination, and refinement

Single-wavelength x-ray diffraction data were collected at 100 K temperature at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source at Argonne National Laboratory using the program SBCcollect. The diffraction images were recorded on the PILATUS3

X 6M detector at 12.662 keV energy (0.9792 Å wavelength) using 0.3° rotation and 0.3 sec exposure. The intensities were integrated and scaled with the HKL3000 suite[32]. Intensities were converted to structure factor amplitudes in the truncate program[33] from the CCP4 package[34]. The structures were determined by molecular replacement using HKL3000 suite incorporating the program MOLREP[35-37]. The coordinates of SARS CoV-2 PLpro in complex with ubiquitin propargylamide (PDB id: 6XAA) and PLpro-C111S with mISG15 (PDB id: 6YVA) were used as the starting models for PLpro-K48-linked Ub_2 and PLpro-ISG15 structure solutions, respectively. The initial solutions were refined, both rigid-body refinement and regular restrained refinement by REFMAC program[38] as a part of HKL3000. Several rounds of manual adjustments of structure models using COOT[39] and refinements with REFMAC program[38] from CCP4 suite[34] were done. The stereochemistry of the structure was validated with PHENIX suite[40] incorporating MOLPROBITY[41] tools. A summary of data collection and refinement statistics is given in Supplementary Table 1.

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 accession codes: 7RBR for PLpro SARS CoV-2, C111S mutant, in complex with a K48-Ub2, 7RBS for PLpro SARS CoV-2, C111S mutant, in complex with a hISG15, 7S6P for human ISG15 alone and 7S6O for K48-linked Ub_2 alone. Diffraction images are available on server in Dr. W. Minor laboratory https://proteindiffraction.org. Plasmid for expression PLpro 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 raw crosslinking mass spectrometry data is available in Data Source 1. All other data generated during the current study including the raw kinetic and biophysical data are available upon request.

NMR data collection and analysis

NMR measurements were performed at 25 °C on Avance III 600 MHz and 800 MHz Bruker NMR spectrometers equipped with cryoprobes. The data were processed using Topspin (Bruker) and analyzed using Sparky 3.190[42]. NMR signal assignments for ISG15 were obtained from Biological Magnetic Resonance Data Bank (BMRB Entry ID 5658) and adjusted to the temperature and buffer conditions used in our studies.

The protein samples for NMR measurements were prepared in 50 mM HEPES buffer or in 20 mM Tris buffer, both at pH 7.42 and containing 100 mM NaCl, 1 mM TCEP, 1 μM $ZnCl_2$, 0.2% (w/v) NaN_3, and 10% (v/v) D_2O. Binding studies by NMR were carried out by adding pre-calculated amounts of unlabeled PLpro (C111S) to ^{15}N-labeled ISG15 up to 1:1 molar ratio, or at 2:1 molar ratio to Ub_1 or K48-Ub_2 (with either distal or proximal Ub ^{15}N labeled) and monitoring changes in 2D ^{1}H-^{15}N SOFAST-HMQC and ^{1}H-^{15}N TROSY spectra. The initial binding studies were performed in HEPES buffer and subsequently repeated in Tris buffer, both produced similar results.

The protein concentrations in NMR studies in Tris buffer were as follows: 150:150 μM for ISG15:PLpro; 88:177 μM for ^{15}N-distal Ub_2:PLpro; 62.5:125 μM for ^{15}N-proximal Ub_2:PLpro; 125:250 μM for Ub_1:PLpro. For NMR measurements in HEPES buffer, the concentrations were 115:115 μM for ISG15:PLpro; 71:142 μM for ^{15}N-distal Ub_2:PLpro; 75:150 μM for ^{15}N-proximal Ub_2:PLpro; and 81:151 μM for Ub_1:PLpro.

Cross-linking mass spectrometry analysis

Our group have developed standardized protocols for cross-linking and data analysis of samples. For complexes between $PLpro_{CoV-2}$ and K48-linked Ub_2 or hISG15, we incubated the protease with the substrate at a 1:4 molar ratio for 1 hour at 25 °C. For DSS reactions, samples were cross-linked with a final 1mM DSS (DSS-d_0 and -d_{12}, Creative Molecules) for three minutes at 37 °C while shaking at 350 rpm. For SuFEx reactions, samples were cross-linked with a final 0.658 mM SuFEx (a kind gift from Dr. William DeGrado, UCSF) for one hour at 37 °C while shaking at 350 rpm. For DMTMM reactions, samples were cross-linked with a final 43 mM DMTMM (Sigma-Aldrich) for 15 minutes at 37 °C while shaking at 600 rpm. All cross-linking reactions were quenched with 172 mM (4 times excess) Ammonium Bicarbonate (AB) for 30 minutes at 37 °C while shaking at 350 rpm. Samples were resolved on SDS-PAGE gels (NuPAGE™ , 4 to 12%, Bis-Tris, 1.5 mm) and, for DMTMM cross-linker set of samples, bands corresponding to $PLpro_{CoV-2}$:substrate heterodimers were gel-extracted following standard protocols[18]. Samples were digested by 1:50 (m/m) trypsin (Promega) overnight at

37 °C while shaking at 600 rpm. 2% (v/v) formic acid was added to acidify the reaction and further purified by reversed-phase Sep-Pak tC18 cartridges (Waters), next flash frozen in liquid nitrogen and lyophilized. The dried samples were resuspended in water/acetonitrile/formic acid (95:5:0.1, v/v/v) to a final concentration of approximately 0.5 μg/μL. 2 μL of each was injected into Eksigent 1D-NanoLC-Ultra HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system at the UTSW Proteomics core.

The mass spectrometry data was analyzed by in-house version of xQuest 2.1.5 pipeline[43]. Thermo RAW data files were first converted to open .mzXML format using msconvert (proteowizard.sourceforge.net). The mass spectra across replicates yielded similar intensities (Supplementary Data 1). Search parameters were set based on DMTMM as the cross-link reagent as follows: maximum number of missed cleavages = 2, peptide length = 5-50 residues, fixed modifications = carbamidomethyl-Cys (mass shift = 57.02146 Da), variable modification = oxidation of methionine (mass shift = 15.99491 Da), mass shift of cross-linker = -18.010595 Da, no monolink mass specified, MS1 tolerance = 15 ppm, and MS2 tolerance = 0.2 Da for common ions and 0.3 Da for cross-link ions; search in enumeration mode. Next, in-house shell script was employed to identify cross-links between lysines and acidic residues, only configuration possible while using DMTMM cross-linker. FDRs were estimated by xProphet[17] to be 0% - 53.8%. For each experiment, five replicate data sets were compared and only cross-link pairs that appeared in all data sets ($PLpro_{CoV-2}$:hISG15) or at least in four data sets ($PLpro_{CoV-2}$:K48-linked Ub_2) were used to generate a consensus data set (Supplementary Data 1).

Energetic analysis of PLpro in complex with hISG15 and K48-Ub_2

The four models of PLpro (SARS-CoV-1 and SARS-CoV-2) bound two different substrates, hISG15 and K48-linked Ub_2, that were used in the subsequent *in silico* alanine scan were prepared as follows. For the $PLpro_{CoV-2}$:hISG15 complex, we used a heterodimer conformation derived from the crystal structure deposited in the protein data bank (PDB ID: 7RBS). To create the complex between $PLpro_{CoV-2}$ and K48-linked Ub_2, we used the conformation and binding mode of K48-linked Ub_2 bound to $PLpro_{CoV-1}$ (PDB ID: 5E6J) as a template[16]. Briefly, $PLpro_{CoV-2}$ was aligned to $PLpro_{CoV-1}$ bound to K48-linked Ub_2 to produce a tentative model of $PLpro_{CoV-2}$:K48-linked Ub_2. For $PLpro_{CoV-1}$:K48-linked Ub_2 complex, we used a heterodimer conformation derived from the crystal structure[16] deposited in the protein data bank (PDB ID: 5E6J), and to create the $PLpro_{CoV-1}$:hISG15 complex, we

employed the same procedure as used to make the $PLpro_{CoV-2}$:K48-linked Ub_2 complex (see above). Next, we applied a relax protocol in Rosetta all four complexes: $PLpro_{CoV-1}$:K48-linked Ub_2, $PLpro_{CoV-2}$:K48-linked Ub_2, $PLpro_{CoV-1}$:hISG15 and $PLpro_{CoV-2}$:hISG15. To guarantee that each instance of relax is being run with different randomizations, groups of nstruct were run with different, randomly generated seeds using random.org. From 100 total structures (4x25 nstruct for computational efficiency) for each heterocomplex, the lowest energy structure was identified and used in further steps. The list of PLpro residues that may be engaged in interacting with its substrate was created by identifying PLpro residues in proximity of 4.0 Å of either hISG15 or K48-linked Ub_2. The union list of interacting residues identified with heterocomplexes PLpro:hISG15 and PLpro:K48-linked Ub_2 was used in the next step. For $PLpro_{CoV-2}$ in complex with hISG15/K48-linked Ub_2 49 positions were used to describe the combined interface. For $PLpro_{CoV-1}$ in complex with hISG15/K48-linked Ub_2 47 positions were selected. Selection of common interface residues was carried out in PyMOL 2.4.1. Flex ΔΔG protocol was used as described previously [24]. The code is available on GitHub: https://github.com/Kortemme-Lab. Briefly, selected interacting residues were mutated to alanines. Alternatively, mutations were made to residues from PLpro from either SARS-CoV-1 or SARS-CoV-2. Parameters were used (all default settings): nstruct = 35, max_minimization_iter = 5000, abs_score_convergence thresh = 1.0, number_backrub_trials = 35000, and to enable earlier time points backrub_trajectory_stride was set to 7000. The $\Delta\Delta G_{binding}$ score for the last iteration was shown in the Results. These simulations were performed using BioHPC computing cluster at UT Southwestern Medical Center. The results, in raw REU (Rosetta energy units), are shown as a heat-map with $\Delta\Delta G_{binding}$ values but also as the difference between the $\Delta\Delta G_{binding}$ from hISG15 compared to K48-linked Ub_2. The plots were made using GraphPad Prism 8 and mapped onto the protease structure using PyMOL 2.4.1. The relax protocol and Flex_ΔΔG used Rosetta v3.13 and v3.12, respectively.

Evolutionary analysis of PLpro sequences

BAT_RaTG13 PLpro sequence was retrieved from ncbi under Genbank ID: MN996532.2. SARS-CoV-1 PLpro protein sequence was derived from 5edj.pdb. *Pteropus alecto* ISG15 sequence was retrieved from ncbi under Genbank ID: 102878693 and was used to produce sequence alignments to human ISG15 (Genbank ID: 9636) and mouse ISG15 (Genbank ID: 100038882) but also across homologous sequences retrieved from blastp [44]. Alignments were produced in clustalo[45] and

visualized in seaview[46]. Protein sequence conservation between PLpro from SARS-CoV-2, SARS-CoV-1 and BAT_RaTG13 was calculated using AL2CO[47]. The per position conservation was mapped onto the $PLpro_{CoV-2}$ structure in the context of a bound K48-linked Ub_2 or hISG15 using PyMOL. SARS-CoV-2 PLpro sequence variation from 1.1 million sequences was derived from the coronavirus3D database [48]. The per residue mutational frequencies were mapped onto a $PLpro_{CoV-2}$ structure in the context of a bound K48-linked Ub_2 or hISG15.

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 for help with MST data analysis. We also appreciate the help of the Proteomics 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 on instruments supported in part by NSF grant DBI1040158. EE was supported by Undergraduate Maryland Summer Scholarship. L.A.J. is supported by an Effie Marie Cain Scholarship in Medical Research. PMW is supported by a O'Donnell Brain Institute Pilot grant.

AUTHOR CONTRIBUTIONS

AJ 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 Ub chains, expressed and purified proteins for NMR studies and carried out NMR studies. BL 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.

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. 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).
4. 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).
5. 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).
6. Sulea, T., Lindner, H. A., Purisima, E. O. & Ménard, R. Binding site-based classification of coronaviral papain-like proteases. *Proteins* **62,** 760–775 (2006).
7. Snyder, N. A. & Silva, G. M. Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response. *J. Biol. Chem.* **297,** 101077 (2021).
8. Swaim, C. D. *et al*. Modulation of Extracellular ISG15 Signaling by Pathogens and Viral Effector Proteins. *CellReports* **31,** 107772 (2020).
9. 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).
10. Reynolds, N. D. *et al*. The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect Dis* **7,** 1483–1502 (2021).
11. 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).
12. 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).
13. Shin, D. *et al*. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587,** 657–662 (2020).

14. 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).
15. Klemm, T. *et al*. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J* **39,** e106275 (2020).
16. 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).
17. Walzthoeni, T. *et al*. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Meth* **9,** 901–903 (2012).
18. Chen, D. *et al*. Tau local structure shields an amyloid-forming motif and controls aggregation propensity. *Nature Communications* **10,** 2493–14 (2019).
19. Mirbaha, H. *et al*. Inert and seed-competent tau monomers suggest structural origins of aggregation. *eLife* **7,** (2018).
20. Sali, A. *et al*. Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure/Folding and Design* **23,** 1156–1167 (2015).
21. Ryder, B. D. *et al*. Regulatory inter-domain interactions influence Hsp70 recruitment to the DnaJB8 chaperone. *Nature Communications* **12,** 946–16 (2021).
22. 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).
23. 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).
24. 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).
25. Karanicolas, J. *et al*. A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Molecular Cell* **42,** 250–260 (2011).
26. Wacharapluesadee, S. *et al*. Longitudinal study of age-specific pattern of coronavirus infection in Lyle's flying fox (Pteropus lylei) in Thailand. *Virol J* **15,** 38–10 (2018).
27. Pickart, C. M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Meth. Enzymol.* **399,** 21–36 (2005).
28. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41,** 207–234 (2005).
29. Varadan, R., Assfalg, M., Raasi, S., Pickart, C. & Fushman, D. Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Molecular Cell* **18,** 687–698 (2005).
30. Varadan, R., Walker, O., Pickart, C. & Fushman, D. Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology* **324,** 637–647 (2002).
31. 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).
32. 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).
33. Padilla, J. E. & Yeates, T. O. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr D Biol Crystallogr* **59,** 1124–1130 (2003).

34. Winn, M. D. *et al*. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67,** 235–242 (2011).
35. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Meth. Enzymol*. **374,** 229–244 (2003).
36. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11,** 49–52 (2004).
37. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66,** 22–25 (2010).
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. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60,** 2126–2132 (2004).
40. Adams, P. D. *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).
41. Chen, V. B. *et al*. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66,** 12–21 (2010).
42. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327 (2015).
43. Rinner, O. *et al*. Identification of cross-linked peptides from large sequence databases. *Nat Meth* **5,** 315–318 (2008).
44. Madden, T. L., Tatusov, R. L. & Zhang, J. Applications of network BLAST server. *Meth. Enzymol*. **266,** 131–141 (1996).
45. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539 (2011).
46. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27,** 221–224 (2010).
47. Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17,** 700–712 (2001).
48. Sedova, M., Jaroszewski, L., Alisoltani, A. & Godzik, A. Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. *Bioinformatics* **36,** 4360–4362 (2020).

FIGURE LEGENDS

Figure 1. PLpro architecture and substrate recognition. (a) Binding mode of K48-linked Ub_2 to $PLpro_{CoV-1}$ and mISG15 to $PLpro_{CoV-2}$. PLpro is shown in surface representation and colored grey. K48-linked Ub_2 is shown in cartoon representation and colored white and pink for the proximal and distal Ubs, respectively. We call 'distal' the Ub unit connected through its C-terminus to K48 of the other Ub unit, called 'proximal'. mISG15 is shown in cartoon representation and colored magenta. In analogy with the two Ubs in Ub_2, the N- and C-terminal UBL (ubiquitin like) domains of ISG15 will be referred to as the distal and proximal UBLs, respectively. (b) Sequence comparison of Ub, hISG15 and mISG15 nonpolar patch important for UBL recognition. Ub and ISG15 are shown in cartoon representation and are colored pink and magenta. Key residues that define the nonpolar patch on Ub and ISG15 (proximal domain) are shown in spacefill and highlighted with arrows. (c) Topology of K48-linked Ub_2 and mISG15 (bottom). LRGG sequence is shown in the C-term. (d) MST binding analysis of substrates to $PLpro_{CoV-2}$. Comparison of $PLpro_{CoV-2}$ binding to hISG15 (magenta, K_d=0.8 μM), Ub_2 (grey, K_d=35 μM) and Ub_1 (pink, K_d=120 μM). (e) $PLpro_{CoV-2}$ cleavage of hISG15-Nsp C-terminal fusions. SDS-PAGE gel reveals that PLpro cleaves off fragments of Nsp2 (AYTRYVDNNF), Nsp3 (APTKVTFGDD), and Nsp4 (KIVNNWLKQL) from hISG15-Nsp fusions linked to the recognition motif "LRGG". (f) $PLpro_{CoV-2}$ cleavage of Ub_3. SDS-PAGE gel reveals that PLpro cleaves Ub_3 into Ub_2 and Ub_1 efficiently.

Figure 2. MX structure of C111S $PLpro_{Cov-2}$ bound to human ISG15 and K48-linked 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. hISG15 is shown in magenta. Proximal and distal Ub domains are pink and white, respectively. Crystal structure of PLpro bound to (a, right) hISG15 and (b, right) K48-linked Ub_2. PLpro is shown in cartoon representation, hISG15 and K48-linked Ub_2 are shown in sticks. PLpro, hISG15 and K48-linked Ub_2 colored as in (a) and (b). The electron density map for PLpro substrates is contoured at 2.0s. (c) Schematic of unbound hISG15 (top) and K48-linked Ub_2 (bottom). PLpro substrates colored as in (a) and (b). (d) Overlay of crystal structures of bound hISG15 (magenta, PDB id 7rbs, this study) and unbound hISG15 (blue, PDB id 7s6p, this study). (e) Overlay of bound conformation of K48-linked Ub_2 as observed with $PLpro_{CoV-1}$ structure (PDB id 5e6j) with unbound conformation of K48-linked Ub_2 (PDB id 7s6o, this study). Proximal

domain of both bound and unbound conformations is shown in pink, distal domain of bound conformation is shown in white, and distal domain of unbound conformation is shown in blue. (f) Overlay of bound conformation of K48-linked Ub_2 as observed with $PLpro_{CoV-1}$ structure (PDB id 5e6j) with unbound "open" conformation of Ub_2 (PDB id 3ns8). Proximal domains of bound and unbound conformations are shown in pink and blue, respectively. Distal domain is shown in white. (g) Overlay of bound conformation of K48-linked Ub_2 as observed with $PLpro_{CoV-1}$ structure (PDB id 5e6j) with unbound "closed" conformation of Ub_2 (PDB id 1aar). Proximal domain of both bound and unbound conformations is shown in pink, distal domain of bound conformation is shown in white, and distal domain of unbound conformation is shown in blue. Self-self interaction surface shown with an arrow.

Figure 3. Comparison of hISG15 and Ub_1 or hISG15 and mISG15 bound to $PLpro_{CoV-2}$. (a) Structural overlay of $PLpro_{CoV-2}$:hISG15 and $PLpro_{CoV-2}$:Ub_1. Proteins are shown in cartoon and colored grey (PLpro), magenta (hISG15) and pink (Ub_1). (b) Zoom in of the boxed area in (a) representing overlay of the proximal domain of hISG15 and Ub_1. Proximal domain of hISG15 and mono are represented as in (a). Rotation of the active surface indicated with arrows. (c,d) Comparison of the active surfaces of hISG15 (c) and Ub_1 (d). Proximal domain of hISG15 and Ub_1 are represented as in (a). Key residues are shown in spacefill. (e) Structural overlay of $PLpro_{CoV-2}$:hISG15 and $PLpro_{CoV-}$2:mISG15. Proteins are shown in cartoon and colored grey (PLpro), magenta (hISG15) and pink (mISG15). (f,g) Zoom in of the boxed area in (e). Comparison of the distal domains of hISG15 (f) and mISG15 (g). hISG15 and mISG15 are represented as in (e). Key residues are shown in spacefill.

Figure 4. NMR data showing PLpro interaction with ISG15 and K48-Ub_2. (a) Overlay of ^{1}H-^{15}N TROSY spectra of ^{15}N-hISG15 alone (blue) and with equimolar unlabeled $PLpro_{Cov-2}$ (C111S) (red); (b) hISG15 residues with strong signal perturbations mapped (yellow) on our structure of $PLpro_{CoV-2}$:hISG15 complex. (c-e) Overlay of fragments of ^{1}H-^{15}N HMQC spectra of the proximal (c) and distal (d) Ub units of K48-Ub_2 and (e) of Ub_1, alone (blue) and upon addition of 2x PLpro (red). Insets zoom on signals of Ub nonpolar patch residues L8, I44, H68, V70. (f) Residues in the proximal Ub that showed signal perturbations in (c) mapped (yellow) on our structure of $PLpro_{CoV-2}$:K48-Ub_2 complex.

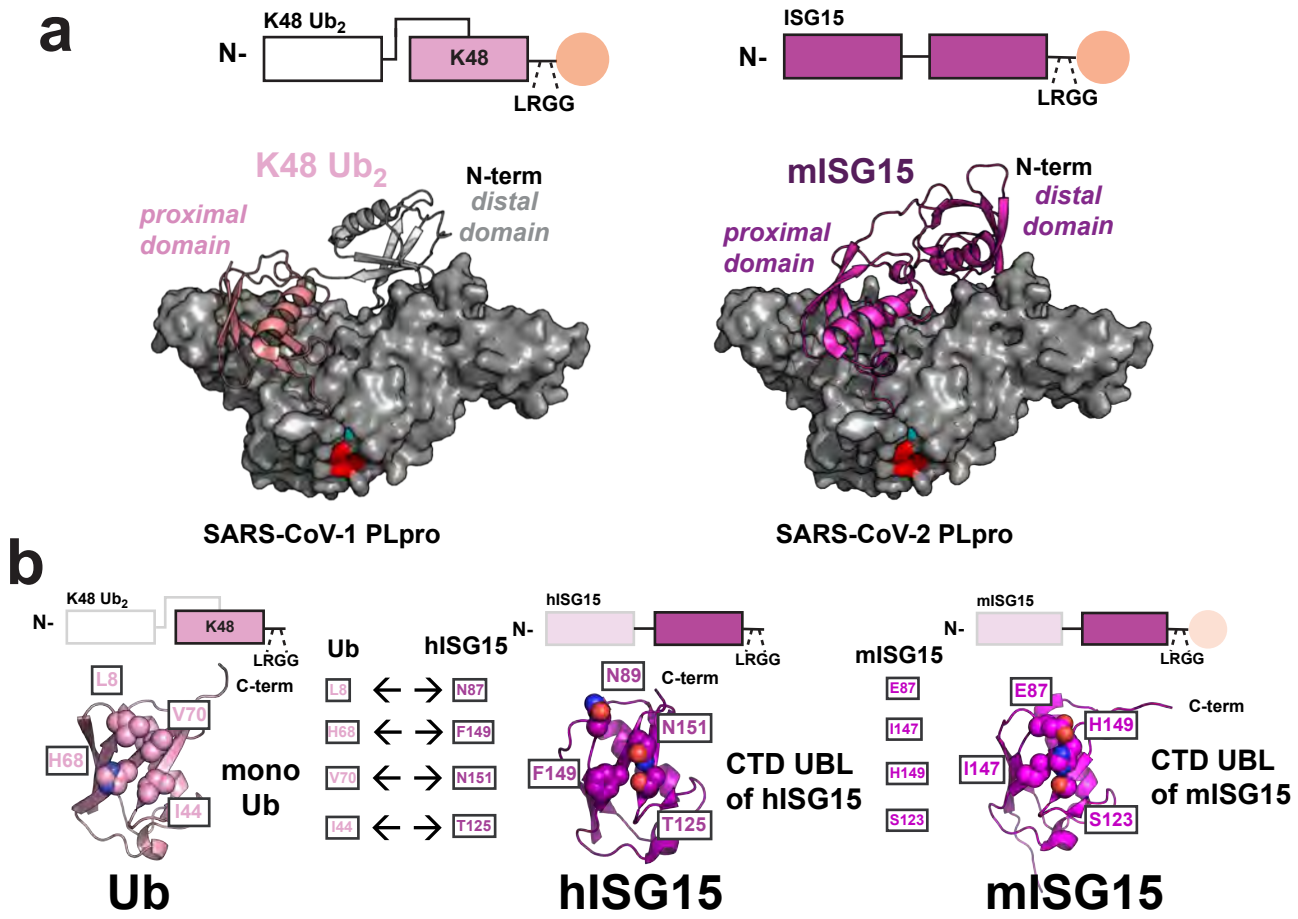
Figure 5. XL-MS analysis of PLpro$_{CoV-2}$ complexes. (a) Frequency comparison of replicate datasets for heterodimer complex. Frequencies of intermolecular cross-links found between D61 and D62 on PLpro$_{CoV-2}$ to K35 on the distal domain of ISG15 are shown in magenta. Frequencies of cross-links found between K45 on PLpro$_{CoV-2}$ to D58 on Ub and between D164 on PLpro$_{CoV-2}$ to K6 on Ub are shown in pink. (b) Interactions between PLpro$_{CoV-2}$ (D61/D62) and the distal ISG15 domain (K35) identified by XL-MS mapped on the structure of heterocomplex. PLpro and hISG15 are shown in cartoon representation. PLpro is shown in grey and hISG15 is shown in magenta. (c) All intramolecular cross-links in PLpro$_{CoV-2}$ from PLpro$_{CoV-2}$:hISG15 dataset. Cross-links between UBL domain and the rest of PLpro are colored red. The other cross-links are colored blue. PLpro is represented as in (b). (d) All intramolecular cross-links in PLpro$_{CoV-2}$ from PLpro$_{CoV-2}$:K48-Ub$_2$ dataset. The cross-link between UBL domain and the fingers domain of PLpro is colored red. The other cross-link (in fingers domain) is colored blue. PLpro is represented as in (b).

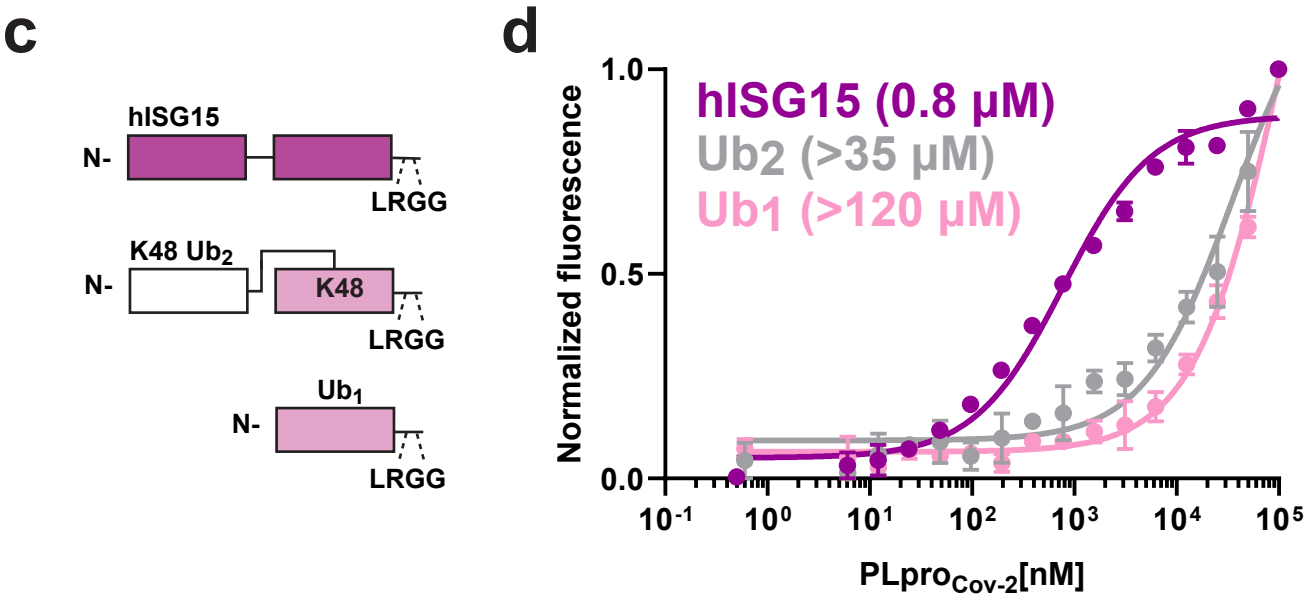
Figure 6. Energetic analysis of PLpro in complex with hISG15 and K48-Ub$_2$. (a,b) Schematic illustration of identification of substrate binding surfaces on PLpro and *in silico* mutagenesis for heterodimer complex of PLpro$_{CoV-2}$ with hISG15 and K48-linked Ub$_2$ (a), or PLpro$_{CoV-1}$ with hISG15 and K48-linked Ub$_2$ (b). (c) Heat-map results of $\Delta\Delta G$ calculations of *in silico* alanine scan for PLpro$_{CoV-2}$ in complex with K48-linked Ub$_2$ or hISG15. 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$_{CoV-2}$:K48-linked Ub$_2$ and PLpro$_{CoV-2}$:hISG15. (d) Heat-map results of $\Delta\Delta G$ calculations of *in silico* alanine scan for PLpro$_{CoV-1}$ in complex with K48-linked Ub$_2$ or hISG15. 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$_{CoV-1}$:hISG15 and PLpro$_{CoV-1}$:K48-linked Ub$_2$. (e) Results of *in silico* mutagenesis for heterodimer complex of PLpro$_{CoV-2}$ with K48-linked Ub$_2$ and hISG15 mapped on surface representation of PLpro$_{CoV-2}$. Hotspot sites identified as driving stability towards K48-linked Ub$_2$ are colored in blue, and hotspot sites identified as driving stability towards hISG15 are colored in red. (f) Results of *in silico* mutagenesis for heterodimer complex of PLpro$_{CoV-1}$ with hISG15 and K48-linked Ub$_2$ mapped on surface representation of PLpro$_{CoV-1}$. Hotspot sites represented as in (e). (g) *In silico* mutagenesis hotspots shown in the context of proximal and distal domains binding sites and interaction surface residues

that vary between SARS-CoV-2 and SARS-CoV-1. PLpro is shown in cartoon representation and in grey. Hotspot residues are shown as red spheres ($\Delta\Delta G_{binding} > 3$), intermediate contribution residues are shown as orange spheres ($\Delta\Delta G_{binding} > 1.5 < 3$), and residues that vary between SARS-CoV-1 and SARS-CoV-2 are shown as blue spheres.

Figure 7. Sequence variation on PLpro underlies changes in substrate selectivity. (a) Comparison of $PLpro_{CoV-2}$ sequence to Bat_RATG13 reveals high conservation with two mutations in the substrate binding region. (b) Comparison of $PLpro_{CoV-1}$ and $PLpro_{CoV-2}$ reveals variation at the substrate binding surface. (c) Sequence variation across 1.1 million $PLpro_{CoV-2}$ isolates reveals variation in both the proximal and distal binding sites. PLpro is shown in space fill representation and is color-coded according to conservation from yellow-red or white-red. Sites of variation are labeled. K48-linked Ub_2 and human ISG15 binding modes are shown for comparison.

Figure 1


a
K48 Ub2
N-
K48
LRGG
ISG15
N-
LRGG
K48 Ub2
proximal domain
N-term
distal domain
mISG15
proximal domain
N-term
distal domain
SARS-CoV-1 PLpro
SARS-CoV-2 PLpro
b
K48 Ub2
N-
K48
LRGG
L8
C-term
V70
H68
I44
mono Ub
Ub
hISG15
L8
N87
H68
F149
V70
N151
I44
T125
hISG15
N-
LRGG
N89
C-term
N151
F149
T125
CTD UBL of hISG15
mISG15
N-
LRGG
mISG15
E87
I147
H149
S123
E87
H149
C-term
I147
S123
CTD UBL of mISG15
Ub
hISG15
mISG15


c
hISG15
N-
LRGG
K48 Ub2
N-
K48
LRGG
Ub1
N-
LRGG
d
hISG15 (0.8 μM)
Ub2 (>35 μM)
Ub1 (>120 μM)
Normalized fluorescence
1.0
0.5
0.0
10-1
100
101
102
103
104
105
PLproCov-2[nM]

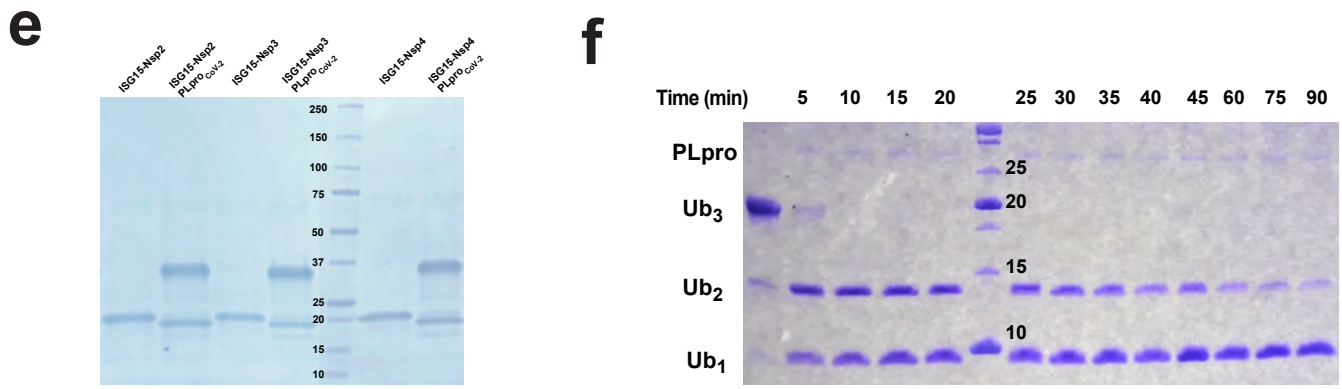

e
ISG15-Nsp2
ISG15-Nsp2 PLpro CoV-2
ISG15-Nsp3
ISG15-Nsp3 PLpro CoV-2
ISG15-Nsp4
ISG15-Nsp4 PLpro CoV-2
250
150
100
75
50
37
25
20
15
10
f
Time (min)
5
10
15
20
25
30
35
40
45
60
75
90
PLpro
Ub3
Ub2
Ub1
25
20
15
10

Figure 2

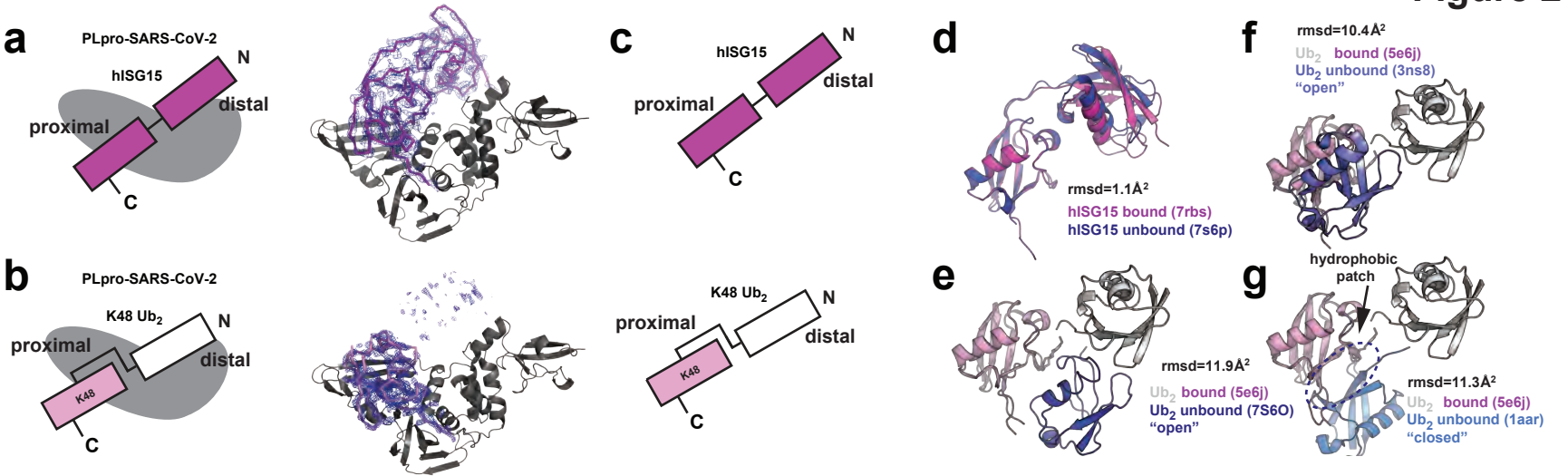

a
PLpro-SARS-CoV-2
hISG15
N
distal
proximal
C
b
PLpro-SARS-CoV-2
K48 Ub2
N
proximal
distal
K48
C
c
hISG15
N
distal
proximal
C
K48 Ub2
N
proximal
distal
K48
C
d
rmsd=1.1Å2
hISG15 bound (7rbs)
hISG15 unbound (7s6p)
e
rmsd=11.9Å2
Ub2 bound (5e6j)
Ub2 unbound (7S6O)
"open"
f
rmsd=10.4Å2
Ub2 bound (5e6j)
Ub2 unbound (3ns8)
"open"
g
hydrophobic patch
rmsd=11.3Å2
Ub2 bound (5e6j)
Ub2 unbound (1aar)
"closed"

Figure 3

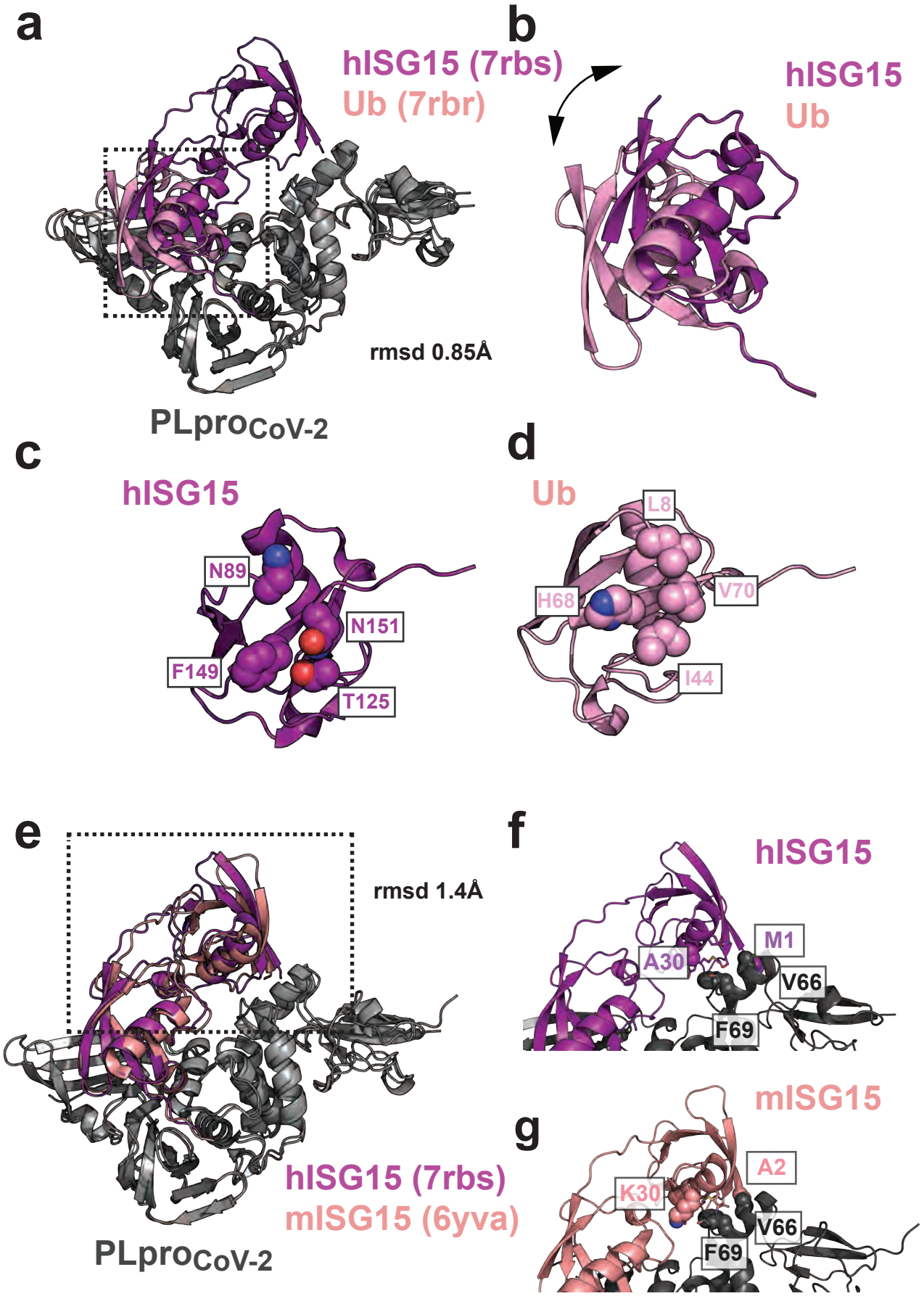

a
hISG15 (7rbs)
Ub (7rbr)
rmsd 0.85Å
PLproCoV-2
b
hISG15
Ub
c
hISG15
N89
N151
F149
T125
d
Ub
L8
H68
V70
I44
e
rmsd 1.4Å
hISG15 (7rbs)
mISG15 (6yva)
PLproCoV-2
f
hISG15
M1
A30
V66
F69
g
mISG15
A2
K30
V66
F69

Figure 4

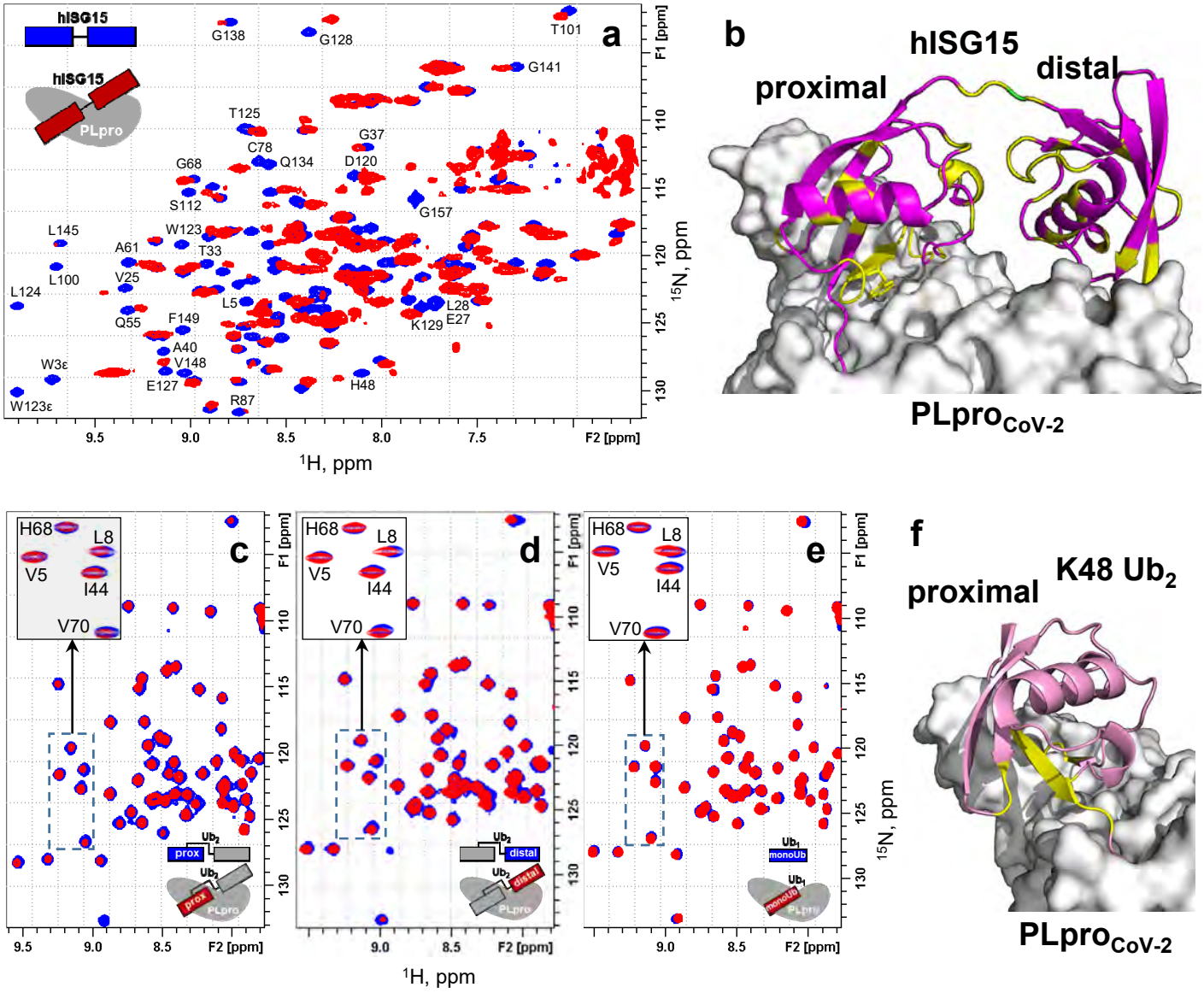

a
hISG15
PLpro
G138
G128
T101
G141
T125
G37
C78
Q134
D120
G68
S112
G157
L145
W123
A61
T33
L100
V25
L124
L5
L28
E27
K129
Q55
F149
A40
V148
W3ε
E127
H48
R87
W123ε
F1 [ppm]
F2 [ppm]
^{15}N, ppm
^{1}H, ppm
b
hISG15
proximal
distal
PLpro$_{CoV-2}$
c
d
e
H68
L8
V5
I44
V70
prox
distal
monoUb
Ub$_2$
Ub$_1$
f
proximal
K48 Ub$_2$
PLpro$_{CoV-2}$

Figure 5

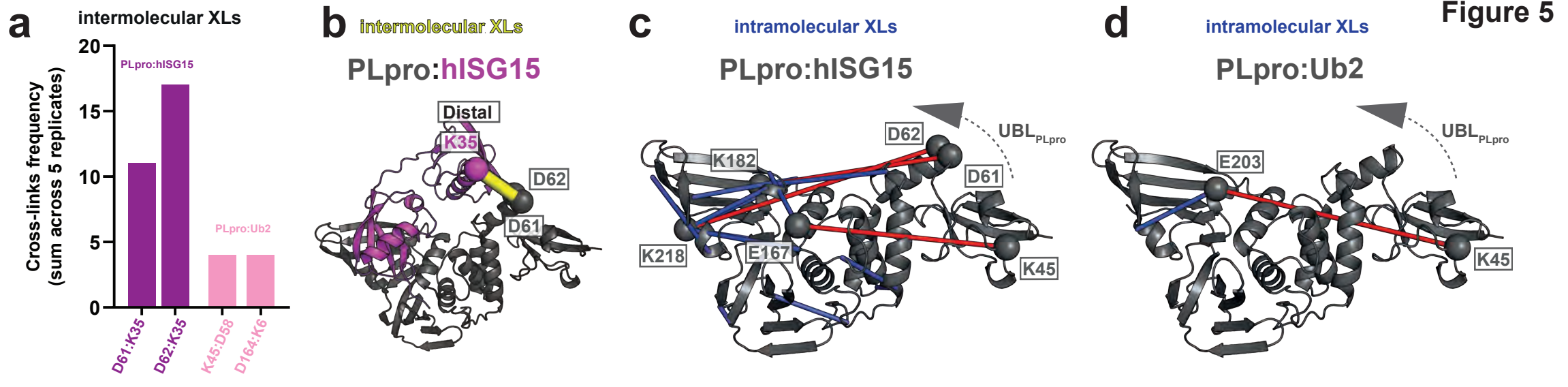

a
intermolecular XLs
Cross-links frequency
(sum across 5 replicates)
20
15
10
5
0
PLpro:hISG15
PLpro:Ub2
D61:K35
D62:K35
K45:D58
D164:K6
b
intermolecular XLs
PLpro:hISG15
Distal
K35
D62
D61
c
intramolecular XLs
PLpro:hISG15
D62
UBL$_{PLpro}$
K182
D61
K218
E167
K45
d
intramolecular XLs
PLpro:Ub2
UBL$_{PLpro}$
E203
K45

Figure 6

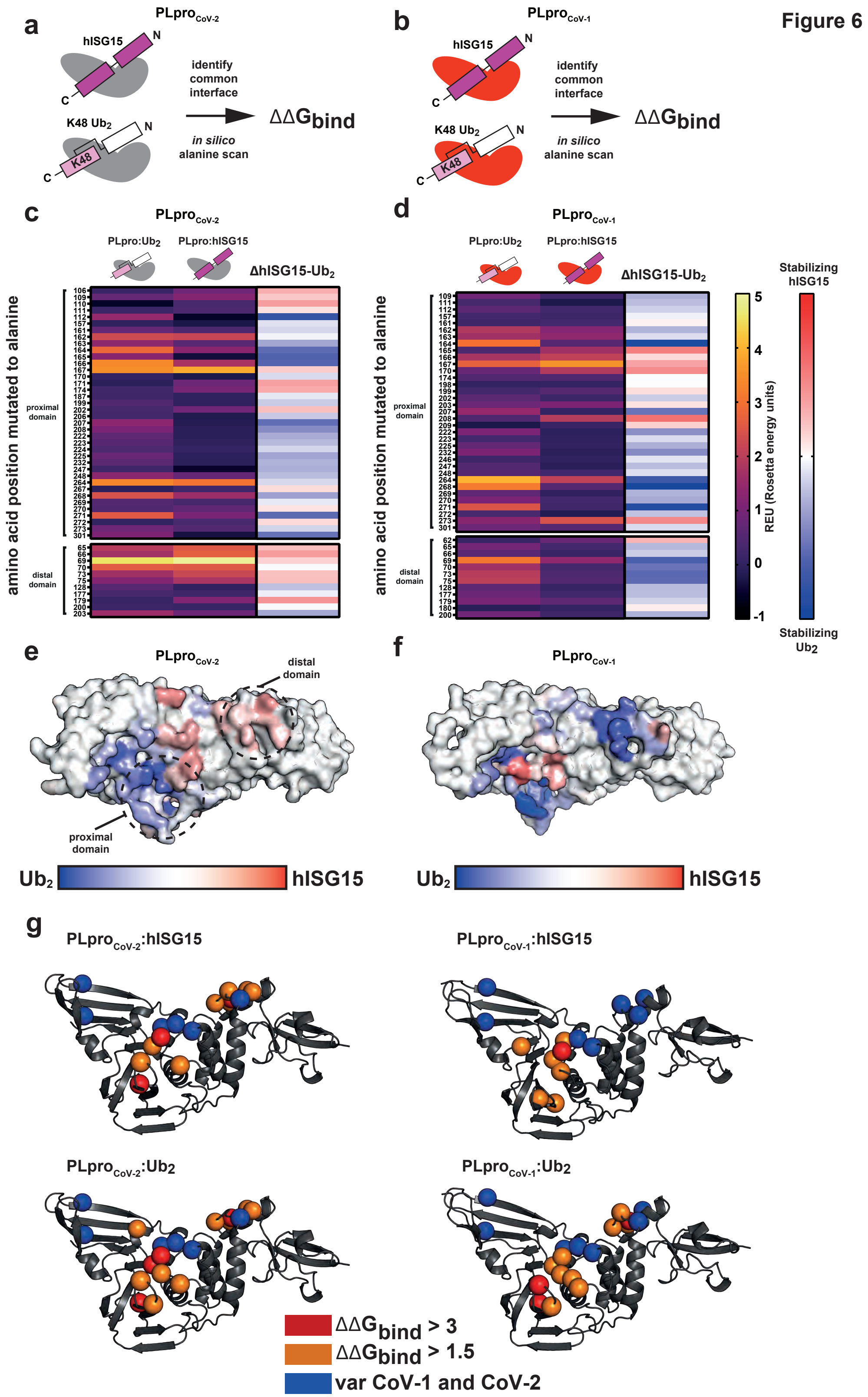
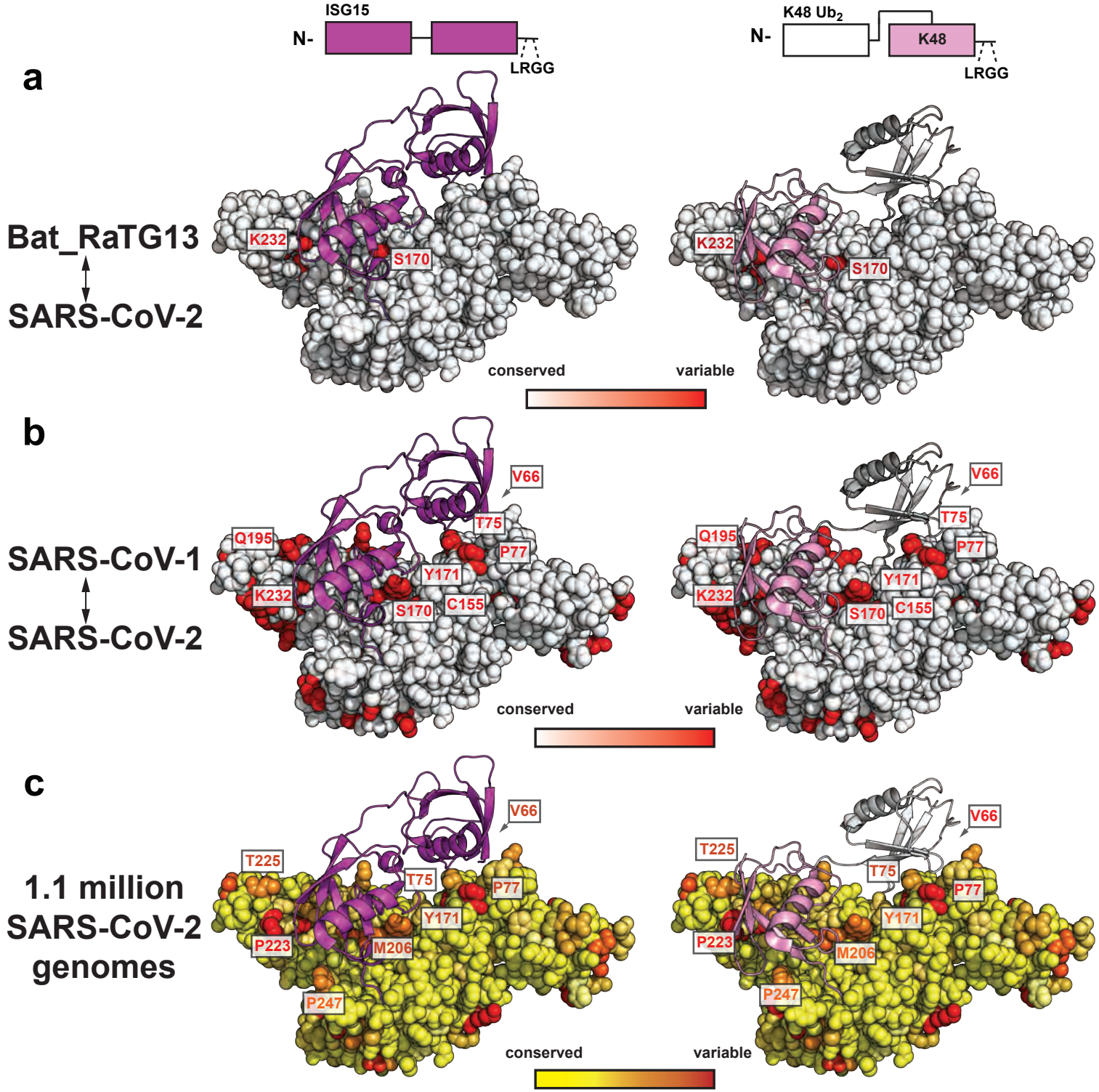

a
PLpro$_{CoV-2}$
hISG15
N
C
K48 Ub$_2$
N
K48
C
identify common interface
in silico alanine scan
ΔΔG$_{bind}$
b
PLpro$_{CoV-1}$
hISG15
N
C
K48 Ub$_2$
N
K48
C
identify common interface
in silico alanine scan
ΔΔG$_{bind}$
c
PLpro$_{CoV-2}$
PLpro:Ub$_2$
PLpro:hISG15
ΔhISG15-Ub$_2$
amino acid position mutated to alanine
proximal domain
distal domain
d
PLpro$_{CoV-1}$
PLpro:Ub$_2$
PLpro:hISG15
ΔhISG15-Ub$_2$
amino acid position mutated to alanine
proximal domain
distal domain
REU (Rosetta energy units)
Stabilizing hISG15
Stabilizing Ub$_2$
e
PLpro$_{CoV-2}$
distal domain
proximal domain
Ub$_2$
hISG15
f
PLpro$_{CoV-1}$
Ub$_2$
hISG15
g
PLpro$_{CoV-2}$:hISG15
PLpro$_{CoV-1}$:hISG15
PLpro$_{CoV-2}$:Ub$_2$
PLpro$_{CoV-1}$:Ub$_2$
ΔΔG$_{bind}$ > 3
ΔΔG$_{bind}$ > 1.5
var CoV-1 and CoV-2

Figure 7


ISG15
N-
LRGG
K48 Ub2
N-
K48
LRGG
a
Bat_RaTG13
SARS-CoV-2
K232
S170
K232
S170
conserved
variable
b
SARS-CoV-1
SARS-CoV-2
V66
T75
Q195
P77
Y171
K232
S170
C155
V66
T75
Q195
P77
Y171
K232
S170
C155
conserved
variable
c
1.1 million
SARS-CoV-2
genomes
V66
T225
T75
P77
Y171
P223
M206
P247
V66
T225
T75
P77
Y171
P223
M206
P247
conserved
variable