SARS-CoV-2 structural coverage map reveals state changes that disrupt host immunity

Seán I. O'Donoghue[1,2,3,8,9], Andrea Schafferhans[4,5], Neblina Sikta[1], Christian Stolte[1], Sandeep Kaur[1,3], Bosco K. Ho[1], Stuart Anderson[2], James Procter[6], Christian Dallago[5], Nicola Bordin[7], Matt Adcock[2], Burkhard Rost[5]

1: Garvan Institute of Medical Research, Sydney, Australia. 2: CSIRO Data61, Sydney, Australia. 3: School of Biotechnology and Biomolecular Sciences (UNSW), Sydney, Australia. 4: Department of Bioengineering Sciences, Weihenstephan-Tr. University of Applied Sciences, Freising, Germany. 5: Department of Informatics, Bioinformatics & Computational Biology, Technical University of Munich, Germany. 6: The University of Dundee, UK. 7: Institute of Structural and Molecular Biology, University College London, UK

[8]Senior author

[9]Lead contact

SUMMARY

In response to the COVID-19 pandemic, many life scientists are focused on SARS-CoV-2. To help them use available structural data, we systematically modeled all viral proteins using all related 3D structures, generating 872 models that provide detail not available elsewhere. To organise these models, we created a structure coverage map: a novel, one-stop visualization summarizing what is — and is not — known about the 3D structure of the viral proteome. The map highlights structural evidence for viral protein interactions, mimicry, and hijacking; it also helps researchers find 3D models of interest, which can then be mapped with UniProt, PredictProtein, or CATH features. The resulting Aquaria-COVID resource (https://aquaria.ws/covid) helps scientists understand molecular mechanisms underlying coronavirus infection. Based on insights gained using our resource, we propose mechanisms by which the virus may enter immune cells, sense the cell type, then switch focus from viral reproduction to disrupting host immune responses.

INTRODUCTION

Due to the COVID-19 pandemic, many life scientists have recently switched focus towards SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). This includes structural biologists, who have so far determined ~300 Protein Data Bank entries (PDB; Berman et al. 2000) for the 27 viral proteins.

These structures are, in turn, driving molecular modeling studies, most focused on the spike glycoprotein (Jaimes et al., 2020; Piplani et al., 2020; Rodrigues et al., 2020). Other modelling studies focus on breadth of coverage, predicting 3D structures for the entire SARS-CoV-2 proteome; this has been done using AlphaFold (Senior et al., 2020), C-I-TASSER (Zheng et al., 2019), Rosetta (Rohl et al., 2004), and SwissModel (Waterhouse et al., 2018). Unfortunately, these predictions vary greatly (Heo and Feig, 2020), raising quality issues explored in the current CASP activity (Kryshtafovych et al., 2019). Additionally, these studies derive few structural states for each viral protein, thereby producing moderate numbers of models (e.g., 24 for AlphaFold and 116 for SwissModel).

We aim to address these limitations via a depth-based strategy that models multiple states for each viral protein, using all related 3D structures in the PDB — thus leveraging structures determined for other coronaviruses, such as SARS-CoV, BtCoV-HKU4 (bat coronavirus HKU4), FCoV (feline coronavirus), IBV (avian infectious bronchitis virus), and MHV-A59 (mouse hepatitis virus A59), as well as distant viruses, such as CHIKV (Chikungunya virus) and FMDV (foot-and-mouth disease virus).

Combining breadth and depth of coverage requires modeling methods with low computational cost; here, we use only sequence profile comparisons (Steinegger et al., 2019) to align viral sequences onto experimentally derived 3D structures (O'Donoghue et al., 2015). This generates what we call minimal models, in which 3D coordinates are not modified, but are mapped onto SARS-CoV-2 sequences using coloring that indicates model quality (Heinrich et al., 2015).

Minimal models have benefits: it is easy to understand how they were derived, helping assess the validity of insights gained. Thus, minimal models are broadly useful, even for researchers who are not modelling experts. Conversely, models generated by more sophisticated methods (e.g., Senior et al., 2020) can be more accurate (Kryshtafovych et al., 2019), but require more time and expertise to assess their accuracy (Heo and Feig, 2020) and the validity of insights gained; this often limits their usefulness.

Large numbers of models can be generated by such minimal strategies, raising a new problem: how to visually organize such complex datasets to be usable. We consider this one instance of a critical issue impeding not just COVID-19 research, but all life sciences (O'Donoghue et al., 2018). Thus, we introduce a novel concept: a one-stop visualization summarizing what is known - and not known - about the 3D structure of the viral proteome. This tailored visualization — called the SARS-CoV-2 structural coverage map — helps researchers find structural models related to specific research questions.

Once a structural model of interest is found, it can be used to explore the spatial arrangement of sequence features — i.e., residue-based annotations, such as nonsynonymous mutations or

post-translational modifications. Here, we integrated SARS-CoV-2 models into Aquaria (O'Donoghue et al., 2015), a molecular graphics system designed to simplify feature mapping and make minimal models broadly accessible to researchers who are not modelling experts. Previously, Aquaria could only map features from UniProt (The UniProt Consortium, 2019); for this work, we added >32,000 SARS-CoV-2 features from additional sources.

The resulting Aquaria-COVID resource (https://aquaria.ws/covid) comprises a large set of SARS-CoV-2 structural information not readily available elsewhere. The resource further identifies structurally dark regions of the proteome, i.e., regions with no significant sequence similarity to any protein region observed by experimental structure determination (Perdigão et al., 2015). Dark proteins perform important biological functions (Schafferhans et al., 2018); identifying them helps direct future research that might reveal currently unknown infection mechanisms.

Below, we describe the resource and systematically review insights it provides across the proteome. In the discussion, we highlight insights gained into how SARS-CoV-2 proteins self-assemble, and how they may mimic (Elde and Malik, 2009; Murphy, 1993) host proteins and hijack (Davey et al., 2011) host processes, including innate and adaptive immunity, nonsense-mediated decay, ubiquitination, and regulation of chromatin and telomeres.

During the COVID-19 pandemic, the Aquaria-COVID resource aims to fulfil a vital role by helping scientists more rapidly understand molecular mechanisms of coronavirus infection that are currently known, and to keep abreast of emerging knowledge, as new 3D structures and sequence features become available.

RESULTS

The Aquaria-COVID resource uses 14 UniProt sequences (The UniProt Consortium, 2019) that comprise the SARS-CoV-2 proteome. We matched these (using HHblits; Steinegger et al., 2019) against sequences of all available 3D structures in the PDB (Berman et al., 2000), resulting in 872 minimal models (Tables S1-S3) that we incorporated into Aquaria (O'Donoghue et al., 2015), where they can be mapped with >32,000 features from UniProt, CATH (Dawson et al., 2017), SNAP2 (Hecht et al., 2015), and PredictProtein (Yachdav et al., 2014). These features include residue-based prediction scores for conservation, disorder, domains, flexibility, mutational propensity, subcellular location, and transmembrane helices (see Methods for details). To help other researchers use these models and features we created a matrix layout showing models for all 14 UniProt sequences (https://aquaria.ws/covid#matrix). We also created a structural coverage map (Figures 1 & 2; https://aquaria.ws/covid) — a novel visual layout based on the viral genome organization. The coverage map summarizes key results obtained, including evidence for viral mimicry (Figure 3) or hijacking of host proteins (Figure 4), as well as viral protein interactions (Figure 5). The map also helps scientists find 3D models related to specific research questions. Below, we introduce each element of the structural coverage map as we use it to describe structural models found in three viral genome regions.

Polyprotein 1a

Polyprotein 1a (a.k.a. PP1a) derives from polyprotein 1ab (a.k.a. PP1ab), and is cleaved into 10 proteins (NSP1–NSP10) that modify viral proteins, disable host defenses, and support viral replication.

NSP1 derives from residues 1–180 of PP1a and is thought to interact with the ribosome, suppressing translation of host mRNAs and promoting their degradation (Kamitani et al., 2009). We found two structures matching most of the NSP1 sequence (Figure 1), both determined for NSP1 from SARS-CoV, which had 85% pairwise sequence identity to the matched region of NSP1 from SARS-CoV-2. These sequence-to-structure alignments were generated using HHblits sequence profiles (Steinegger et al., 2019), and had a significance score (E-value) of 10^{-20}, giving ≥ 97% likelihood that NSP1 from SARS-CoV and SARS-CoV-2 adopt similar structures (O'Donoghue et al., 2015). Both matching structures were from the same researchers (Almeida et al., 2007), so the more recent structure (P0DTC1/2gdt) was selected for the representative image, which was colored to convey alignment quality (Heinrich et al., 2015) (Figure 1). Unfortunately, neither structure showed interactions with other proteins or RNA and — unusually — they provided few functional insights (Almeida et al., 2007), partly because they had a unique fold with no matches in CATH (Dawson et al., 2017). NSP1 also had two short dark regions resulting from alignment gaps (Figure 1); the N-terminal dark region may be accounted for by high flexibility (average predicted B-value = 60; see Methods).

NSP2 (PP1a 181–818) may disrupt intracellular signaling by interaction with host proteins (Cornillez-Ty et al., 2009). Unfortunately, these interactions are not captured in the single matching structure (P0DTC1/3ld1; Figure 1) derived for NSP2 from IBV, a remote homolog with 11% pairwise sequence identity to the matched region of PP1a (residues 104–562). However, the HHblits E-value was 10^{-32}, giving ≥ 98.5% likelihood that SARS-CoV-2 NSP2 adopts a similar structure — a stronger prediction than for NSP1 (above), largely because the matched region was much longer. Interestingly, the matched region overlapped NSP1 (PP1a 104–180). NSP2 also had multiple short dark regions, resulting from alignment gaps, plus a long C-terminal dark region (Figure 1). The Aquaria webpage for this matching structure showed that — very unusually — there is no scientific publication associated with this 10 year-old dataset. Thus, while it may be of some use for researchers focused on NSP2, for most researchers this structure should probably be excluded, making NSP2 an entirely dark protein.

NSP3 (PP1a 819–2763) is a large, multidomain protein thought to perform many functions, including anchoring the viral replication complex to double-membrane vesicles derived from the endoplasmic reticulum (Lei et al., 2018). NSP3 had 188 matching structures clustered in nine distinct sequence regions.

NSP3 region 1 (a.k.a. Ubl1; PP1a 819–929) was the least conserved NSP3 region (average ConSurf score = 3.7; see Methods), suggesting it adapts to host-specific defenses. Ubl1 is thought to bind single-stranded RNA and the viral nucleocapsid protein (Lei et al., 2018). Unfortunately, these interactions are not shown in the three matching structures found, which all adopt a ubiquitin-like topology (CATH 3.10.20.350). Although it has distinct structural differences, Ubl1 may mimic host ubiquitin (Lei et al., 2018); however, we found no matches to structures of human ubiquitin, undermining the mimicry hypothesis. For the representative image (Figure 2), we used the top-ranked matching structure provided by Aquaria:

P0DTC1/2gri, derived for NSP3 from SARS-CoV, which aligned onto SARS-CoV-2 NSP3 with 77% identity and with a significance of E = 10^{-23}.

NSP3 region 2 (PP1a 930–1026) had no matches in CATH and no matching structures. This was the NSP3 region with lowest predicted sensitivity to mutation (median sensitivity 0%; see Methods), highest predicted flexibility (average B-factor = 55), highest fraction of disordered residues (47%), and highest fraction of residues predicted to be solvent-accessible (99%). We speculate that this region acts as a flexible linker and may contain post-translational modification sites hijacking host signalling, as are often found in viral disordered regions (Davey et al., 2011).

NSP3 region 3 (PP1a 1027–1193) has a macro domain (a.k.a. Mac1 or X domain; CATH 3.40.220.10) that may counteract innate immunity via interfering with ADP-ribose (ADPr) modification (Lei et al., 2018; Rack et al., 2016). This was the second least conserved NSP3 region (ConSurf = 3.9) and had the highest fraction of mutationally sensitive residues (29%), suggesting it is well adapted to specific hosts. This region had 144 matching structures; as a representative image we used P0DTC1/6woj, derived from SARS-CoV-2 (Figure 2). Of the matching structures, 47 were of human proteins (Figures 2 & 3A), suggesting these host proteins may be mimicked by the virus. The potentially mimicked human proteins were: GDAP2 (P0DTC1/4uml, 19% identity to NSP3 macro domain, E = 10^{-22}), MACROD1 (P0DTC1/2x47, 25% identity, E = 10^{-22}), MACROD2 (P0DTC1/4iqy, 24% identity, E = 10^{-22}), MACROH2A1 (P0DTC1/1zr5, 19% identity, E = 10^{-18}), MACROH2A2 (P0DTC1/2xd7, 19% identity, E = 10^{-21}), OARD1 (P0DTC1/2eee, 13% identity, E = 10^{-12}), PARP9 (P0DTC1/5ail, 22% identity, E = 10^{-17}), PARP14 (P0DTC1/3q6z, 25% identity, E = 10^{-19}), and PARP15 (P0DTC1/3v2b, 18% identity, E = 10^{-16}). An additional 73 structures matched to viral proteins, two in complex with RNA (P0DTD1/4tu0, 24% identity, E = 10^{-17} and P0DTD1/3gpq, 24% identity, E = 10^{-17}; both from CHIKV). For brevity here and in Figure 2, we have omitted the remaining 41 matching structures derived from other host organisms (Table S1).

NSP3 region 4 (PP1a 1027–1368) comprised a disordered region plus a macro-like domain called SUD-N (a.k.a. Mac2; CATH 3.40.220.30) that may bind RNA (Lei et al., 2018). This region had no matching structures.

NSP3 region 5 (PP1a 1389–1493) comprised another macro-like domain called SUD-M (a.k.a. Mac3; CATH 3.40.220.20) that may bind both RNA and host proteins, and take part in viral replication (Lei et al., 2018). However, no interactions are shown in the seven matching structures found (P0DTC1/2jzd, 81% identity, E = 10^{-34}), all determined using SARS-CoV NSP3. Comparing these with structures matching NSP3 region 3, we see considerable differences, and no evidence of mimicry of host macro domains (Figure 2).

NSP3 region 6 (PP1a 1494–1562) comprised a domain called SUD-C (CATH 2.30.30.590) and had two matching structures: one from SARS-CoV (P0DTC1/2kqw, 76% identity, E = 10^{-41}) and one matching a distant virus (P0DTC1/4ypt from MHV-A59, 28% identity, E = 10^{-62}), which also spanned NSP3 region 7. Using these structures, Lei et al. speculated that this region may bind metal ions and induce oxidative stress (Lei et al., 2018).

NSP3 region 7 (PP1a 1563–1878) comprises a papain-like protease (a.k.a. PL-Pro), made up of three domains (CATH 3.10.20.540, CATH 1.10.8.1190, and CATH 3.90.70.90) thought to cleave NSP1–NSP3 from the polyprotein and to cleave ubiquitin-like modifications from host proteins (Figure 4A), thereby undermining interferon-induced antiviral activity (Dzimianski et al., 2019;

Lei et al., 2018). PL-Pro had 38 matching structures (P0DTC1/6wrh from SARS-CoV-2, 100% identity), of which 11 show binding to human ubiquitin-like proteins (Figures 2 & 4A): four showed binding to ISG15 (P0DTC1-5tl6 from SARS-CoV, 82% identity, E = 10^{-70}); two showed binding to UBA52 (P0DTC1/4rf0 from MERS-CoV, 29% identity, E = 10^{-65}); two showed binding to UBB (P0DTC1/5wfi from MHV-A59, 31% identity, E = 10^{-61}); one showed binding to UBC (P0DTC1/4mm3 from SARS-CoV, 83% identity, E = 10^{-62}), and one showed binding to both UBB and UBC (P0DTC1/5e6j from SARS-CoV, 83% identity, E = 10^{-62}). Note these identity scores and E-values indicate similarity between SARS-CoV-2 NSP3 and the viral proteins in each matching structure, not to the human proteins. Two additional matching structures showed SARS-CoV-2 NSP3 in complex with inhibitory peptides (P0DTC1/6wuu, 100% identity, P0DTC1/6wx4, 100% identity).

NSP3 region 8 (PP1a 1879–2020) comprises a nucleic-acid binding (a.k.a. NAB) domain thought to bind single-stranded RNA and to unwind double-stranded DNA (Lei et al., 2018). NAB had only one matching structure (P0DTC1/2k87 from SARS-CoV, 82% identity, E = 10^{-24}) with a fold not seen in any other PDB structure (CATH 3.40.50.11020).

NSP3 region 9 (PP1a 2021–2763) may anchor NSP3 to double-membrane vesicles (Lei et al., 2018). This was the most conserved NSP3 region (ConSurf = 4.9), suggesting it is less adapted to specific hosts. This region had no CATH matches and no matching structures.

NSP4 (PP1a 2764–3263) may act with NSP3 and NSP6 to create the double-membrane vesicles required for viral replication (Angelini et al., 2014). NSP4 mostly comprised a dark region (PP1a 2764–3172) with no CATH matches, no disorder, and multiple transmembrane helices. The C-terminal region (PP1a 3173–3263) comprised a domain called NSP4C (CATH 1.10.150.420) that had three matching structures (P0DTC1/3vcb from MHV-A59, 59% identity, E = 10^{-35}); all were homodimers, although NSP4C is thought to act as a monomer (Xu et al., 2009).

NSP5 (a.k.a. 3C-like protease or 3CL-Pro; PP1a 3264–3569) is a two-domain protein (CATH (CATH 2.40.10.10 and CATH 1.10.1840.10) thought to cleave the viral polyprotein at 11 sites, resulting in NSP4–NSP16. NSP5 had 256 matching structures (P0DTC1/5rfa from SARS-CoV-2, 100% identity) — more than any other viral protein — of which, 15 showed binding to inhibitory peptides (P0DTC1/7bqy from SARS-CoV-2, 100% identity).

NSP6 (PP1a 3570–3859) is a transmembrane protein thought to act with NSP3 and NSP4 to create double-membrane vesicles (Angelini et al., 2014). Of the 15 PP1ab proteins, NSP6 was the only completely dark protein and had the lowest conservation (ConSurf = 2.9). NSP6 also had no CATH matches and no disordered regions.

NSP7 (PP1a 3860–3942) may support viral genome replication (te Velthuis et al., 2012); it had 15 matching structures, some showing interactions with other viral proteins (Figures 2 & 5A). In two structures, NSP7 occurs as a monomer (P0DTC1/2kys, 98% identity, E = 10^{-32} and P0DTC1/1ysy, 98% identity, E = 10^{-32} — both from SARS-CoV), while the remaining 13 structures showed NSP7 bound to NSP8 (P0DTC1/3ub0 from FCoV, 43% identity, E = 10^{-79}). Six of these structures also showed binding to NSP12 (P0DTC1/7bv1 from SARS-CoV-2, 99% identity); two of these six structures also showed binding to RNA (P0DTC1/7bv2, 99% identity and P0DTC1/6yyt, 97% identity, E = 10^{-32} — both from SARS-CoV-2). NSP7 comprises an

antiparallel helical bundle (CATH 1.10.8.370) that adopts distinct substates, depending on its interaction partners (Figure 5A).

NSP8 (PP1a 3943–4140) is also thought to support viral genome replication (te Velthuis et al., 2012). It features a highly conserved (ConSurf = 7.3) 'tail' segment (PP1a 3943–4041), predominantly helical with some disordered residues and no CATH matches, followed by a less conserved (ConSurf = 5.7) 'head' domain (PP1a 4042–4140), with a beta barrel fold (CATH 2.40.10.290). NSP8 had 14 matching structures, all showing interactions to other viral proteins (Figures 2 & 5A). One structure showed binding to NSP12 only (P0DTC1/6nus from SARS-CoV, 97% identity, $E = 10^{-78}$), while the remaining 13 structures all showed binding to NSP7 (P0DTC1/3ub0). Six of these 13 structures also showed binding to NSP12, and two of these six showed binding to RNA as well.

NSP9 (PP1a 4141–4253) may bind single-stranded RNA and take part in genome replication (Miknis et al., 2009). NSP9 had 14 matching structures (P0DTC1/6wxd from SARS-CoV-2, 98% identity, $E = 10^{-44}$), all with beta barrel architecture (CATH 2.40.10.250) and mostly homodimers — thought to be the functional state (Miknis et al., 2009).

NSP10 (PP1a 4254–4392) is thought to act with NSP14 and NSP16 to cap and proofread RNA during genome replication (Bouvet et al., 2012). NSP10 had no CATH matches, yet had 33 matching structures, some showing interactions to other viral proteins (Figures 2 & 5B). In one matching structures, NSP10 was monomeric (P0DTC1/2fyg from SARS-CoV, 97% identity, $E = 10^{-57}$), while in two structures, NSP10 was a dodecamer (Figure 1; P0DTC1/2ga6 from SARS-CoV, 96% identity, $E = 10^{-72}$ and P0DTC1/2g9t from SARS-CoV, 96% identity, $E = 10^{-72}$), forming a hollow sphere with 12 C-terminal zinc finger motifs on the outer surface, and another 12 on the inner surface (Su et al., 2006). Four matching structures showed binding to NSP14 (P0DTC1/5c8u from SARS-CoV, 97% identity, $E = 10^{-68}$), while the remaining 26 structures showed binding to NSP16 (P0DTC1/6w61 from SARS-CoV-2, 100% identity).

Polyprotein 1b

Polyprotein 1b (a.k.a. PP1b) is cleaved by 3CL-PRO into five proteins (NSP12–NSP16) that drive replication of viral RNA. These proteins were predicted to have no disordered regions, no transmembrane helices, very few dark regions, and to be highly conserved (ConSurf = 5.3–6.6), compared with PP1a proteins (ConSurf = 2.9–5.6).

NSP12 (PP1ab 4393–5324) is an RNA-directed RNA polymerase (a.k.a. RdRp) thought to drive viral genome replication (Yin et al., 2020). NSP12 was the second most conserved PP1ab protein (ConSurf = 6.5), and had 47 matching structures, many showing interactions to other viral proteins (Figures 2 & 5A). One structure showed binding to NSP8 (P0DTC1/6nus from SARS-CoV, 97% identity, $E = 10^{-77}$), while four structures showed binding to both NSP8 and NSP7 (P0DTD1/7bv1 from SARS-CoV-2, 99% identity, $E = 10^{-230}$); two of these four also showed binding to RNA (P0DTD1/6yyt from SARS-CoV-2, 100% identity). An additional 15 structures showed binding to RNA only (P0DTD1/3kmq from FMDV, 18% identity, $E = 10^{-11}$), while one structure showed binding to both RNA and DNA (P0DTD1/4k4v from poliovirus, 17% identity, $E = 10^{-11}$).

NSP13 (PP1ab 5325–5925) is a multi-functional helicase thought to unwind double-stranded RNA and DNA (Adedeji et al., 2012) and to be a core component of the viral replication complex

(Subissi et al., 2014). The N-terminal half of NSP13 (PP1ab 5325–5577) had no matches in CATH, while the C-terminal half contained two Rossman fold domains (CATH 3.40.50.300). NSP13 had 64 matching structures, of which 23 showed potential mimicry of four human helicase proteins (Figures 2 & 3B). Two of these structures showed matching between PIF1 (P0DTD1/6hpt, 20% identity, E = 10^{-12}) and the first Rossman fold domain (PP1ab 5597–5733). A further 11 structures showed matching between AQR (P0DTD1/4pj3, 19% identity, E = 10^{-30}) and the second Rossman fold domain plus part of the first (PP1ab 5689–5909); of these structures, 10 showed AQR bound to the spliceosome (P0DTD1/6id0, 20% identity, E = 10^{-30}). The remaining 10 structures matched to both Rossman fold domains: eight of these structures showed mimicry of UPF1 (P0DTD1/2xzo, 24% identity, E = 10^{--32}), of which two structures also showed binding to UPF2 (P0DTD1/2wjy, 23% identity, E = 10^{-34}); finally, two structures showed mimicry of IGHMBP2 (P0DTD1/4b3f, 23% identity, E = 10^{-35}), of which one structure also showed binding with RNA (P0DTD1/4b3g, 25% identity, E = 10^{-33}). An further 41 structures showed matches to viral proteins (P0DTD1/6jyt from SARS-CoV, 100% identity), of which four structures also showed binding to DNA (P0DTD1/4n0o, 22% identity, E = 10^{-19}).

NSP14 (PP1ab 5926–6452) is a proofreading exoribonuclease (a.k.a. ExoN) thought to remove 3'-terminal nucleotides from RNA, thereby reducing mutations during viral genome replication (Minskaia et al., 2006). NSP14 had no matches in CATH, but had four matching structures (P0DTD1/5nfy, 100% identity), all in complex with NSP10 (Figures 2 & 5B).

NSP15 (PP1ab 6453–6798) is a uridylate-specific endoribonuclease (a.k.a. NendoU) thought to support viral genome replication (Ricagno et al., 2006). The N-terminal region of NSP15 had two domains (CATH 2.20.25.360 and CATH 3.40.50.11580), while the C-terminal region (PP1ab 6642–6798) had none. NSP15 had 19 matching structures (P0DTD1/6wxc, 99% identity), none showing matching to or interaction with human proteins, or with other viral proteins (Figure 2).

NSP16 (PP1ab 6799–7096) may methylate viral mRNA caps, following replication, which is thought to be important for evading host immune defenses (Bouvet et al., 2010). This was also the most conserved (ConSurf = 6.6) PP1ab protein. NSP16 had a Rossman fold domain (CATH 3.40.50.150) that had 26 matching structures (P0DTC1/6w61, 100% identity), all showing binding to NSP10 (Figures 2 & 5B).

Virion and accessory proteins

The 3' end of the genome encodes 12 proteins, many involved in virion assembly. Remarkably, our analysis found no interactions between these proteins.

Spike glycoprotein (a.k.a. S protein) binds host receptors, thereby initiating membrane fusion and viral entry (Hoffmann et al., 2020). This protein has four UniProt domains, of which one (S1-CTD; residues 334–527) binds to host receptors (Figure 4B). Aquaria found 136 matching structures (P0DTC2/6vxx from SARS-CoV-2, 99% identity) clustered in two regions (Figure 2). One was a heptad repeat region (HR2), with 15 matching structures (P0DTC2/2fxp from SARS-CoV, 98% identity, E = 10^{-14}), of which four structures showed binding to antibodies (P0DTC2/6pxh from MERS-CoV, 22% identity, E = 10^{-11}). The second region had 121 matching structures, of which 68 showed binding to antibodies (P0DTC2/6w41 from SARS-CoV-2, 100% identity) and two structures showed binding to inhibitory peptides (P0DTC2/5zvm from SARS-CoV, 88% identity, E = 10^{-32}; P0DTC2/5zvk from SARS-CoV, 55% identity, E = 10^{-32}). Of

the remaining structures, 18 showing interaction with human proteins (Figures 2 & 4B): 15 showed binding to ACE2 (P0DTC2/6acg from SARS-CoV, 77% identity, E = 10^{-321}), of which one structure also showed binding to SLC6A19 (P0DTC2/6m17 from SARS-CoV-2, 100% identity); finally, three structures showed binding to DPP4 (P0DTC2/4qzv from BtCoV-HKU4, 21% identity, E = 10^{-45}).

ORF3a may act as a homotetramer, forming an ion channel in host cell membranes that helps with virion release (Lu et al., 2006). ORF3a had no matching structures.

The envelope protein (a.k.a. E protein) matched two structures, both from SARS-CoV. One was a monomer (P0DTC4/2mm4, 91% identity, E = 10^{-26}) while the other was a pentamer (P0DTC4/5x29, 86% identity, E = 10^{-27}), forming a ion channel thought to span the viral envelope (Surya et al., 2018; Vennema et al., 1996).

The matrix glycoprotein (a.k.a. M protein) is also part of the viral envelope (Vennema et al., 1996). The matrix glycoprotein had no matching structures.

ORF6 may block expression of interferon stimulated genes (e.g., ISG15) that have antiviral activity (Frieman et al., 2007). ORF6 had no matching structures.

ORF7a may interfere with the host cell surface protein BST2, preventing it from tethering virions (Taylor et al., 2015). ORF7a had three matching structures: one from SARS-CoV-2 (P0DTC7/6w37, 100% identity) and two from SARS-CoV (P0DTC7/1yo4, 90% identity, E = 10^{-59}; P0DTC7/1xak, 88% identity, E = 10^{-59}).

ORF7b is an integral membrane protein thought to localize to the Golgi compartment and the virion envelope (Schaecher et al., 2007). ORF7b had no matching structures.

ORF8 is thought to inhibit type 1 interferon signaling (J.-Y. Li et al., 2020); it is also very different to proteins from other coronaviruses. ORF8 had no matching structures.

The nucleocapsid protein (a.k.a. N or NP protein) is thought to package the viral genome during virion assembly through interaction with the matrix glycoprotein (Saikatendu et al., 2007), and also to become ADP-ribosylated (Grunewald et al., 2018). Depending on its phosphorylation state, this protein may also switch function, translocating to the nucleus and interacting with the host genome (Surjit et al., 2005). This protein had 35 matching structures clustered in two regions: the N-terminal region had 24 matching structures (P0DTC9/6yi3 from SARS-CoV-2, 100% identity), of which one was a tetramer, five were dimers, and the rest were monomers; the C-terminal region had 13 matching structures (P0DTC9/6wji from SARS-CoV-2, 98% identity), all dimers.

ORF9b is a lipid-binding protein thought to interact with mitochondrial proteins, thereby suppressing interferon-driven innate immunity responses (Shi et al., 2014). ORF9b matched two structures (P0DTD2/2cme from SARS-CoV, 78% identity, E = 10^{-48}) that showed binding to a lipid analog (Meier et al., 2006).

ORF9c (a.k.a. ORF14) is currently uncharacterized experimentally; it is predicted to have a single-pass transmembrane helix. ORF9c had no matching structures.

ORF10 is a predicted protein that currently has limited evidence of translation (Gordon et al., 2020), has no reported similarity to other coronavirus proteins, and has no matching structures.

DISCUSSION

The 872 models derived in this study capture essentially all structural states of SARS-CoV-2 proteins with supporting experimental evidence. We used these states to create a structural coverage map (Figure 2): a concise yet comprehensive visual summary of what is known about the 3D structure of the viral proteome. Remarkably, we found so few states showing viral self-assembly (Figure 5), mimicry (Figure 3), or hijacking (Figure 4) that — excluding non-human host proteins — all states could be included in the coverage map via several simple graphs. This may indicate that host interactions are rarely used in COVID-19 infection, consistent with the notion that viral activity is largely shielded from the host. However, other experimental techniques have found many more interactions between viral proteins (Pan et al., 2008) and with host proteins (Gordon et al., 2020). Thus, the small number of interactions found in this work likely indicates limitations in current structural data.

Using the current structural data (Figure 2), we can divide the 27 SARS-CoV-2 proteins into four categories: mimics, hijackers, teams, and suspects — below, we highlight insights derived in this work for each of these categories.

Mimics

We found structural evidence for mimicry of human proteins for only two SARS-CoV-2 proteins: NSP3 and NSP13 (Figures 2 & 3; Table S4).

NSP3 may mimic host proteins containing macro domains, interfering with ADP-ribose (ADPr) modification and thereby suppressing host innate immunity (Lei et al., 2018; Rack et al., 2016). We found nine potentially mimicked proteins (Figure 3A); the top ranked matches (MACROD2 and MACROD1) remove ADPr from proteins (Hottiger, 2015; O'Sullivan et al., 2019), reversing the effect of ADPr writers (e.g., PARP9, PARP14, and PARP15 in lymphoid tissues), and affecting ADPr readers (e.g., the core core histone proteins MACROH2A1, and MACROH2A2, found in most cells). Thus we speculate that, in infected cells, ADPr erasure by NSP3 may influence epigenetic regulation of chromatin state (Schäfer and Baric, 2017), potentially contributing to the high variation in COVID-19 patient outcomes (Callaway et al., 2020). Furthermore, in infected macrophages, activation by PARP9 and PARP14 may be undermined by NSP3's erasure of ADPr, resulting in vascular disorders (Iwata et al., 2016), as seen in COVID-19 (Varga et al., 2020).

NSP13 (a.k.a. viral helicase) may mimic four human helicases, based on stronger alignment evidence than for mimicry by NSP3 (Figure 3). Three of these helicases are associated with DNA repair and recombination, thus we speculate mimicry of these proteins could either recruit host proteins to assist in viral recombination (Graham and Baric, 2010), or alternatively, could dysregulate recombination of host DNA. However, we found no evidence for mimicry of the ~100 other human helicases (Umate et al., 2011), many also involved in recombination, suggesting that mimicry may hijack more specific functions performed by the four helicases. The top ranked helicase was IGHMBP2 (a.k.a. SMBP2), which interacts with single-stranded DNA in the class switching region of the genome (Yu et al., 2011), close to *IGMH* (the gene coding the

constant region of immunoglobulin heavy chains); we speculate that mimicry of IGHMBP2 may explain the dysregulation of immunoglobulin-class switching observed clinically (Bauer, 2020). The second ranked helicase was UPF1 (a.k.a. up-frameshift suppressor 1, RENT1, or regulator of nonsense transcripts 1), which acts in the cytoplasm as part of the nonsense-mediated mRNA decay pathway known to counteract coronavirus infection (Wada et al., 2018); we speculate that mimicry of UPF1 may hijack this pathway, thus impeding host defenses. UPF1 also has acts in the nucleus, interacting with telomeres — as does the helicase PIF1; we speculate that mimicry of UPF1 or PIF1 may be implicated to the connection seen between COVID-19 severity, age, and telomere length (Aviv, 2020). In summary, we propose that (as mentioned above for nucleocapsid protein) NSP13 may sometimes switch from its key role in viral replication to a state focused on host genome interactions that undermine host immunity.

Hijackers

We found direct structural evidence for hijacking of human proteins for only two SARS-CoV-2 proteins: NSP3 and spike glycoprotein (Figures 2 & 4; Table S4).

NSP3 is believed to cleave ubiquitin from host proteins (Barretto et al., 2005), thereby suppressing innate immune response and disrupting proteasome-mediated degradation (Lei et al., 2018). We found six structures showing NSP3 PL-Pro bound to ubiquitin domains from UBB, UBC, or UBA52 (a.k.a. ubiquitin-60S ribosomal protein L40). Four additional structures showed binding to the ubiquitin-like domains of ISG15 (a.k.a. interferon stimulated gene 15), which is attached to most newly synthesized proteins; ISGylation does not induce degradation but likely disturbs virion assembly (Dzimianski et al., 2019; Held et al., 2020; Lei et al., 2018). Together, these 10 structures show in detail how NSP3 reverses either ubiquitination or ISGylation (Figure 4B); a further two structures show how NSP3 can be blocked with inhibitors (Figure 2).

Spike glycoprotein is known to bind receptors on the host cell membrane, thereby initiating membrane fusion and viral entry (Hoffmann et al., 2020). We found 16 matching structures showing hijacking of ACE2, a carboxypeptidase that cleaves vasoactive peptides. One of these structures also shows binding to SLC6A19 (a.k.a. B^0AT1), an amino acid transporter reported to act with ACE2 (Camargo et al., 2009). Most structures show only short fragments of spike; however, by consolidating several structures and sequence features, we created an integrated view showing the cellular context of the pre-fusion event, when spike first binds the ACE + SLC6A19 complex (Figure 4B). Three further matching structures showed interactions between spike (from MERS-CoV and BtCoV-HKU4) and host membrane receptor DPP4 (a.k.a. CD26), an aminopeptidase that SARS-CoV-2 may use (Y. Li et al., 2020) to enter host immune cells (Radzikowska et al., 2020). If so, DPP4 and ACE2 may interact with spike at distinct but overlapping binding sites (Figure 4B) — making this region a potential therapeutic target. Once inside immune cells, the virus may switch to a non-reproductive state aimed at host cell disruption — as reported for MERS and SARS (Chu et al., 2016; Gu et al., 2005) — thereby contributing to lymphopenia in severe COVID-19 (Tan et al., 2020). We speculate that this cell-dependent state switching may be driven by NSP3 modifications of either ADP ribosylation, ubiquitination, or ISGlyation, and may involve the state switches noted above for NSP13 and nucleocapsid protein.

Teams

We found structural evidence for interaction between only six SARS-CoV-2 proteins (Figures 2 & 5); they divided into two distinct teams, each with three proteins.

Team 1 comprised NSP7, NSP8, and NSP12, all members of the viral RNA replication complex (Figure 5A). NSP12 alone can replicate RNA, as can NSP7 + NSP8 acting together (te Velthuis et al., 2012). However, replication is greatly stimulated by cooperative interactions between these proteins (Kirchdoerfer and Ward, 2019). Most combinations of these proteins occur in the 60 matching structures found (Figure 5A), although models for NSP12 (alone or with RNA) all derived from very remote viruses (e.g., enterovirus). The matrix of combinations (Figure 5A) gives insight into how the complex may assemble, and shows that several proteins that may be involved in genome replication (Subissi et al., 2014) are missing (NSP3, NSP9, NSP10, NSP13, NSP14, and NSP16). These outcomes again demonstrate the value of modelling all available structural states, compared with strategies that select only one or few structural states per protein.

Team 2 comprised NSP10, NSP14, and NSP16 (Figure 5B). All matching structures found for NSP14 and NSP16 showed binding with NSP10, consistent with the belief that NSP10 is required for NSP16 RNA-cap methyltransferase activity (Decroly et al., 2011), and also for NSP14 methyltransferase and exoribonuclease activities (Ma et al., 2015). In addition, NSP10 was found to form a homododecamer; however, these three states of NSP10 share a common binding region (Figure 5B; Bouvet et al., 2014). This suggests NSP10, NSP14, and NSP16 interact competitively — in contrast to cooperative interactions seen in team 1. We speculate that NSP10 is produced at higher abundance than NSP14 or NSP16, or may be rate limiting for viral replication.

Finally, it is noteworthy that no interactions were found between the 12 virion or accessory proteins (Figure 1, bottom third). This, again, highlights limitations in currently available structural data.

Suspects

This leaves 18 of the 27 viral proteins in a final category we call suspects: these are proteins thought to play key roles in infection, but having no structural evidence of interaction with RNA, DNA, or other proteins (viral or human). We divided the suspects into two groups, based on matching structures.

Group 1 proteins were those with at least one matching structure: NSP1, NSP2, NSP4, NSP9, NSP15, E, ORF7a, and ORF9a. Some of these have been well studied (e.g., NSP5 had 256 matching structures). Yet none of these proteins had significant similarity to any experimentally determined 3D structure involving human proteins, or to any structure showing interactions between viral proteins — based on the methods used in this work.

Group 2 proteins were those with no matching structures: NSP6, ORF3a, matrix glycoprotein (a.k.a. M protein), ORF6, ORF8, and ORF9c, and ORF10. As noted previously, NSP2 could, arguably, be included in this group. These are structurally dark proteins (Perdigão et al., 2015), meaning not only is their structure unknown, but also that they have no significant sequence similarity to any experimentally determined 3D structure — based on the methods used in this

work. These proteins are ripe candidates for advanced modelling strategies, e.g., using predicted residue contacts combined with deep learning (e.g., Senior et al. 2020).

Conclusions

We have assembled a wealth of structural data, not available elsewhere, about the viral proteome. We used these data to construct a structural coverage map summarizing what is known — and not known — about the 3D structure of SARS-CoV-2 proteins. The coverage map also summarizes structural evidence for viral protein interaction teams and for mimicry and hijacking of host proteins — and identifies suspect viral proteins, with no evidence of interactions to other macromolecules. Additionally, the map helps researchers find 3D models of interest, which can then be mapped with a large set of sequence features to provide insights into protein function. The resulting Aquaria-COVID resource (https://aquaria.ws/covid) aims to fulfil a vital role during the COVID-19 pandemic, helping scientists use emerging structural data to understand the molecular mechanisms underlying coronavirus infection. Based on insights gained using our resource, we propose mechanisms by which the virus may enter immune cells (via spike hijacking of DPP4), sense the cell type (via posttranslational modifications by NSP3), then switch focus from viral reproduction to disrupting host immune responses (via hijacking by NSP13 and nucleocapsid protein). Our resource lets researchers easily explore and assess the evidence for these mechanisms — or for others they may propose themselves — thereby helping direct future research.

METHOD DETAILS

SARS-CoV-2 Sequences

This study was based on the 14 protein sequences provided in UniProtKB/Swiss-Prot version 2020_03 (released April 22, 2020; https://www.uniprot.org/statistics/) as comprising the SARS-CoV-2 proteome. Swiss-Prot provides polyproteins 1a and 1ab (a.k.a. PP1a and PP1ab) as two separate entries, both identical for the first 4401 residues; PP1a then has four additional residues ('GFAV') not in PP1ab, which has 2695 additional residues not in PP1a. Swiss-Prot also indicates residue positions at which the polyproteins become cleaved in the cell, resulting in 16 protein fragments, named NSP1 though NSP16. The NSP11 fragment comprises the last 13 residues of PP1a (4393–4405). The first 9 residues of NSP12 are identical to the first nine of NSP11, but the rest of that 919 residue long protein continues with a different sequence due to a functionally important frameshift between ORF1a and ORF1b (Nakagawa et al., 2016). Thus, following cleavage, the proteome comprises a final total of 27 separate proteins.

Sequence-to-Structure Alignments

The 14 SARS-CoV-2 sequences were then systematically compared with sequences derived from all known 3D structures from all organisms, based on PDB released on May 30, 2020. These comparisons used the latest version of HHblits (Steinegger et al., 2019) — an alignment method employing iterative comparisons of hidden Markov models (HMMs) — and using the processing pipeline defined previously (O'Donoghue et al., 2015), accepting all sequence-to-structure alignments with a significance threshold of $E \leq 10^{-10}$. The result is a database of protein sequence-to-structure homologies called PSSH2, which is used in the Aquaria web resource.

HHblits has been substantially updated (Mirdita et al., 2017) since we last assessed the specificity and sensitivity of PSSH2 (O'Donoghue et al., 2015), therefore we repeated our assessment of resulting alignments. We used CATH (Dawson et al., 2017) to replace the discontinued COPS database (Suhrer et al., 2009) used previously. Our test data set comprised 23,028 sequences from the CATH nr40 data set. We built individual sequence profiles against UniClust30 and used these profiles to search against "PDB_full", a database of HMMs for all PDB sequences. We then evaluated how many false positives were retrieved at an E-value lower than 10^{-10}, where a false positive was seen to be a structure with a different CATH code at the level of Homologous superfamily (H) or Topology (T). We compared the ratio of false positives received with HH-suite3 and UniClust30 with a similar analysis for data produced in 2017 with HH-suite2 and UniProt20, and found that in both cases the false positive rate was at 2.5% at the homology level (H), and 1.9% at the topology level (T). The recovery rate, i.e. the ratio of proteins from the CATH nr40 data (with less than 40% sequence identity) found by our method that have the same CATH code, was slightly higher with HH-suite3 (20.8% vs. 19.4%).

For each sequence-to-structure alignment, the Aquaria interface gives a pairwise sequence identity score, thus providing an intuitive indication of how closely related the given region of SARS-CoV-2 is to the sequence of the matched structure. However, to more accurately assess

the quality of the match, Aquaria also gives an E-value, calculated by comparing two HMMs, one generated for each of these two sequences.

PredictProtein Features

To facilitate analysis of SARS-CoV-2 sequences, we enhanced the Aquaria resource to include PredictProtein features (Yachdav et al., 2014), thus providing a very rich set of predicted features for all Swiss-Protein protein sequences. In Aquaria, the PredictProtein feature collection is fetched directly by the browser via:

```
https://api.predictprotein.org/v1/results/molart/:uniprot_id
```

The PredictProtein feature sets used in the analysis presented in this work are described below.

Conservation. This feature set is generated by ConSurf (Ashkenazy et al., 2010; Celniker et al., 2013) and gives, for each residue, a score between 1 and 9, corresponding to very low and very high conservation, respectively. These scores estimate the evolutionary rate in protein families, based on evolutionary relatedness between the query protein and its homologues from UniProt using empirical Bayesian methods (Mayrose et al., 2004). The strength of these methods is that they rely on the phylogeny of the sequences and thus can accurately distinguish between conservation due to short evolutionary time and conservation resulting from importance for maintaining protein foldability and function.

Disordered Regions. This feature set gives consensus predictions generated by Meta-Disorder (Schlessinger and Rost, 2005), which combines outputs of several structure-based disorder predictors (Schlessinger et al., 2009) to classify each residue as either disordered or not disordered.

Relative B-values. This feature set predicts, for each residue, normalized B-factor values (Masmaliyeva and Murshudov, 2019) expected to be observed in an X-ray-derived structure. The predictions were generated by PROFbval (Schlessinger et al., 2006) which gives, for each residue, a score between 0 and 100; residues with a score of 50 are predicted to have average flexibility, while those with ≥71 are considered to be highly flexible.

Topology. This feature set is generated by TMSEG (Bernhofer et al., 2016), a machine learning model that uses evolutionary-derived information to predict regions of a protein that traverse membranes, as well as the subcellular locations of the complementary (non-transmembrane) regions.

SNAP2 Features

We further enhanced Aquaria to include SNAP2 features, which provides details on the mutational propensities for each residue position (Hecht et al., 2015). In Aquaria, the SNAP2 feature collection for each protein sequence is fetched directly by the browser via:

```
https://rostlab.org/services/aquaria/snap4aquaria/json.php?uniprotAcc=:uniprot_id
```

Two SNAP2 feature sets were used in this work:

Mutational Sensitivity. For each residue position, this feature set provides 20 scores indicating the predicted functional consequences of the position being occupied by each of the 20

standard amino acids. Large, positive scores (up to 100) indicate substitutions likely to have deleterious changes, while negative scores (down to -100) indicate no likely functional change. From these 20 values a single summary score is calculated based on the total fraction of substitutions predicted to have a deleterious effect on function, taken to be those with a score > 40. The summary scores are used to generate a red to blue color map, indicating residues with highest to least functional importance, respectively.

Mutational Score. This feature set is based on the same 20 scores above, but calculates the single summary score for each residue as the average of the individual scores for each of the 20 standard amino acids.

UniProt Features

UniProt features are curated annotations, and therefore largely complement the automatically generated PredictProtein features. In Aquaria, for each protein sequence, the UniProt feature collection is fetched directly by the browser via:

```
https://www.uniprot.org/uniprot/:uniprot_id.xml
```

CATH Features

For this work, we further enhanced Aquaria to include CATH domain annotations (Dawson et al., 2017). For most protein sequences, Aquaria fetches these annotations directly from the browser via APIs given at:

```
https://github.com/UCLOrengoGroup/cath-api-docs
```

For SARS-CoV-2 proteins, however, CATH annotations are not yet fully available via the above APIs. In this work, we used a pre-release version of these annotations, derived by scanning the UniProt pre-release sequences against the CATH-Gene3D v4.3 FunFams HMM library (Dawson et al., 2017; Lewis et al., 2018) using HMMsearch with inclusion thresholds cut-offs (Mistry et al., 2013). Domain assignments were obtained using cath-resolve-hits and curated manually (Lewis et al., 2019). For the SARS-CoV-2 sequences, these data are fetched directly from the browser via:

```
https://aquaria.ws/covid19cath/P0DTC2
```

Two CATH feature sets were used in this work:

Superfamilies. These identify regions of protein sequences across a wide variety of organisms that are expected to have very similar 3D structures and to have general biological functions in common.

Functional Families. Also known as FunFams, these domains partition each superfamily into subsets expected to have more specific biological functions in common (Dawson et al., 2017; Lewis et al., 2018).

When examining a specific superfamily or functional family domain in Aquaria, the browser uses additional CATH API endpoints (see link above) to create compact, interactive data visualizations that give access to detailed information on the biological function and phylogenetic distribution of proteins containing this domain.

Average Feature Scores

In the results, we include average scores for some of the above feature sets, calculated using the APIs defined above. The average scores given for NSP3 regions were calculated using PP1a features. The whole-protein average scores given for each of the 15 PP1ab proteins were calculated using PP1ab features.

Intermolecular Contacts

We determined the set of residues comprising intermolecular contacts by selecting all residues for one protein, then using the 'Neighbours' function in jolecule. This highlights all residues in which any atom is within 5 Angstrom.

Structure Matrix

We created a web page featuring a matrix of structures that summarizes the total number of matches found for each viral protein sequence, and allows navigation to the corresponding Aquaria page for each protein.

Structure Coverage Map

We created an additional web page with a layout derived from the organization of the viral genome. For each region of the viral proteome with matching structures, we selected a single representative structure (Figure 2), primarily based on identity to the SARS-CoV-2 sequence. However, in some cases, representatives were chosen that best illustrated the consensus biological assembly seen across all related matching structures or showed the simplest assembly. Under the name of each viral protein, the total number of matching structures found in PSSH2 (O'Donoghue et al., 2015), is indicated. Below each structure, a tree graph is drawn to indicate cases of mimicry (where the viral sequence aligns onto human proteins), hijacking (where the matching structures show binding between viral protein and human proteins, DNA, RNA, antibodies, or inhibitory factors), or interaction with other viral proteins. When these tree graphs are missing, none of the matching structures meet these criteria.

ACKNOWLEDGEMENTS

Thanks to Tim Mercer and Giulia Wang (Garvan Institute, Australia), Phil Austin (University of Sydney, Australia) and Lucy van Dorp (UCL Genetics Institute, UK) for helpful feedback and discussions, to Ian Sillitoe (UCL, UK) for helpful advice regarding the CATH API, to Michael Bernhofer and Maria Littmann (TUM) for advice regarding the PredictProtein server. We are very grateful to Max Ott (CSIRO, Australia) for detailed advice on improving the performance and reliability of the Aquaria web application, and to Tim Karl (TU Munich, Germany) for contributing towards this same goal, even during parental leave.

REFERENCES

Adedeji, A.O., Marchand, B., te Velthuis, A.J.W., Snijder, E.J., Weiss, S., Eoff, R.L., Singh, K., Sarafianos, S.G., 2012. Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. PLoS ONE 7, e36521. https://doi.org/10.1371/journal.pone.0036521

Almeida, M.S., Johnson, M.A., Herrmann, T., Geralt, M., Wüthrich, K., 2007. Novel β-Barrel Fold

in the Nuclear Magnetic Resonance Structure of the Replicase Nonstructural Protein 1 from the Severe Acute Respiratory Syndrome Coronavirus. J. Virol. 81, 3151–3161. https://doi.org/10.1128/JVI.01939-06

Angelini, M.M., Neuman, B.W., Buchmeier, M.J., 2014. Untangling Membrane Rearrangement in the Nidovirales. DNA Cell Biol. 33, 122–127. https://doi.org/10.1089/dna.2013.2304

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, W529–W533. https://doi.org/10.1093/nar/gkq399

Aviv, A., 2020. Telomeres and COVID‐19. FASEB J. 34, 7247–7252. https://doi.org/10.1096/fj.202001025

Bailey-Elkin, B.A., Knaap, R.C.M., Johnson, G.G., Dalebout, T.J., Ninaber, D.K., van Kasteren, P.B., Bredenbeek, P.J., Snijder, E.J., Kikkert, M., Mark, B.L., 2014. Crystal Structure of the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Papain-like Protease Bound to Ubiquitin Facilitates Targeted Disruption of Deubiquitinating Activity to Demonstrate Its Role in Innate Immune Suppression. J. Biol. Chem. 289, 34667–34682. https://doi.org/10.1074/jbc.M114.609644

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D., Baker, S.C., 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. J. Virol. 79, 15189–15198. https://doi.org/10.1128/JVI.79.24.15189-15198.2005

Bauer, G., 2020. The variability of the serological response to SARS‐corona virus‐2: Potential resolution of ambiguity through determination of avidity (functional affinity). J. Med. Virol. jmv.26262. https://doi.org/10.1002/jmv.26262

Békés, M., van der Heden van Noort, G.J., Ekkebus, R., Ovaa, H., Huang, T.T., Lima, C.D., 2016. Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. Mol. Cell 62, 572–585. https://doi.org/10.1016/j.molcel.2016.04.016

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–242. https://doi.org/10.1093/nar/28.1.235

Bernhofer, M., Kloppmann, E., Reeb, J., Rost, B., 2016. TMSEG: Novel prediction of transmembrane helices. Proteins Struct. Funct. Bioinforma. 84, 1706–1716. https://doi.org/10.1002/prot.25155

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E.J., Canard, B., Decroly, E., 2010. In Vitro Reconstitution of SARS-Coronavirus mRNA Cap Methylation. PLoS Pathog. 6, e1000863. https://doi.org/10.1371/journal.ppat.1000863

Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., Decroly, E., 2012. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proc. Natl. Acad. Sci. 109, 9372–9377. https://doi.org/10.1073/pnas.1201130109

Bouvet, M., Lugari, A., Posthuma, C.C., Zevenhoven, J.C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J.-C., Lécine, P., Pfefferle, S., Drosten, C., Snijder, E.J., Decroly, E., Morelli, X., 2014. Coronavirus Nsp10, a Critical Co-factor for Activation of Multiple Replicative Enzymes. J. Biol. Chem. 289, 25783–25796. https://doi.org/10.1074/jbc.M114.577353

Callaway, E., Ledford, H., Mallapaty, S., 2020. Six months of coronavirus: the mysteries scientists are still racing to solve. Nature 583, 178–179. https://doi.org/10.1038/d41586-020-01989-z

Camargo, S.M.R., Singer, D., Makrides, V., Huggel, K., Pos, K.M., Wagner, C.A., Kuba, K., Danilczyk, U., Skovby, F., Kleta, R., Penninger, J.M., Verrey, F., 2009. Tissue-Specific Amino Acid Transporter Partners ACE2 and Collectrin Differentially Interact With

Hartnup Mutations. Gastroenterology 136, 872-882.e3. https://doi.org/10.1053/j.gastro.2008.10.055
Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2013. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Isr. J. Chem. 53, 199–206. https://doi.org/10.1002/ijch.201200096
Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., Conti, E., 2011. Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. Mol. Cell 41, 693–703. https://doi.org/10.1016/j.molcel.2011.02.010
Chatterjee, A., Johnson, M.A., Serrano, P., Pedrini, B., Joseph, J.S., Neuman, B.W., Saikatendu, K., Buchmeier, M.J., Kuhn, P., Wüthrich, K., 2009. Nuclear Magnetic Resonance Structure Shows that the Severe Acute Respiratory Syndrome Coronavirus-Unique Domain Contains a Macrodomain Fold. J. Virol. 83, 1823–1836. https://doi.org/10.1128/JVI.01781-08
Chen, D., Vollmar, M., Rossi, M.N., Phillips, C., Kraehenbuehl, R., Slade, D., Mehrotra, P.V., von Delft, F., Crosthwaite, S.K., Gileadi, O., Denu, J.M., Ahel, I., 2011. Identification of Macrodomain Proteins as Novel *O* -Acetyl-ADP-ribose Deacetylases. J. Biol. Chem. 286, 13261–13271. https://doi.org/10.1074/jbc.M110.206771
Chen, Y., Savinov, S.N., Mielech, A.M., Cao, T., Baker, S.C., Mesecar, A.D., 2015. X-ray Structural and Functional Studies of the Three Tandemly Linked Domains of Non-structural Protein 3 (nsp3) from Murine Hepatitis Virus Reveal Conserved Functions. J. Biol. Chem. 290, 25293–25306. https://doi.org/10.1074/jbc.M115.662130
Chu, H., Zhou, J., Wong, B.H.-Y., Li, Cun, Chan, J.F.-W., Cheng, Z.-S., Yang, D., Wang, D., Lee, A.C.-Y., Li, Chuangen, Yeung, M.-L., Cai, J.-P., Chan, I.H.-Y., Ho, W.-K., To, K.K.-W., Zheng, B.-J., Yao, Y., Qin, C., Yuen, K.-Y., 2016. Middle East Respiratory Syndrome Coronavirus Efficiently Infects Human Primary T Lymphocytes and Activates the Extrinsic and Intrinsic Apoptosis Pathways. J. Infect. Dis. 213, 904–914. https://doi.org/10.1093/infdis/jiv380
Clerici, M., Mourão, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., Cusack, S., 2009. Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J. 28, 2293–2306. https://doi.org/10.1038/emboj.2009.175
Cornillez-Ty, C.T., Liao, L., Yates, J.R., Kuhn, P., Buchmeier, M.J., 2009. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. J. Virol. 83, 10314–10318. https://doi.org/10.1128/JVI.00842-09
Daczkowski, C.M., Dzimianski, J.V., Clasman, J.R., Goodwin, O., Mesecar, A.D., Pegan, S.D., 2017. Structural Insights into the Interaction of Coronavirus Papain-Like Proteases and Interferon-Stimulated Gene Product 15 from Different Species. J. Mol. Biol. 429, 1661–1683. https://doi.org/10.1016/j.jmb.2017.04.011
Davey, N.E., Travé, G., Gibson, T.J., 2011. How viruses hijack cell regulation. Trends Biochem. Sci. 36, 159–169. https://doi.org/10.1016/j.tibs.2010.10.002
Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A., Sillitoe, I., 2017. CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res. 45, D289–D295. https://doi.org/10.1093/nar/gkw1098
De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Lührmann, R., Pena, V., 2015. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. Nat. Struct. Mol. Biol. 22, 138–144. https://doi.org/10.1038/nsmb.2951
Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., Canard, B.,

2011. Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2′-O-Methyltransferase nsp10/nsp16 Complex. PLoS Pathog. 7, e1002059. https://doi.org/10.1371/journal.ppat.1002059

Dehghani-Tafti, S., Levdikov, V., Antson, A.A., Bax, B., Sanders, C.M., 2019. Structural and functional analysis of the nucleotide and DNA binding activities of the human PIF1 helicase. Nucleic Acids Res. 47, 3208–3222. https://doi.org/10.1093/nar/gkz028

Deng, Z., Lehmann, K.C., Li, X., Feng, C., Wang, G., Zhang, Q., Qi, X., Yu, L., Zhang, X., Feng, W., Wu, W., Gong, P., Tao, Y., Posthuma, C.C., Snijder, E.J., Gorbalenya, A.E., Chen, Z., 2014. Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. Nucleic Acids Res. 42, 3464–3477. https://doi.org/10.1093/nar/gkt1310

Dzimianski, J.V., Scholte, F.E.M., Bergeron, É., Pegan, S.D., 2019. ISG15: It's Complicated. J. Mol. Biol. 431, 4203–4216. https://doi.org/10.1016/j.jmb.2019.03.013

Elde, N.C., Malik, H.S., 2009. The evolutionary conundrum of pathogen mimicry. Nat. Rev. Microbiol. 7, 787–797. https://doi.org/10.1038/nrmicro2222

Ferrer-Orta, C., Sierra, M., Agudo, R., de la Higuera, I., Arias, A., Pérez-Luque, R., Escarmís, C., Domingo, E., Verdaguer, N., 2010. Structure of Foot-and-Mouth Disease Virus Mutant Polymerases with Reduced Sensitivity to Ribavirin. J. Virol. 84, 6188–6199. https://doi.org/10.1128/JVI.02420-09

Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., Imbert, I., 2018. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. Proc. Natl. Acad. Sci. 115, E162–E171. https://doi.org/10.1073/pnas.1718806115

Forst, A.H., Karlberg, T., Herzog, N., Thorsell, A.-G., Gross, A., Feijs, K.L.H., Verheugd, P., Kursula, P., Nijmeijer, B., Kremmer, E., Kleine, H., Ladurner, A.G., Schüler, H., Lüscher, B., 2013. Recognition of Mono-ADP-Ribosylated ARTD10 Substrates by ARTD8 Macrodomains. Structure 21, 462–475. https://doi.org/10.1016/j.str.2012.12.019

Frieman, M., Yount, B., Heise, M., Kopecky-Bromberg, S.A., Palese, P., Baric, R.S., 2007. Severe Acute Respiratory Syndrome Coronavirus ORF6 Antagonizes STAT1 Function by Sequestering Nuclear Import Factors on the Rough Endoplasmic Reticulum/Golgi Membrane. J. Virol. 81, 9812–9824. https://doi.org/10.1128/JVI.01012-07

Gong, P., Kortus, M.G., Nix, J.C., Davis, R.E., Peersen, O.B., 2013. Structures of Coxsackievirus, Rhinovirus, and Poliovirus Polymerase Elongation Complexes Solved by Engineering RNA Mediated Crystal Contacts. PLoS ONE 8, e60272. https://doi.org/10.1371/journal.pone.0060272

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., Tummino, T.A., Hüttenhain, R., Kaake, R.M., Richards, A.L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B.J., Braberg, H., Fabius, J.M., Eckhardt, M., Soucheray, M., Bennett, M.J., Cakir, M., McGregor, M.J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z.Z.C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I.T., Melnyk, J.E., Chorba, J.S., Lou, K., Dai, S.A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C.J.P., Perica, T., Pilla, K.B., Ganesan, S.J., Saltzberg, D.J., Rakesh, R., Liu, X., Rosenthal, S.B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X.-P., Liu, Y., Wankowicz, S.A., Bohn, M., Safari, M., Ugur, F.S., Koh, C., Savar, N.S., Tran, Q.D., Shengjuler, D., Fletcher, S.J., O'Neal, M.C., Cai, Y., Chang, J.C.J., Broadhurst, D.J., Klippsten, S., Sharp, P.P., Wenzell, N.A., Kuzuoglu-Ozturk, D., Wang, H.-Y., Trenker, R., Young, J.M., Cavero, D.A., Hiatt, J., Roth, T.L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R.M., Frankel, A.D., Rosenberg, O.S., Verba, K.A., Agard, D.A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O.,

d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H.S., Fujimori, D.G., Ideker, T., Craik, C.S., Floor, S.N., Fraser, J.S., Gross, J.D., Sali, A., Roth, B.L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K.M., Shoichet, B.K., Krogan, N.J., 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459–468. https://doi.org/10.1038/s41586-020-2286-9

Graham, R.L., Baric, R.S., 2010. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. J. Virol. 84, 3134–3146. https://doi.org/10.1128/JVI.01394-09

Grunewald, M.E., Fehr, A.R., Athmer, J., Perlman, S., 2018. The coronavirus nucleocapsid protein is ADP-ribosylated. Virology 517, 62–68. https://doi.org/10.1016/j.virol.2017.11.020

Gu, J., Gong, E., Zhang, B., Zheng, J., Gao, Z., Zhong, Y., Zou, W., Zhan, J., Wang, S., Xie, Z., Zhuang, H., Wu, B., Zhong, H., Shao, H., Fang, W., Gao, D., Pei, F., Li, X., He, Z., Xu, D., Shi, X., Anderson, V.M., Leong, A.S.-Y., 2005. Multiple organ infection and the pathogenesis of SARS. J. Exp. Med. 202, 415–424. https://doi.org/10.1084/jem.20050828

Hakansson-McReynolds, S., Jiang, S., Rong, L., Caffrey, M., 2006. Solution Structure of the Severe Acute Respiratory Syndrome-Coronavirus Heptad Repeat 2 Domain in the Prefusion State. J. Biol. Chem. 281, 11965–11971. https://doi.org/10.1074/jbc.M601174200

Hänel, K., Stangler, T., Stoldt, M., Willbold, D., 2006. Solution structure of the X4 protein coded by the SARS related coronavirus reveals an immunoglobulin like fold and suggests a binding activity to integrin I domains. J. Biomed. Sci. 13, 281–293. https://doi.org/10.1007/s11373-005-9043-9

Hecht, M., Bromberg, Y., Rost, B., 2015. Better prediction of functional effects for sequence variants. BMC Genomics 16, S1. https://doi.org/10.1186/1471-2164-16-S8-S1

Heinrich, J., Kaur, S., O'Donoghue, S., 2015. Evaluating the Effectiveness of Color to Convey Alignment Quality in Macromolecular Structures. Presented at the Symposium on Big Data Visual Analytics, IEEE, Hobart, Australia.

Held, T., Basler, M., Knobeloch, K., Groettrup, M., 2020. Evidence for an involvement of the ubiquitin-like modifier ISG15 in MHC class I antigen presentation. Eur. J. Immunol. eji.202048646. https://doi.org/10.1002/eji.202048646

Heo, L., Feig, M., 2020. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. bioRxiv 2020.03.25.008904. https://doi.org/10.1101/2020.03.25.008904

Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. Nature 584, 154–156. https://doi.org/10.1038/s41586-020-2368-8

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M.A., Drosten, C., Pöhlmann, S., 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Hottiger, M.O., 2015. SnapShot: ADP-Ribosylation Signaling. Mol. Cell 58, 1134-1134.e1. https://doi.org/10.1016/j.molcel.2015.06.001

Iwata, H., Goettsch, C., Sharma, A., Ricchiuto, P., Goh, W.W.B., Halu, A., Yamada, I., Yoshida, H., Hara, T., Wei, M., Inoue, N., Fukuda, D., Mojcher, A., Mattson, P.C., Barabási, A.-L., Boothby, M., Aikawa, E., Singh, S.A., Aikawa, M., 2016. PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nat. Commun. 7, 12849. https://doi.org/10.1038/ncomms12849

Jaimes, J.A., André, N.M., Chappie, J.S., Millet, J.K., Whittaker, G.R., 2020. Phylogenetic

Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. J. Mol. Biol. 432, 3309–3325. https://doi.org/10.1016/j.jmb.2020.04.009

Jankevicius, G., Hassler, M., Golia, B., Rybin, V., Zacharias, M., Timinszky, G., Ladurner, A.G., 2013. A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. Nat. Struct. Mol. Biol. 20, 508–514. https://doi.org/10.1038/nsmb.2523

Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., Rao, Z., 2019. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. Nucleic Acids Res. 47, 6538–6550. https://doi.org/10.1093/nar/gkz409

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, Xinglou, You, T., Liu, Xiaoce, Yang, Xiuna, Bai, F., Liu, H., Liu, Xiang, Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., Yang, H., 2020. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 582, 289–293. https://doi.org/10.1038/s41586-020-2223-y

Johnson, M.A., Chatterjee, A., Neuman, B.W., Wüthrich, K., 2010a. SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding. J. Mol. Biol. 400, 724–742. https://doi.org/10.1016/j.jmb.2010.05.027

Johnson, M.A., Jaudzems, K., Wüthrich, K., 2010b. NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. J. Mol. Biol. 402, 619–628. https://doi.org/10.1016/j.jmb.2010.07.043

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M., Moy, K., Yadav, M.K., Velasquez, J., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2006. Crystal Structure of Nonstructural Protein 10 from the Severe Acute Respiratory Syndrome Coronavirus Reveals a Novel Fold with Two Zinc-Binding Motifs. J. Virol. 80, 7894–7901. https://doi.org/10.1128/JVI.00467-06

Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., Makino, S., 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat. Struct. Mol. Biol. 16, 1134–1140. https://doi.org/10.1038/nsmb.1680

Kawabata, T., 2016. HOMCOS: an updated server to search and model complex 3D structures. J. Struct. Funct. Genomics 17, 83–99. https://doi.org/10.1007/s10969-016-9208-y

Kirchdoerfer, R.N., Ward, A.B., 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat. Commun. 10, 2342. https://doi.org/10.1038/s41467-019-10280-3

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., Moult, J., 2019. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins Struct. Funct. Bioinforma. 87, 1011–1020. https://doi.org/10.1002/prot.25823

Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K., Ladurner, A.G., 2005. Splicing regulates NAD metabolite binding to histone macroH2A. Nat. Struct. Mol. Biol. 12, 624–625. https://doi.org/10.1038/nsmb956

Lei, J., Kusov, Y., Hilgenfeld, R., 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res. 149, 58–74. https://doi.org/10.1016/j.antiviral.2017.11.001

Lewis, T.E., Sillitoe, I., Dawson, N., Lam, S.D., Clarke, T., Lee, D., Orengo, C., Lees, J., 2018. Gene3D: Extensive prediction of globular domains in proteins. Nucleic Acids Res. 46, D435–D439. https://doi.org/10.1093/nar/gkx1069

Lewis, T.E., Sillitoe, I., Lees, J.G., 2019. cath-resolve-hits: a new tool that resolves domain matches suspiciously quickly. Bioinformatics 35, 1766–1767. https://doi.org/10.1093/bioinformatics/bty863

Li, J.-Y., Liao, C.-H., Wang, Q., Tan, Y.-J., Luo, R., Qiu, Y., Ge, X.-Y., 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway.

Virus Res. 286, 198074. https://doi.org/10.1016/j.virusres.2020.198074
Li, Y., Surya, W., Claudine, S., Torres, J., 2014. Structure of a Conserved Golgi Complex-targeting Signal in Coronavirus Envelope Proteins. J. Biol. Chem. 289, 12535–12549. https://doi.org/10.1074/jbc.M114.560094
Li, Y., Zhang, Z., Yang, L., Lian, X., Xie, Y., Li, S., Xin, S., Cao, P., Lu, J., 2020. The MERS-CoV Receptor DPP4 as a Candidate Binding Target of the SARS-CoV-2 Spike. iScience 23, 101160. https://doi.org/10.1016/j.isci.2020.101160
Lim, S.C., Bowler, M.W., Lai, T.F., Song, H., 2012. The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. Nucleic Acids Res. 40, 11009–11022. https://doi.org/10.1093/nar/gks792
Lu, W., Zheng, B.-J., Xu, K., Schwarz, W., Du, L., Wong, C.K.L., Chen, J., Duan, S., Deubel, V., Sun, B., 2006. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. Proc. Natl. Acad. Sci. 103, 12540–12545. https://doi.org/10.1073/pnas.0605402103
Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., Rao, Z., 2015. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. Proc. Natl. Acad. Sci. 112, 9436–9441. https://doi.org/10.1073/pnas.1508686112
Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E.A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A.E., de Lamballerie, X., Canard, B., 2009. The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket. J. Virol. 83, 6534–6545. https://doi.org/10.1128/JVI.00189-09
Masmaliyeva, R.C., Murshudov, G.N., 2019. Analysis and validation of macromolecular B values. Acta Crystallogr. Sect. Struct. Biol. 75, 505–518. https://doi.org/10.1107/S2059798319004807
Mayrose, I., Graur, D., Ben-Tal, N., Pupko, T., 2004. Comparison of Site-Specific Rate-Inference Methods for Protein Sequences: Empirical Bayesian Methods Are Superior. Mol. Biol. Evol. 21, 1781–1791. https://doi.org/10.1093/molbev/msh194
Meier, C., Aricescu, A.R., Assenberg, R., Aplin, R.T., Gilbert, R.J.C., Grimes, J.M., Stuart, D.I., 2006. The Crystal Structure of ORF-9b, a Lipid Binding Protein from the SARS Coronavirus. Structure 14, 1157–1165. https://doi.org/10.1016/j.str.2006.05.012
Miknis, Z.J., Donaldson, E.F., Umland, T.C., Rimmer, R.A., Baric, R.S., Schultz, L.W., 2009. Severe Acute Respiratory Syndrome Coronavirus nsp9 Dimerization Is Essential for Efficient Viral Growth. J. Virol. 83, 3007–3018. https://doi.org/10.1128/JVI.01505-08
Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., Ziebuhr, J., 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc. Natl. Acad. Sci. 103, 5108–5113. https://doi.org/10.1073/pnas.0508200103
Mirdita, M., von den Driesch, L., Galiez, C., Martin, M.J., Söding, J., Steinegger, M., 2017. Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res. 45, D170–D176. https://doi.org/10.1093/nar/gkw1081
Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41, e121. https://doi.org/10.1093/nar/gkt263
Murphy, P.M., 1993. Molecular mimicry and the generation of host defense protein diversity. Cell 72, 823–826. https://doi.org/10.1016/0092-8674(93)90571-7
Nakagawa, K., Lokugamage, K.G., Makino, S., 2016. Chapter Five - Viral and Cellular mRNA Translation in Coronavirus-Infected Cells, in: Ziebuhr, J. (Ed.), Coronaviruses, Advances in Virus Research. Academic Press, pp. 165–192. https://doi.org/10.1016/bs.aivir.2016.08.001
Nelson, C.A., Pekosz, A., Lee, C.A., Diamond, M.S., Fremont, D.H., 2005. Structure and

Intracellular Targeting of the SARS-Coronavirus Orf7a Accessory Protein. Structure 13, 75–85. https://doi.org/10.1016/j.str.2004.10.010
O'Donoghue, S.I., Baldi, B.F., Clark, S.J., Darling, A.E., Hogan, J.M., Kaur, S., Maier-Hein, L., McCarthy, D.J., Moore, W.J., Stenau, E., Swedlow, J.R., Vuong, J., Procter, J.B., 2018. Visualization of Biomedical Data. Annu. Rev. Biomed. Data Sci. 1, 275–304. https://doi.org/10.1146/annurev-biodatasci-080917-013424
O'Donoghue, S.I., Sabir, K.S., Kalemanov, M., Stolte, C., Wellmann, B., Ho, V., Roos, M., Perdigão, N., Buske, F.A., Heinrich, J., Rost, B., Schafferhans, A., 2015. Aquaria: simplifying discovery and insight from protein structures. Nat. Methods 12, 98–99. https://doi.org/10.1038/nmeth.3258
O'Sullivan, J., Tedim Ferreira, M., Gagné, J.-P., Sharma, A.K., Hendzel, M.J., Masson, J.-Y., Poirier, G.G., 2019. Emerging roles of eraser enzymes in the dynamic control of protein ADP-ribosylation. Nat. Commun. 10, 1182. https://doi.org/10.1038/s41467-019-08859-x
Pan, J., Peng, X., Gao, Y., Li, Z., Lu, X., Chen, Y., Ishaq, M., Liu, D., DeDiego, M.L., Enjuanes, L., Guo, D., 2008. Genome-Wide Analysis of Protein-Protein Interactions and Involvement of Viral Proteins in SARS-CoV Replication. PLOS ONE 3, e3299. https://doi.org/10.1371/journal.pone.0003299
Perdigão, N., Heinrich, J., Stolte, C., Sabir, K.S., Buckley, M.J., Tabor, B., Signal, B., Gloss, B.S., Hammang, C.J., Rost, B., Schafferhans, A., O'Donoghue, S.I., 2015. Unexpected features of the dark proteome. Proc. Natl. Acad. Sci. 112, 15898–15903. https://doi.org/10.1073/pnas.1508380112
Peti, W., Johnson, M.A., Herrmann, T., Neuman, B.W., Buchmeier, M.J., Nelson, M., Joseph, J., Page, R., Stevens, R.C., Kuhn, P., Wüthrich, K., 2005. Structural Genomics of the Severe Acute Respiratory Syndrome Coronavirus: Nuclear Magnetic Resonance Structure of the Protein nsP7. J. Virol. 79, 12905–12913. https://doi.org/10.1128/JVI.79.20.12905-12913.2005
Piplani, S., Singh, P.K., Winkler, D.A., Petrovsky, N., 2020. In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus. ArXiv200506199 Q-Bio.
Rack, J.G.M., Perina, D., Ahel, I., 2016. Macrodomains: Structure, Function, Evolution, and Catalytic Activities. Annu. Rev. Biochem. 85, 431–454. https://doi.org/10.1146/annurev-biochem-060815-014935
Radzikowska, U., Ding, M., Tan, G., Zhakparov, D., Peng, Y., Wawrzyniak, P., Wang, M., Li, S., Morita, H., Altunbulakli, C., Reiger, M., Neumann, A.U., Lunjani, N., Traidl-Hoffmann, C., Nadeau, K.C., O'Mahony, L., Akdis, C., Sokolowska, M., 2020. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. Allergy all.14429. https://doi.org/10.1111/all.14429
Ratia, K., Kilianski, A., Baez-Santos, Y.M., Baker, S.C., Mesecar, A., 2014. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. PLoS Pathog. 10, e1004113. https://doi.org/10.1371/journal.ppat.1004113
Ricagno, S., Egloff, M.-P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J., Canard, B., 2006. Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. Proc. Natl. Acad. Sci. 103, 11892–11897. https://doi.org/10.1073/pnas.0601708103
Rodrigues, J.P., Barrera-Vilarmau, S., Teixeira, J.M., Seckel, E., Kastritis, P., Levitt, M., 2020. Insights on cross-species transmission of SARS-CoV-2 from structural modeling (preprint). Bioinformatics. https://doi.org/10.1101/2020.06.05.136861
Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., 2004. Protein Structure Prediction Using Rosetta, in: Methods in Enzymology, Numerical Computer Methods, Part D. Academic Press, pp. 66–93. https://doi.org/10.1016/S0076-6879(04)83004-0

Saikatendu, K.S., Joseph, J.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2007. Ribonucleocapsid Formation of Severe Acute Respiratory Syndrome Coronavirus through Molecular Action of the N-Terminal Domain of N Protein. J. Virol. 81, 3913–3921. https://doi.org/10.1128/JVI.02236-06

Schäfer, A., Baric, R., 2017. Epigenetic Landscape during Coronavirus Infection. Pathogens 6, 8. https://doi.org/10.3390/pathogens6010008

Schafferhans, A., O'Donoghue, S.I., Heinzinger, M., Rost, B., 2018. Dark Proteins Important for Cellular Function. PROTEOMICS 18, 1800227. https://doi.org/10.1002/pmic.201800227

Schlessinger, A., Punta, M., Yachdav, G., Kajan, L., Rost, B., 2009. Improved Disorder Prediction by Combination of Orthogonal Approaches. PLOS ONE 4, e4433. https://doi.org/10.1371/journal.pone.0004433

Schlessinger, A., Rost, B., 2005. Protein flexibility and rigidity predicted from sequence. Proteins Struct. Funct. Bioinforma. 61, 115–126. https://doi.org/10.1002/prot.20587

Schlessinger, A., Yachdav, G., Rost, B., 2006. PROFbval: predict flexible and rigid residues in proteins. Bioinformatics 22, 891–893. https://doi.org/10.1093/bioinformatics/btl032

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D.T., Silver, D., Kavukcuoglu, K., Hassabis, D., 2020. Improved protein structure prediction using potentials from deep learning. Nature 577, 706–710. https://doi.org/10.1038/s41586-019-1923-7

Serrano, P., Johnson, M.A., Almeida, M.S., Horst, R., Herrmann, T., Joseph, J.S., Neuman, B.W., Subramanian, V., Saikatendu, K.S., Buchmeier, M.J., Stevens, R.C., Kuhn, P., Wüthrich, K., 2007. Nuclear Magnetic Resonance Structure of the N-Terminal Domain of Nonstructural Protein 3 from the Severe Acute Respiratory Syndrome Coronavirus. J. Virol. 81, 12049–12060. https://doi.org/10.1128/JVI.00969-07

Serrano, P., Johnson, M.A., Chatterjee, A., Neuman, B.W., Joseph, J.S., Buchmeier, M.J., Kuhn, P., Wüthrich, K., 2009. Nuclear Magnetic Resonance Structure of the Nucleic Acid-Binding Domain of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 3. J. Virol. 83, 12998–13008. https://doi.org/10.1128/JVI.01253-09

Shi, C.-S., Qi, H.-Y., Boularan, C., Huang, N.-N., Abu-Asab, M., Shelhamer, J.H., Kehrl, J.H., 2014. SARS-Coronavirus Open Reading Frame-9b Suppresses Innate Immunity by Targeting Mitochondria and the MAVS/TRAF3/TRAF6 Signalosome. J. Immunol. 193, 3080–3089. https://doi.org/10.4049/jimmunol.1303196

Song, W., Gui, M., Wang, X., Xiang, Y., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLOS Pathog. 14, e1007236. https://doi.org/10.1371/journal.ppat.1007236

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., Söding, J., 2019. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20, 473. https://doi.org/10.1186/s12859-019-3019-7

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M., Rao, Z., 2006. Dodecamer Structure of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10. J. Virol. 80, 7902–7908. https://doi.org/10.1128/JVI.00483-06

Subissi, L., Imbert, I., Ferron, F., Collet, A., Coutard, B., Decroly, E., Canard, B., 2014. SARS-CoV ORF1b-encoded nonstructural proteins 12–16: Replicative enzymes as antiviral targets. Antiviral Res. 101, 122–130. https://doi.org/10.1016/j.antiviral.2013.11.006

Suhrer, S.J., Wiederstein, M., Gruber, M., Sippl, M.J., 2009. COPS—a novel workbench for explorations in fold space. Nucleic Acids Res. 37, W539–W544. https://doi.org/10.1093/nar/gkp411

Surjit, M., Kumar, R., Mishra, R.N., Reddy, M.K., Chow, V.T.K., Lal, S.K., 2005. The Severe

Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein Is Phosphorylated and Localizes in the Cytoplasm by 14-3-3-Mediated Translocation. J. Virol. 79, 11476–11486. https://doi.org/10.1128/JVI.79.17.11476-11486.2005

Surya, W., Li, Y., Torres, J., 2018. Structural model of the SARS coronavirus E channel in LMPG micelles. Biochim. Biophys. Acta BBA - Biomembr. 1860, 1309–1317. https://doi.org/10.1016/j.bbamem.2018.02.017

Tan, L., Wang, Qi, Zhang, D., Ding, J., Huang, Q., Tang, Y.-Q., Wang, Qiongshu, Miao, H., 2020. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct. Target. Ther. 5, 33. https://doi.org/10.1038/s41392-020-0148-4

Taylor, J.K., Coleman, C.M., Postel, S., Sisk, J.M., Bernbaum, J.G., Venkataraman, T., Sundberg, E.J., Frieman, M.B., 2015. Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of Glycosylation Interference. J. Virol. 89, 11820–11833. https://doi.org/10.1128/JVI.02274-15

te Velthuis, A.J.W., van den Worm, S.H.E., Snijder, E.J., 2012. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res. 40, 1737–1747. https://doi.org/10.1093/nar/gkr893

The UniProt Consortium, 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Umate, P., Tuteja, N., Tuteja, R., 2011. Genome-wide comprehensive analysis of human helicases. Commun. Integr. Biol. 4, 118–137. https://doi.org/10.4161/cib.4.1.13844

Varga, Z., Flammer, A.J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A.S., Mehra, M.R., Schuepbach, R.A., Ruschitzka, F., Moch, H., 2020. Endothelial cell infection and endotheliitis in COVID-19. The Lancet 395, 1417–1418. https://doi.org/10.1016/S0140-6736(20)30937-5

Vennema, H., Godeke, G.J., Rossen, J.W., Voorhout, W.F., Horzinek, M.C., Opstelten, D.J., Rottier, P.J., 1996. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. EMBO J. 15, 2020–2028. https://doi.org/10.1002/j.1460-2075.1996.tb00553.x

Wada, M., Lokugamage, K.G., Nakagawa, K., Narayanan, K., Makino, S., 2018. Interplay between coronavirus, a cytoplasmic RNA virus, and nonsense-mediated mRNA decay pathway. Proc. Natl. Acad. Sci. 115, E10157–E10166. https://doi.org/10.1073/pnas.1811675115

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058

Wang, N., Rosen, O., Wang, L., Turner, H.L., Stevens, L.J., Corbett, K.S., Bowman, C.A., Pallesen, J., Shi, W., Zhang, Y., Leung, K., Kirchdoerfer, R.N., Becker, M.M., Denison, M.R., Chappell, J.D., Ward, A.B., Graham, B.S., McLellan, J.S., 2019. Structural Definition of a Neutralization-Sensitive Epitope on the MERS-CoV S1-NTD. Cell Rep. 28, 3395-3405.e6. https://doi.org/10.1016/j.celrep.2019.08.052

Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J., Iwamoto, A., Woo, P.C.Y., Yuen, K.-Y., Yan, J., Lu, G., Gao, G.F., 2014. Bat Origins of MERS-CoV Supported by Bat Coronavirus HKU4 Usage of Human Receptor CD26. Cell Host Microbe 16, 328–337. https://doi.org/10.1016/j.chom.2014.08.009

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427

Xia, S., Yan, L., Xu, W., Agrawal, A.S., Algaissi, A., Tseng, C.-T.K., Wang, Q., Du, L., Tan, W.,

Wilson, I.A., Jiang, S., Yang, B., Lu, L., 2019. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Sci. Adv. 5, eaav4580. https://doi.org/10.1126/sciadv.aav4580

Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G., Hilgenfeld, R., 2012. Nonstructural Proteins 7 and 8 of Feline Coronavirus Form a 2:1 Heterotrimer That Exhibits Primer-Independent RNA Polymerase Activity. J. Virol. 86, 4444–4454. https://doi.org/10.1128/JVI.06635-11

Xu, X., Lou, Z., Ma, Y., Chen, X., Yang, Z., Tong, X., Zhao, Q., Xu, Y., Deng, H., Bartlam, M., Rao, Z., 2009. Crystal Structure of the C-Terminal Cytoplasmic Domain of Non-Structural Protein 4 from Mouse Hepatitis Virus A59. PLoS ONE 4, e6217. https://doi.org/10.1371/journal.pone.0006217

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X.C., Liao, M., Bartlam, M., Rao, Z., 2008. Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design. J. Virol. 82, 2515–2527. https://doi.org/10.1128/JVI.02114-07

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., Rost, B., 2014. PredictProtein—an open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 42, W337–W343. https://doi.org/10.1093/nar/gku366

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., Zhou, Q., 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367, 1444–1448. https://doi.org/10.1126/science.abb2762

Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y.-C., Tian, G., Jiang, H.-W., Tao, S.-C., Shen, J., Jiang, Y., Jiang, H., Xu, Y., Zhang, S., Zhang, Y., Xu, H.E., 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science 368, 1499–1504. https://doi.org/10.1126/science.abc1560

Yu, H.-H., Chu, K.-H., Yang, Y.-H., Lee, J.-H., Wang, L.-C., Lin, Y.-T., Chiang, B.-L., 2011. Genetics and Immunopathogenesis of IgA Nephropathy. Clin. Rev. Allergy Immunol. 41, 198–213. https://doi.org/10.1007/s12016-010-8232-0

Yuan, M., Wu, N.C., Zhu, X., Lee, C.-C.D., So, R.T.Y., Lv, H., Mok, C.K.P., Wilson, I.A., 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science 368, 630–633. https://doi.org/10.1126/science.abb7269

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., Rao, Z., 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. Nat. Struct. Mol. Biol. 12, 980–986. https://doi.org/10.1038/nsmb999

Zhang, X., Zhan, X., Yan, C., Zhang, W., Liu, D., Lei, J., Shi, Y., 2019. Structures of the human spliceosomes before and after release of the ligated exon. Cell Res. 29, 274–285. https://doi.org/10.1038/s41422-019-0143-x

Zheng, W., Li, Y., Zhang, C., Pearce, R., Mortuza, S.M., Zhang, Y., 2019. Deep-learning contact-map guided protein structure prediction in CASP13. Proteins 87, 1149–1164. https://doi.org/10.1002/prot.25792

FIGURES

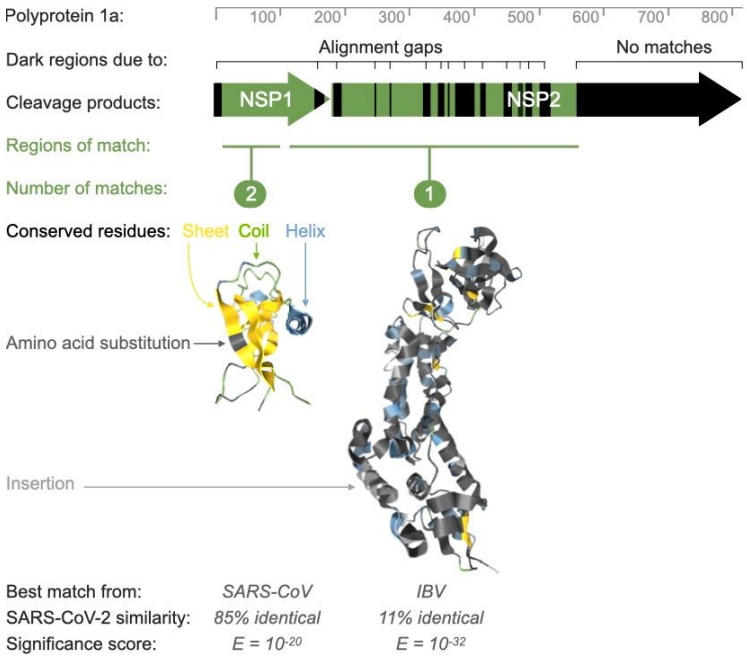

Figure 1. Graphical elements of the structure coverage map

The ruler (top) indicates sequence residue numbering. Dark sequence regions indicate either alignment gaps, or regions that do not match any experimentally-determined 3D structure. Regions with matching structure are indicated in green, and with a representative 3D structure position below, with conserved residues colored by secondary structure, with amino acid substitutions in dark gray, and with insertions in light gray. This coloring makes it clear that the NSP1 model is similar in sequence to SARS-CoV-2, while the NSP2 model is remote. However, a better measure of model significance is the HHblits E-value, calculated by aligning sequence profiles for SARS-CoV-2 onto profiles for each 3D structure. The NSP2 model is more significant as it has a longer alignment.

Made using Aquaria and edited with Keynote.

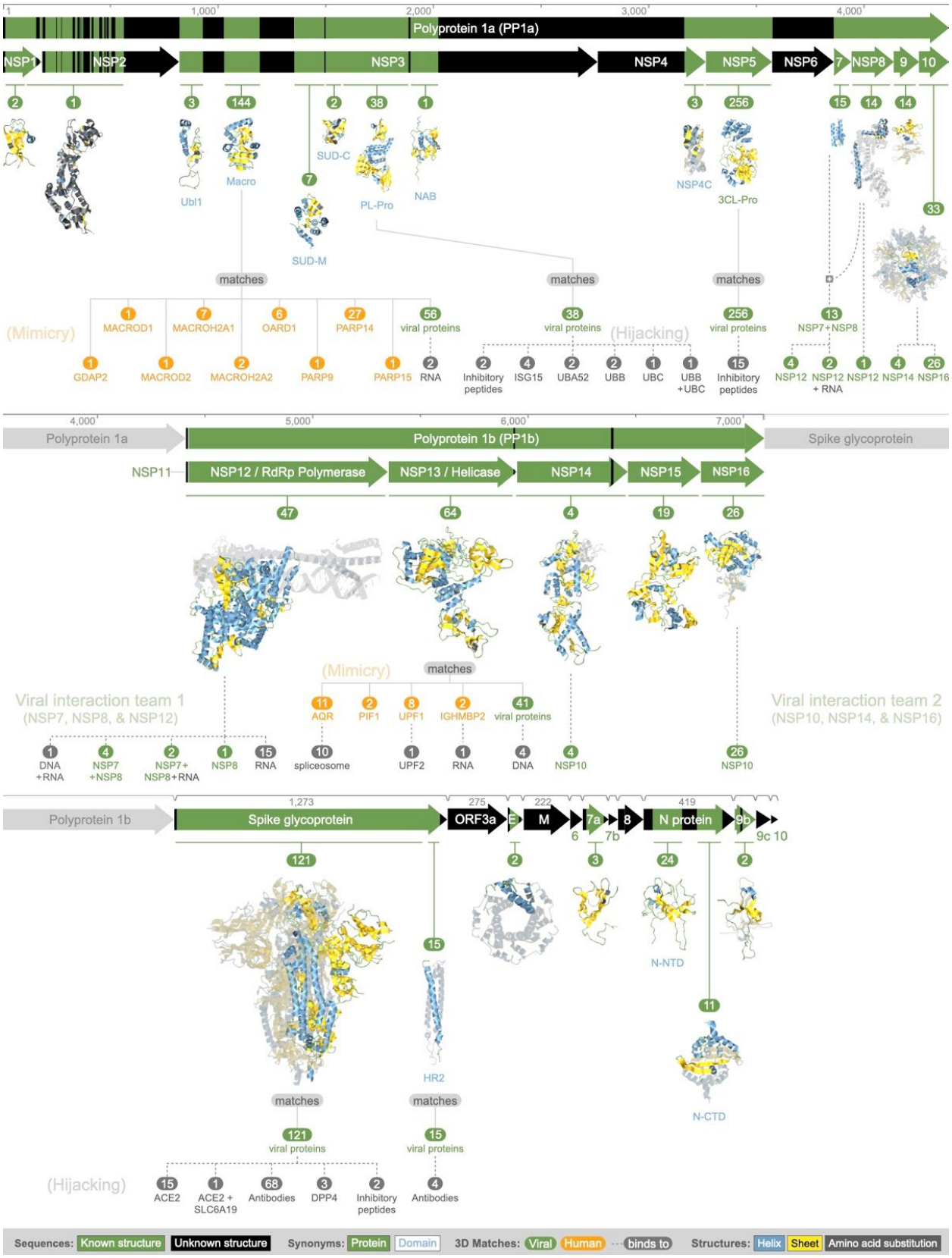
Polyprotein 1a (PP1a)
NSP1
NSP2
NSP3
NSP4
NSP5
NSP6
7
NSP8
9
10
Ubl1
Macro
SUD-C
SUD-M
PL-Pro
NAB
NSP4C
3CL-Pro
matches
(Mimicry)
MACROD1
MACROH2A1
OARD1
PARP14
viral proteins
GDAP2
MACROD2
MACROH2A2
PARP9
PARP15
RNA
(Hijacking)
Inhibitory peptides
ISG15
UBA52
UBB
UBC
UBB +UBC
NSP7+NSP8
NSP12
NSP12 + RNA
NSP14
NSP16
Polyprotein 1a
Polyprotein 1b (PP1b)
Spike glycoprotein
NSP11
NSP12 / RdRp Polymerase
NSP13 / Helicase
NSP14
NSP15
NSP16
Viral interaction team 1
(NSP7, NSP8, & NSP12)
Viral interaction team 2
(NSP10, NSP14, & NSP16)
AQR
PIF1
UPF1
IGHMBP2
DNA +RNA
NSP7 +NSP8
NSP7+ NSP8+RNA
NSP8
spliceosome
UPF2
DNA
NSP10
Polyprotein 1b
ORF3a
E
M
6
7a
7b
8
N protein
9b
9c 10
N-NTD
N-CTD
HR2
ACE2
ACE2 + SLC6A19
Antibodies
DPP4
Sequences: Known structure Unknown structure
Synonyms: Protein Domain
3D Matches: Viral Human binds to
Structures: Helix Sheet Amino acid substitution

Figure 2. SARS-CoV-2 structure coverage map

Integrated visual summary of 3D structural evidence about viral protein states. Proteins are shown as arrows, scaled by sequence length, ordered by genomic location, and divided into three groups: (1) polyprotein 1a proteins (top); (2) polyprotein 1b proteins (middle); and (3) virion and accessory proteins (bottom). Dark coloring indicates sequence regions that do not match any experimentally-determined 3D structure. Regions with matching structure are indicated in green, and with representative 3D structures positioned below (Figure 1), all at about the same scale. NSP3 macro domain and NSP13 may mimic the human proteins shown as orange nodes connected by solid lines (Figure 3). NSP3 PL-Pro domain and spike glycoprotein may hijack the human proteins shown as gray nodes connected by dotted lines (Figure 4). Interactions between viral proteins are shown using green colored nodes connected by dotted lines; these interactions occur in two disjoint teams: (1) NSP7, NSP8, and NSP12; and (2) NSP10, NSP14, and NSP16. For the remaining 18 viral proteins, there was no structural evidence for interactions with other viral proteins, or for mimicry or hijacking of human proteins. Seven of the 18 are structurally dark proteins: NSP6, ORF3a, matrix glycoprotein, ORF6, ORF8, ORF9c, and ORF10.

Made using Aquaria and edited with Keynote.

A NSP3 Mimicry of Host Proteins

Host Protein	NSP3 Alignment E value	NSP3 Alignment Identity	Domain Topology (Aligned Domains)	Potentially Hijacked Function
MACROD2	2×10^{-22}	24%	1 Macro domain (5051) 425	Erases ADPr, PAR, and AAR from D and E amino acids
MACROD1	3×10^{-22}	25%	1 Macro domain (5051) 325	Erases ADPr, PAR, and AAR from D and E amino acids
GDAP2	7×10^{-22}	19%	1 Macro domain (483) CRAL-TRIO domain 497	Binds ADPr; Transports lipids between membranes
MACROH2A2	2×10^{-21}	19%	1 Histone subunit A Macro domain (4935) 372	Binds ADPr, PAR, and AAR; Modifies chromatin state
PARP14	4×10^{-19}	25%	1 783 Macro domain (5051)	
	2×10^{-17}	18%	Macro domain (4917)	
	6×10^{-17}	17%	Macro domain (4911) PARP catalytic domain 1,801	Binds ADPr; Adds PAR to D and E amino acids
MACROH2A1	1×10^{-18}	19%	1 Histone subunit A Macro domain (4935) 372	Binds ADPr; Modifies chromatin state
PARP9	*No 3D structure*		1 Macro domain (5051)	
	2×10^{-17}	22%	Macro domain (260) PARP catalytic domain 854	Binds ADPr; Adds PAR to D and E amino acids
PARP15	*No 3D structure*		1 Macro domain (206)	
	6×10^{-16}	8%	Macro domain (1028) PARP catalytic domain 678	Binds ADPr; Adds PAR to D and E amino acids
OARD1	2×10^{-12}	13%	1 Macro domain (5173) 152	Erases ADPr, PAR, and AAR from D and E amino acids

B NSP13 Mimicry of Host Proteins

Host Protein	NSP13 Alignment E value	NSP13 Alignment Identity	Sequence-to-Structure Alignment	3D Structure	Color Legend	Potentially Hijacked Function
IGHMBP2	5×10^{-35}	23%	*NSP13 sequence aligned onto 3D structure of IGHMBP2*		Amino acid substitution Conserved residue: Helix Conserved residue: Sheet Conserved residue: Coil RNA Does not align to NSP13	Binds zinc Hydrolyzes ATP Unwinds RNA & DNA Repairs & recombines DNA Regulates transcription & translation Binds immunoglobulin mu chain switch region of genome
UPF1	2×10^{-32}	24%	*NSP13 sequence aligned onto 3D structure of UPF1*		UPF2 Amino acid substitution Conserved residue: Helix Conserved residue: Sheet Conserved residue: Coil Does not align to NSP13	Binds zinc Unwinds RNA Hydrolyzes ATP Binds chromatin Binds telomeric DNA Regulates nonsense-mediated decay of mRNAs
AQR	4×10^{-30}	19%	*AQR and PIF1 sequences aligned onto 3D structure of NSP13*		Aligns to AQR only Aligns to PIF1 only Aligns to both AQR and PIF1 Does not align to AQR or PIF1	Unwinds RNA Splices mRNA Hydrolyzes ATP Repairs & recombines DNA
PIF1	2×10^{-12}	20%				Hydrolyzes ATP Binds telomeric DNA Repairs & recombines DNA Unwinds RNA-DNA hybrids

Figure 3. Viral mimicry of human proteins

(A) Lists domain topology for nine human proteins potentially mimicked by NSP3 macro domain (ranked by alignment significance) along with functions that may be affected by mimicry. Each macro domain is numbered to indicate its CATH functional family. The top ranked proteins (MACROD2 and MACROD1) remove ADPr from proteins, reversing the effect of ADPr writers (PARP14, PARP15, and PARP9), and affecting ADPr readers (GDAP2, MACROH2A2, and MACROH2A1).

(B) Lists four human helicase proteins potentially mimicked by NSP13 (ranked by alignment significance) along with functions that may be affected by mimicry. Based on alignment significance, there is stronger evidence for mimicry by NSP13 than by NSP3. For the two top ranked proteins, we show 3D structures of the human proteins, colored to indicate alignment (Figure 1) with NSP13. For IGHMBP2 (P0DTD1/4b3g), we see that RNA binds to the region matched by NSP13. For UPF1 (P0DTD1/2wjv), we see that UPF2 binds to a region not matched by NSP13, suggesting that NSP13 does not bind UPF2. For AQR and PIF1, we show a structure of NSP13 (P0DTD1/6jyt), colored to indicate regions matching AQR and PIF1.

Made using features and structures from Aquaria, and edited in Photoshop and Keynote.

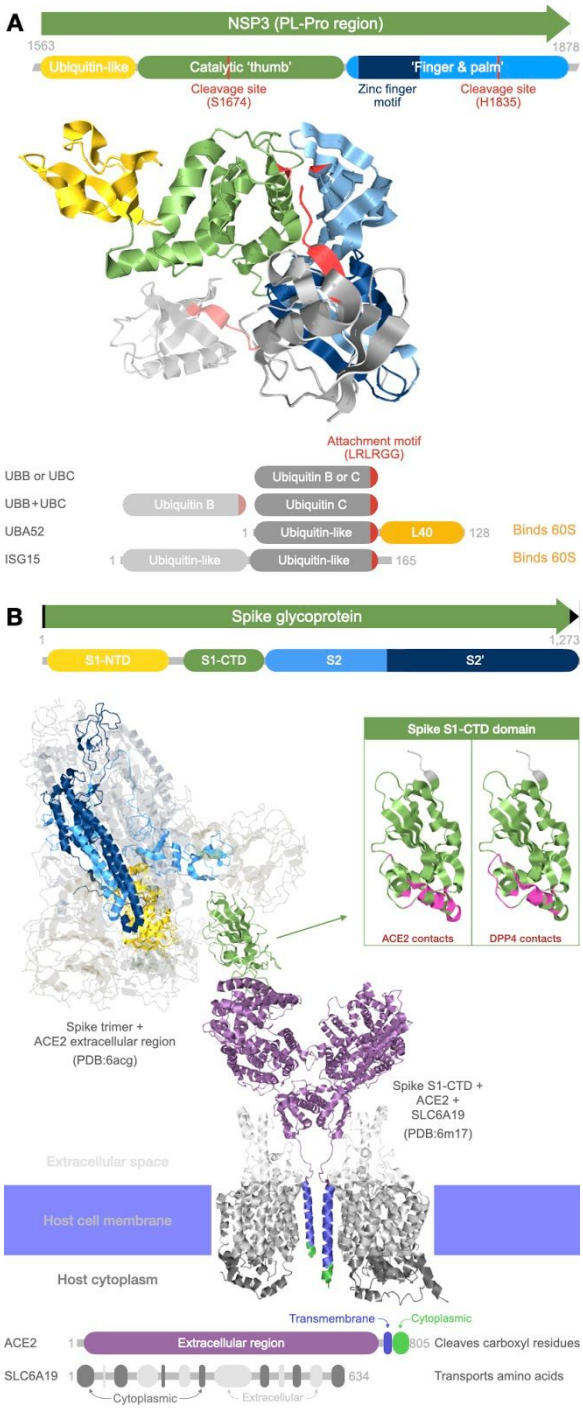

Figure 4. Viral hijacking of human proteins

(A) NSP3 (PL-Pro region) has three domains (yellow, green, and blue) and two binding sites for ubiquitin-like (Ubl) domains (gray). In most matching structures, a single Ubl domain occupies the primary binding site (dark gray), with its C-terminal LRLRGG motif (red) positioned between the PL-Pro thumb and palm domains and ending at the proteolytic cleavage site (red). The LRLRGG motif can attach to other proteins, including other copies of ubiquitin. PL-Pro cleaves this attachment, reversing mono- or poly-ubiquitination. One matching structure showed binding of two ubiquitin domains (UBB and UBC), with UBB occupying the secondary site (light gray). Other structures showed binding to the N- and C-terminal Ubl domains of ISG15 (interferon stimulated gene 15) at the secondary and primary sites, respectively.

(B) The spike trimer structure (top left; P0DTC2/6acg) shows the pre-fusion state after cleavage at the S1/S2 boundary, and with one S1-CTD domain (green) bound to the extracellular region of the carboxypeptidase ACE2 (purple). Below that, we have manually aligned a second structure (P0DTC2/6m17) showing S1-CTD in complex with ACE2 and the amino acid transporter SLC6A19. By integrating sequence features (top and bottom) with 3D structures, the resulting figure shows the cellular context of the pre-fusion event, when spike first binds the ACE + SLC6A19 complex. The insert highlights S1-CTD residues that contact ACE2 (magenta). These are compared to residues that contact the N-terminal dipeptide peptidase DPP4 in a structure from BtCoV-HKU4 (P0DTC2/4qzv). DPP4 may also be targeted by SARS-CoV-2 (Y. Li et al., 2020); if so, our analysis shows that the S1-CTD + DPP4 binding site may partly overlap the S1-CTD + ACE2 binding site.

Made using features and structures from Aquaria, and edited in Photoshop and Keynote.

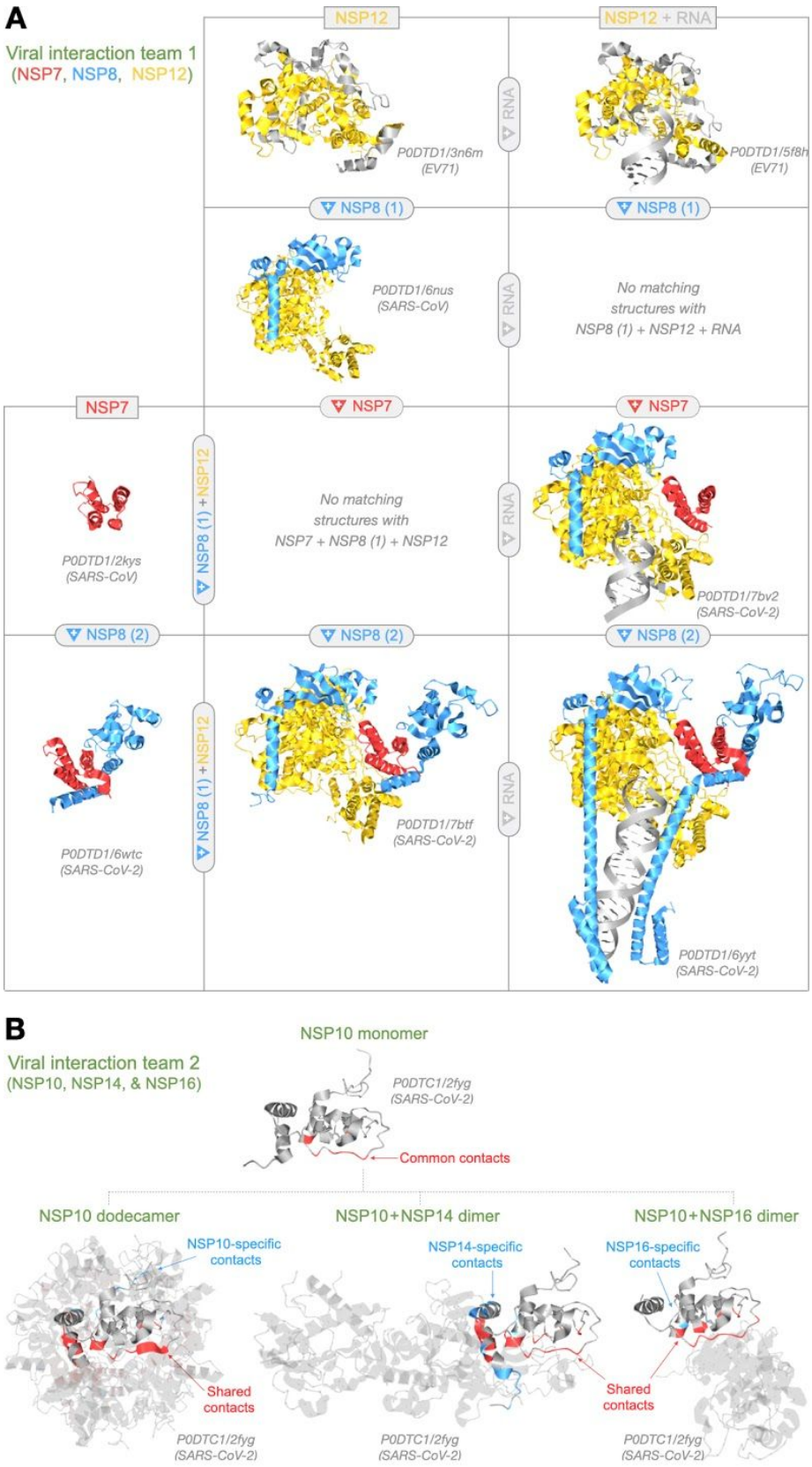

Figure 5. Viral protein interactions

(A) Summarizes all structural states observed for NSP7 (red), NSP8 (blue), and NSP12 (yellow). NSP12 alone (top row, left) can replicate RNA (top row, right); however, all matching structures were from distantly related viruses (pairwise identity ≤ 18%), such as EV71 (enterovirus). NSP8 binds NSP12 at two sites: (1) at the NSP12 core (2nd row, left); and (2) via NSP7-mediated cooperative interactions at the NSP12 periphery (bottom row, center), greatly enhancing RNA replication (bottom row, right). By themselves, NSP7 + NSP8 (bottom row, left) can also replicate RNA (te Velthuis et al., 2012). Six of the matching structures showed NSP7 + NSP8 as a dimer; however, one structure showed a tetramer and one other showed a hexadecamer (P0DTD1/3ub0 and P0DTD1/2ahm, respectively).

(B) Summarizes all structural states observed for NSP10, NSP14, and NSP16 (Team 2). One matching structure showed an NSP10 monomer (top), while two structures showed NSP10 as a dodecamer (bottom, left). All structures matching NSP14 showed a heterodimer with NSP10 (bottom, middle). All structures matching NSP16 showed a heterodimer with NSP10 (bottom, right). Nine NSP10 residues (shown in red on the monomer) were common intermolecular contacts in all three oligomers. Each oligomer also had specific NSP10 contacts (blue); but most NSP10 contacts were shared with at least one other oligomer (shown in red on each oligomer). This suggests that NSP10, NSP14, and NSP16 interact competitively.

Made using Aquaria and edited with Keynote.

SUPPLEMENTAL INFORMATION

Table S1 | Matching structures for SARS-CoV-2 polyprotein 1a

NSP1: P0DTC1/2gdt, P0DTC1/2hsx, P0DTC1/3ld1

NSP2: P0DTC1/3ld1

NSP3 (Ubl1): P0DTC1/2gri, P0DTC1/2idy, P0DTC1/2m0a

NSP3 (Macro):
P0DTC1/6woj, P0DTC1/6ywk, P0DTC1/6ywl, P0DTC1/6ywm, P0DTC1/6vxs, P0DTC1/6w6y, P0DTC1/6wen, P0DTC1/6w02, P0DTC1/6wcf, P0DTC1/6wey, P0DTC1/2fav, P0DTC1/2acf, P0DTC1/5zua, P0DTC1/5dus, P0DTC1/5zu7, P0DTC1/5zu9, P0DTC1/5zub, P0DTC1/5hih, P0DTC1/5hol, P0DTC1/6mea, P0DTC1/6meb, P0DTC1/6men, P0DTC1/2vri, P0DTC1/3ejg, P0DTC1/3ewq, P0DTC1/3ewr, P0DTC1/3gqe, P0DTC1/3gqo, P0DTC1/5isn, P0DTC1/5mqx, P0DTC1/1spv, P0DTC1/5cms, P0DTC1/5cb3, P0DTC1/6fx7, P0DTC1/5cb5, P0DTC1/1vhu, P0DTC1/1hjz, P0DTC1/2bfq, P0DTC1/2bfr, P0DTC1/4gua, P0DTC1/5iq5, P0DTC1/4gvw, P0DTC1/5lcc, P0DTC1/3ew5, P0DTC1/3jzt, P0DTC1/2x47, P0DTC1/3q6z, P0DTC1/3eti, P0DTC1/3vfq, P0DTC1/4d86, P0DTC1/5fud, P0DTC1/5l9k, P0DTC1/5l9q, P0DTC1/3ewo, P0DTC1/3ewp, P0DTC1/4etj, P0DTC1/3v45, P0DTC1/4etk, P0DTC1/2dx6, P0DTC1/4iqy, P0DTC1/5iit, P0DTC1/5lnc, P0DTC1/1yd9, P0DTC1/5lau, P0DTC1/5lbp, P0DTC1/5fsz, P0DTC1/4gvv, P0DTC1/4jll, P0DTC1/4jca, P0DTC1/4kyb, P0DTC1/4k0c, P0DTC1/4ess, P0DTC1/6r0r, P0DTC1/6r0g, P0DTC1/6r0p, P0DTC1/6r0t, P0DTC1/6qzu, P0DTC1/6r0f, P0DTC1/3ejf, P0DTC1/3eke, P0DTC1/3iid, P0DTC1/3iif, P0DTC1/1zr3, P0DTC1/2fxk, P0DTC1/4j4z, P0DTC1/4jgk, P0DTC1/4jvv, P0DTC1/3gpg, P0DTC1/3gpo, P0DTC1/4tu0, P0DTC1/3gpq, P0DTC1/5kiv, P0DTC1/5ail, P0DTC1/5fsu, P0DTC1/5fsx, P0DTC1/5fsv, P0DTC1/5fsy, P0DTC1/6fy5, P0DTC1/2xd7, P0DTC1/1zr5, P0DTC1/4uml, P0DTC1/5o2d, P0DTC1/3q71, P0DTC1/3v2b, P0DTC1/1njr, P0DTC1/1txz, P0DTC1/1ty8, P0DTC1/3vfq, P0DTC1/4d86, P0DTC1/4abk, P0DTC1/5qht, P0DTC1/5qhu, P0DTC1/5qhv, P0DTC1/5qhw, P0DTC1/5qhx, P0DTC1/5qhy, P0DTC1/5qhz, P0DTC1/5qi0, P0DTC1/5qi1, P0DTC1/5qi2, P0DTC1/5qi3, P0DTC1/5qi4, P0DTC1/5qi5, P0DTC1/5qi6, P0DTC1/5qi7, P0DTC1/5qi8, P0DTC1/5qi9, P0DTC1/5qia, P0DTC1/4abl, P0DTC1/5m31, P0DTC1/5m3e, P0DTC1/5e3b, P0DTC1/2w2g, P0DTC1/2wct, P0DTC1/5lw0, P0DTC1/5m3i, P0DTC1/2eee, P0DTC1/2lgr, P0DTC1/4j5q, P0DTC1/2l8r, P0DTC1/4j5s, P0DTC1/4j5r, P0DTC1/2fg1, P0DTC1/2afc

NSP3 (SUD-M): P0DTC1/2jzd, P0DTC1/2jze, P0DTC1/2kqv, P0DTC1/2jzf, P0DTC1/2rnk, P0DTC1/2w2g, P0DTC1/2wct

NSP3 (SUD-C): P0DTC1/2kqw, P0DTC1/4ypt

NSP3 - PL-Pro:
P0DTC1/6w9c, P0DTC1/6wuu, P0DTC1/6wx4, P0DTC1/6yva, P0DTC1/6wrh, P0DTC1/5e6j, P0DTC1/3e9s, P0DTC1/4mm3, P0DTC1/3mj5, P0DTC1/4ow0, P0DTC1/4ovz, P0DTC1/2fe8, P0DTC1/5y3e, P0DTC1/5y3q, P0DTC1/4m0w, P0DTC1/5tl7, P0DTC1/5tl6, P0DTC1/5wfi, P0DTC1/4r3d, P0DTC1/5v6a, P0DTC1/5v69, P0DTC1/4rf1, P0DTC1/4rez, P0DTC1/4rf0, P0DTC1/4p16, P0DTC1/4rna, P0DTC1/4pt5, P0DTC1/5w8t, P0DTC1/5w8u, P0DTC1/4wur, P0DTC1/5ko3, P0DTC1/6bi8, P0DTC1/5bz0, P0DTC1/6noz, P0DTC1/4ypt, P0DTC1/3mp2, P0DTC1/4x2z, P0DTC1/6l5t

NSP3 (NAB): P0DTC1/2k87

NSP4: P0DTC1/3vcb, P0DTC1/3vc8, P0DTC1/3gzf

NSP5:
P0DTC1/5rfa, P0DTC1/6y2g, P0DTC1/7bqy, P0DTC1/5r7y, P0DTC1/5r7z, P0DTC1/5r80, P0DTC1/5r81, P0DTC1/5r82, P0DTC1/5r83, P0DTC1/5r84, P0DTC1/5r8t, P0DTC1/5re4, P0DTC1/5re5, P0DTC1/5re6, P0DTC1/5re7, P0DTC1/5re8, P0DTC1/5re9, P0DTC1/5rea, P0DTC1/5reb, P0DTC1/5rec, P0DTC1/5red, P0DTC1/5ree, P0DTC1/5ref, P0DTC1/5reg, P0DTC1/5reh, P0DTC1/5rei, P0DTC1/5rej, P0DTC1/5rek, P0DTC1/5rel, P0DTC1/5rem, P0DTC1/5ren, P0DTC1/5reo, P0DTC1/5rep, P0DTC1/5rer, P0DTC1/5res, P0DTC1/5ret, P0DTC1/5reu, P0DTC1/5rev, P0DTC1/5rew, P0DTC1/5rex, P0DTC1/5rey, P0DTC1/5rez, P0DTC1/5rf0, P0DTC1/5rf1, P0DTC1/5rf2, P0DTC1/5rf3, P0DTC1/5rf4, P0DTC1/5rf5, P0DTC1/5rf6, P0DTC1/5rf7, P0DTC1/5rf8, P0DTC1/5rf9, P0DTC1/5rfb, P0DTC1/5rfc, P0DTC1/5rfd, P0DTC1/5rfe, P0DTC1/5rff, P0DTC1/5rfg, P0DTC1/5rfh, P0DTC1/5rfi, P0DTC1/5rfj, P0DTC1/5rfk, P0DTC1/5rfl, P0DTC1/5rfm, P0DTC1/5rfn, P0DTC1/5rfo, P0DTC1/5rfp, P0DTC1/5rfq, P0DTC1/5rfr, P0DTC1/5rfs, P0DTC1/5rft, P0DTC1/5rfu, P0DTC1/5rfv, P0DTC1/5rfw, P0DTC1/5rfx, P0DTC1/5rfy, P0DTC1/5rfz, P0DTC1/5rg0, P0DTC1/5rg1, P0DTC1/5rg2, P0DTC1/5rg3, P0DTC1/5rgg, P0DTC1/5rgh, P0DTC1/5rgi, P0DTC1/5rgj, P0DTC1/5rgk, P0DTC1/5rgl, P0DTC1/5rgm, P0DTC1/5rgn, P0DTC1/5rgo, P0DTC1/5rgp, P0DTC1/5rgq, P0DTC1/5rgr, P0DTC1/5rgs, P0DTC1/6y84, P0DTC1/6yz6, P0DTC1/6m2q, P0DTC1/6w63, P0DTC1/6lu7, P0DTC1/6m03, P0DTC1/6m2n, P0DTC1/6wnp, P0DTC1/6wqf, P0DTC1/6wtj, P0DTC1/6wtk, P0DTC1/6wtm, P0DTC1/6y2e, P0DTC1/6yb7, P0DTC1/6ynq, P0DTC1/6yt8, P0DTC1/6yvf, P0DTC1/7buy, P0DTC1/6y2f, P0DTC1/7brp, P0DTC1/7bro, P0DTC1/7brr, P0DTC1/6wtt, P0DTC1/6m0k, P0DTC1/6lze, P0DTC1/3m3v, P0DTC1/1wof, P0DTC1/2amd, P0DTC1/2d2d, P0DTC1/2amq, P0DTC1/3m3s, P0DTC1/1z1i, P0DTC1/3iwm, P0DTC1/3vb3, P0DTC1/3vb4, P0DTC1/3vb5, P0DTC1/3vb6, P0DTC1/3vb7, P0DTC1/1uk2, P0DTC1/2c3s, P0DTC1/2z9j, P0DTC1/2z9k, P0DTC1/2a5a, P0DTC1/2a5i, P0DTC1/2duc, P0DTC1/2gx4, P0DTC1/2gz7, P0DTC1/2gz8, P0DTC1/2gz9, P0DTC1/2h2z, P0DTC1/2hob, P0DTC1/2v6n, P0DTC1/2z3c, P0DTC1/2z3d, P0DTC1/2z3e, P0DTC1/2z94, P0DTC1/2z9g, P0DTC1/2zu4, P0DTC1/2zu5, P0DTC1/3sn8, P0DTC1/3snb, P0DTC1/3snc, P0DTC1/3snd, P0DTC1/3sne, P0DTC1/3szn, P0DTC1/3tit, P0DTC1/3tiu, P0DTC1/3tns, P0DTC1/3tnt, P0DTC1/3v3m, P0DTC1/5n19, P0DTC1/5n5o, P0DTC1/6y7m, P0DTC1/2gt7, P0DTC1/2z9l, P0DTC1/2bx3, P0DTC1/2bx4, P0DTC1/1ujl, P0DTC1/1uk3, P0DTC1/1uk4, P0DTC1/1z1j, P0DTC1/3atw, P0DTC1/3avz, P0DTC1/5c5o, P0DTC1/3awl, P0DTC1/3aw0, P0DTC1/4tww, P0DTC1/4twy, P0DTC1/4wy3, P0DTC1/5c5n, P0DTC1/2a5k, P0DTC1/2gtb, P0DTC1/2gt8, P0DTC1/3E+91, P0DTC1/3ea7, P0DTC1/3ea8, P0DTC1/3ea9, P0DTC1/3eaj, P0DTC1/2qcy, P0DTC1/2q6g, P0DTC1/1q2w, P0DTC1/2vj1, P0DTC1/4mds, P0DTC1/3sna, P0DTC1/2op9, P0DTC1/3fzd, P0DTC1/2qiq, P0DTC1/3d62, P0DTC1/4hi3, P0DTC1/3f9e, P0DTC1/3f9f, P0DTC1/3f9h, P0DTC1/2qc2, P0DTC1/3m3t, P0DTC1/2pwx, P0DTC1/2alv, P0DTC1/3f9g, P0DTC1/5b6o, P0DTC1/5wkk, P0DTC1/5wkm, P0DTC1/5wkj, P0DTC1/5wkl, P0DTC1/5c3n, P0DTC1/4rsp, P0DTC1/4ylu, P0DTC1/4wmf, P0DTC1/4wmd, P0DTC1/4wme, P0DTC1/2yna, P0DTC1/2ynb, P0DTC1/4yo9, P0DTC1/4yog, P0DTC1/4yoi, P0DTC1/4yoj, P0DTC1/6jij, P0DTC1/3d23, P0DTC1/4f49, P0DTC1/5gwz, P0DTC1/2amp, P0DTC1/5eu8, P0DTC1/1lvo, P0DTC1/1p9u, P0DTC1/5hyo, P0DTC1/4xfq, P0DTC1/6l70, P0DTC1/4zuh, P0DTC1/4zro, P0DTC1/5zqg, P0DTC1/5gwy, P0DTC1/3tlo, P0DTC1/6fv1, P0DTC1/6fv2, P0DTC1/5nh0, P0DTC1/5nh0, P0DTC1/5nh0, P0DTC1/2q6d, P0DTC1/2q6f, P0DTC1/2zu2, P0DTC1/1p9s, P0DTC1/2k7x, P0DTC1/2liz, P0DTC1/3ebn

NSP6: No matching structures

NSP7:
P0DTC1/7btf, P0DTC1/6m71, P0DTC1/6yhu, P0DTC1/7bv1, P0DTC1/7bv2, P0DTC1/6nur, P0DTC1/6wqd, P0DTC1/1ysy, P0DTC1/2kys, P0DTC1/6wiq, P0DTC1/6wtc, P0DTC1/6yyt, P0DTC1/2ahm, P0DTC1/5f22, P0DTC1/3ub0

NSP8:
P0DTC1/7btf, P0DTC1/6wiq, P0DTC1/6wqd, P0DTC1/6wtc, P0DTC1/6yhu, P0DTC1/7bv1, P0DTC1/7bv2, P0DTC1/6nur, P0DTC1/6nus, P0DTC1/5f22, P0DTC1/pdb_id, P0DTC1/6m71, P0DTC1/pdb_id, P0DTC1/6yyt, P0DTC1/2ahm, P0DTC1/3ub0

NSP9:
P0DTC1/6wxd, P0DTC1/6w4b, P0DTC1/1qz8, P0DTC1/3ee7, P0DTC1/5hiz, P0DTC1/5hiy, P0DTC1/2j97, P0DTC1/2j98, P0DTC1/2j98, P0DTC1/5c94, P0DTC1/5ym6, P0DTC1/6w9q, P0DTC1/1uw7, P0DTC1/5ym8

NSP10:
P0DTC1/6wks, P0DTC1/6w61, P0DTC1/6w4h, P0DTC1/6yz1, P0DTC1/6wvn, P0DTC1/6wrz, P0DTC1/6wq3, P0DTC1/6w75, P0DTC1/6wjt, P0DTC1/6wkq, P0DTC1/2xyv, P0DTC1/2xyq, P0DTC1/2xyr, P0DTC1/2fyg, P0DTC1/2g9t, P0DTC1/2ga6, P0DTC1/3r24, P0DTC1/5ynn, P0DTC1/5yno, P0DTC1/5ynb, P0DTC1/5ynp, P0DTC1/5yn6, P0DTC1/5ynj, P0DTC1/5ynm, P0DTC1/5ynq, P0DTC1/5yn8, P0DTC1/5ynf, P0DTC1/5yn5, P0DTC1/5ynj, P0DTC1/5nfy, P0DTC1/5c8s, P0DTC1/5c8t, P0DTC1/5c8u

Table S2 | Matching structures for SARS-CoV-2 polyprotein 1b

NSP12:
P0DTD1/6yyt, P0DTD1/6m71, P0DTD1/7btf, P0DTD1/7bv1, P0DTD1/7bv2, P0DTD1/6nur, P0DTD1/6nus, P0DTD1/3nky, P0DTD1/3kmq, P0DTD1/3kms, P0DTD1/4y2c, P0DTD1/3n6l, P0DTD1/3n6m, P0DTD1/3n6n, P0DTD1/4zp6, P0DTD1/4zp8, P0DTD1/4y34, P0DTD1/4zpb, P0DTD1/3nl0, P0DTD1/4zp7, P0DTD1/4wfx, P0DTD1/4wfy, P0DTD1/4wyw, P0DTD1/3kna, P0DTD1/3koa, P0DTD1/4zpd, P0DTD1/4r0e, P0DTD1/4k4w, P0DTD1/4k4u, P0DTD1/4k4v, P0DTD1/1xr7, P0DTD1/1tp7, P0DTD1/4k50, P0DTD1/4y3c, P0DTD1/4nz0, P0DTD1/5f8g, P0DTD1/5f8h, P0DTD1/5f8i, P0DTD1/5f8j, P0DTD1/5f8l, P0DTD1/5f8m, P0DTD1/5f8n, P0DTD1/4ika, P0DTD1/4nyz, P0DTD1/2ckw, P0DTD1/2uut, P0DTD1/1raj

NSP13:
P0DTD1/6jyt, P0DTD1/5wwp, P0DTD1/3k70, P0DTD1/6sjb, P0DTD1/6sje, P0DTD1/6sjf, P0DTD1/6sjg, P0DTD1/6t2u, P0DTD1/6t2v, P0DTD1/1w36, P0DTD1/5mbv, P0DTD1/5ld2, P0DTD1/6hpt, P0DTD1/5fhh, P0DTD1/4b3g, P0DTD1/2xzo, P0DTD1/2xzp, P0DTD1/2xzl, P0DTD1/4b3f, P0DTD1/2wjv, P0DTD1/2wjy, P0DTD1/2gjk, P0DTD1/2gk6, P0DTD1/2gk7, P0DTD1/6ej5, P0DTD1/5mzn, P0DTD1/6i59, P0DTD1/3gpl, P0DTD1/3gp8, P0DTD1/3e1s, P0DTD1/3wrx, P0DTD1/3wry, P0DTD1/3vkw, P0DTD1/4n0o, P0DTD1/4n0n, P0DTD1/6jdu, P0DTD1/6jdr, P0DTD1/5eaw, P0DTD1/5ean, P0DTD1/5eax, P0DTD1/4rs4, P0DTD1/4s1t, P0DTD1/6w01, P0DTD1/6vww, P0DTD1/6wlc, P0DTD1/6wxc, P0DTD1/2ozk, P0DTD1/2gth, P0DTD1/5yvd, P0DTD1/5xjc, P0DTD1/5yzg, P0DTD1/6icz, P0DTD1/6id0, P0DTD1/6id1, P0DTD1/6qdv, P0DTD1/5mqf, P0DTD1/6ff7, P0DTD1/5z56, P0DTD1/5z57, P0DTD1/4pj3, P0DTD1/6jim, P0DTD1/3jb9, P0DTD1/3upu, P0DTD1/6jds

NSP14: P0DTD1/5nfy, P0DTD1/5c8s, P0DTD1/5c8t, P0DTD1/5c8u

NSP15:
P0DTD1/6wlc, P0DTD1/6wxc, P0DTD1/6w01, P0DTD1/6vww, P0DTD1/2ozk, P0DTD1/2rhb, P0DTD1/2h85, P0DTD1/5yvd, P0DTD1/2gth, P0DTD1/2gti, P0DTD1/4s1t, P0DTD1/4rs4, P0DTD1/5f17, P0DTD1/5hbz, P0DTD1/5hc1, P0DTD1/5eyi, P0DTD1/5da1, P0DTD1/5wwp, P0DTD1/6jyt

NSP16:
P0DTD1/6w4h, P0DTD1/6wq3, P0DTD1/6wrz, P0DTD1/6wvn, P0DTD1/6w75, P0DTD1/6wjt, P0DTD1/6wkq, P0DTD1/6wks, P0DTD1/6w61, P0DTD1/6yz1, P0DTD1/2xyv, P0DTD1/2xyr, P0DTD1/2xyq, P0DTD1/3r24, P0DTD1/5yn5, P0DTD1/5ynb, P0DTD1/5ynm, P0DTD1/5ynf, P0DTD1/5ynj, P0DTD1/5ynp, P0DTD1/5ynn, P0DTD1/5yn6, P0DTD1/5yno, P0DTD1/5ynq, P0DTD1/5yni, P0DTD1/5yn8

Table S3 | Matching structures for SARS-CoV-2 Virion and Accessory Proteins

Spike glycoprotein:
P0DTC2/6m17, P0DTC2/6lzg, P0DTC2/6m0j, P0DTC2/6w41, P0DTC2/6vw1, P0DTC2/3sci, P0DTC2/2ghw, P0DTC2/2ghv, P0DTC2/3d0g, P0DTC2/3d0h, P0DTC2/3scj, P0DTC2/3sck, P0DTC2/3scl, P0DTC2/2dd8, P0DTC2/4l3n, P0DTC2/4kqz, P0DTC2/4kr0, P0DTC2/5do2, P0DTC2/6c6z, P0DTC2/4zs6, P0DTC2/4zpt, P0DTC2/4zpw, P0DTC2/5gmq, P0DTC2/6yor, P0DTC2/7bz5, P0DTC2/6yla, P0DTC2/6ym0, P0DTC2/5gyq, P0DTC2/5xgr, P0DTC2/4qzv, P0DTC2/4xak, P0DTC2/6l8q, P0DTC2/6war, P0DTC2/5yy5, P0DTC2/5zxv, P0DTC2/1zvb, P0DTC2/2ruo, P0DTC2/6vsb, P0DTC2/6vxx, P0DTC2/6vyb, P0DTC2/6acc, P0DTC2/6acd, P0DTC2/6acg, P0DTC2/6acj, P0DTC2/6ack, P0DTC2/6crv, P0DTC2/6cs2, P0DTC2/6crw, P0DTC2/6crz, P0DTC2/6cs0, P0DTC2/6crx, P0DTC2/6cs1, P0DTC2/6nb6, P0DTC2/6nb7, P0DTC2/5xlr, P0DTC2/5x5b, P0DTC2/5x58, P0DTC2/6q04, P0DTC2/6q05, P0DTC2/6q06, P0DTC2/6q07, P0DTC2/6nb3, P0DTC2/6nb4, P0DTC2/6nzk, P0DTC2/6ohw, P0DTC2/3jcl, P0DTC2/6vsj, P0DTC2/5x59, P0DTC2/5x5c, P0DTC2/5x5f, P0DTC2/5i08, P0DTC2/5zvm, P0DTC2/5zvk, P0DTC2/2bez, P0DTC2/1wdf, P0DTC2/5yl9, P0DTC2/1wdg, P0DTC2/1zva, P0DTC2/5xjk, P0DTC2/1zv8, P0DTC2/1wnc, P0DTC2/5zhy, P0DTC2/2ajf, P0DTC2/3d0i, P0DTC2/5wrg, P0DTC2/6lxt, P0DTC2/1wyy, P0DTC2/4njl, P0DTC2/5x4s, P0DTC2/6qfy, P0DTC2/6jhy, P0DTC2/4h14, P0DTC2/5zuv, P0DTC2/6pz8, P0DTC2/5w9l, P0DTC2/5w9h, P0DTC2/5w9i, P0DTC2/5w9j, P0DTC2/5w9m, P0DTC2/5w9n, P0DTC2/5w9o, P0DTC2/5w9p, P0DTC2/5w9k, P0DTC2/6b3o, P0DTC2/2ieq, P0DTC2/4mod, P0DTC2/6u7h, P0DTC2/6u7k, P0DTC2/6vv5, P0DTC2/6cv0, P0DTC2/6bfu, P0DTC2/5szs, P0DTC2/6b7n, P0DTC2/6jx7, P0DTC2/6pz8, P0DTC2/5gnb, P0DTC2/5kwb, P0DTC2/6pxh, P0DTC2/5x4r, P0DTC2/3r4d, P0DTC2/4l72, P0DTC2/1zv7, P0DTC2/6lvn, P0DTC2/1zva, P0DTC2/2bez, P0DTC2/6lxt, P0DTC2/1zv8, P0DTC2/1wnc, P0DTC2/2fxp, P0DTC2/2beq, P0DTC2/1wyy, P0DTC2/2run, P0DTC2/5x4r, P0DTC2/6j11, P0DTC2/6pxh, P0DTC2/5vyh

ORF3a: No matching structures

Envelope protein: P0DTC4/2mm4, P0DTC4/5x29

Matrix glycoprotein: No matching structures

ORF6: No matching structures

ORF7a: P0DTC7/6w37, P0DTC7/1xak, P0DTC7/1yo4

ORF8: No matching structures

ORF7b: No matching structures

N protein (NTD):
P0DTC9/6yi3, P0DTC9/6vyo, P0DTC9/6m3m, P0DTC9/2og3, P0DTC9/2ofz, P0DTC9/1ssk, P0DTC9/6kl2, P0DTC9/6kl6, P0DTC9/6kl5, P0DTC9/4ud1, P0DTC9/3hd4, P0DTC9/4j3k, P0DTC9/4kxj, P0DTC9/4li4, P0DTC9/4lm7, P0DTC9/4lmt, P0DTC9/4lm9, P0DTC9/4lmc, P0DTC9/2bxx, P0DTC9/2c86, P0DTC9/2gec, P0DTC9/2btl, P0DTC9/5n4k, P0DTC9/6wkp

N protein (CTD):
P0DTC9/6wji, P0DTC9/7c22, P0DTC9/2gib, P0DTC9/6g13, P0DTC9/5epw, P0DTC9/2ge7, P0DTC9/2ge8, P0DTC9/6yun, P0DTC9/2jw8, P0DTC9/2cjr, P0DTC9/2ca1

ORF9b: P0DTD2/2cme, P0DTD2/2cme

ORF10: No matching structures

ORF14: No matching structures

Table S4 | Structural evidence for mimicry and hijacking of human proteins

NSP3 (macro) may mimic:

GDAP2: P0DTC1/4uml

MACROD1: P0DTC1/2x47

MACROD2: P0DTC1/4iqy

MACROH2A1: P0DTC1/1zr5, P0DTC2/5iit, P0DTC3/5lnc

MACROH2A2: P0DTC1/2xd7, P0DTC1/2xd7

OARD1: P0DTC1/2eee, P0DTC1/2l8r, P0DTC1/2lgr, P0DTC1/4j5q, P0DTC1/4j5r, P0DTC1/4j5s

PARP9: P0DTC1/5ail

PARP14:
P0DTC1/3q6z, P0DTC1/3q71, P0DTC1/3vfq, P0DTC1/3vfq, P0DTC1/4abk, P0DTC1/4abl, P0DTC1/4d86, P0DTC1/4d86, P0DTC1/5o2d, P0DTC1/5qht, P0DTC1/5qhu, P0DTC1/5qhv, P0DTC1/5qhw, P0DTC1/5qhx, P0DTC1/5qhy, P0DTC1/5qhz, P0DTC1/5qi0, P0DTC1/5qi1, P0DTC1/5qi2, P0DTC1/5qi3, P0DTC1/5qi4, P0DTC1/5qi5, P0DTC1/5qi6, P0DTC1/5qi7, P0DTC1/5qi8, P0DTC1/5qi9, P0DTC1/5qia

PARP15: P0DTC1/3v2b

NSP3 (PL-Pro) may hijack:

ISG15: P0DTC1/5tl6, P0DTC1/5w8t, P0DTC1/5w8u, P0DTC1/6bi8

UBA52: P0DTC1/4rf0, P0DTC1/4rf1

UBB: P0DTC1/4wur, P0DTC1/5e6j, P0DTC1/5wfi

UBC: P0DTC1/4mm3, P0DTC1/5e6j

NSP13 may mimic:

AQR: P0DTD1/4pj3
... and thereby hijack the spliceosome:
P0DTD1/5yzg, P0DTD1/6icz, P0DTD1/6id0, P0DTD1/6id1, P0DTD1/6qdv, P0DTD1/6ff7, P0DTD1/5mqf, P0DTD1/5z56, P0DTD1/5z57, P0DTD1/5xjc

PIF1: P0DTD1/6hpt

UPF1: P0DTD1/2xzo, P0DTD1/2xzp, P0DTD1/2wjy, P0DTD1/2gjk, P0DTD1/2gk6, P0DTD1/2gk7, P0DTD1/6ej5
... and thereby hijack *UPF2*: P0DTD1/2wjv

IGHMBP2: P0DTD1/4b3g, P0DTD1/4b3f

Spike glycoprotein may hijack:

ACE2:
P0DTC2/6lzg, P0DTC2/6m17, P0DTC2/6m0j, P0DTC2/6vw1, P0DTC2/3sci, P0DTC2/3d0g, P0DTC2/3d0h, P0DTC2/3sck, P0DTC2/3scj, P0DTC2/3scl, P0DTC2/6acg, P0DTC2/6acj, P0DTC2/6ack, P0DTC2/6cs2, P0DTC2/2ajf, P0DTC2/3d0i

ACE2 + ***SLC6A19***: P0DTC2/6m17

DPP4: P0DTC2/4qzv, P0DTC2/4l72, P0DTC2/4qzv