

## **Identification of motif-based interactions between SARS-CoV-2 protein domains and human peptide ligands pinpoint antiviral targets.**

Filip Mihalic<sup>1</sup>, Caroline Benz<sup>2</sup>, Eszter Kassa<sup>2</sup>, Richard Lindqvist<sup>3,4</sup>, Leandro Simonetti<sup>2</sup>,  
Raviteja Inturi<sup>1</sup>, Hanna Aronsson<sup>1</sup>, Eva Andersson<sup>1</sup>, Celestine N. Chi<sup>1</sup>, Norman E. Davey<sup>5</sup>,  
Anna K. Överby<sup>3,4</sup>, Per Jemth<sup>1,\*</sup>, Ylva Ivarsson<sup>2,\*</sup>

<sup>1</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, Husargatan 3, 751 23 Uppsala, Sweden

<sup>2</sup>Department of Chemistry - BMC, Uppsala University, Box 576, Husargatan 3, 751 23 Uppsala, Sweden

<sup>3</sup>Department of Clinical Microbiology, Umeå University, 90185 Umeå, Sweden

<sup>4</sup>Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, 90187 Umeå

<sup>5</sup>Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

\*Correspondence: [Per.Jemth@imbim.uu.se](mailto:Per.Jemth@imbim.uu.se) or [ylva.ivarsson@kemi.uu.se](mailto:ylva.ivarsson@kemi.uu.se)

**The infection and replication cycle of all viruses depend on interactions between viral and host proteins. Each of these protein-protein interactions is therefore a potential drug target. These host-virus interactions often involve a disordered protein region on one side of the interface and a folded protein domain on the other. Here, we used proteomic peptide phage display (ProP-PD) to identify peptides from the intrinsically disordered regions of the human proteome that bind to folded protein domains encoded by the SARS-CoV-2 genome. Eleven folded domains of SARS-CoV-2 proteins were found to bind peptides from human proteins. Of 281 high/medium confidence peptides, 23 interactions involving eight SARS-CoV-2 protein domains were tested by fluorescence polarization, and binding was observed with affinities spanning the whole micromolar range. The key specificity determinants were established for six of these domains, two based on ProP-PD and four by alanine scanning SPOT arrays. Finally, two cell-penetrating peptides, targeting Nsp9 and Nsp16, respectively, were shown to function as inhibitors of viral replication. Our findings demonstrate how high-throughput peptide binding screens simultaneously provide information on potential host-virus interactions and identify ligands with antiviral properties.**

## Introduction

The coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus has overwhelmed hospitals around the world and pushed public health facilities to their limits in the past years. Rapid vaccine development has drastically reduced this burden, but given the high rate of vaccine breakthrough cases, better post-infection therapeutic interventions are needed. Currently, many academic and industrial laboratories are working to develop new drugs or testing the effect of repurposing existing drugs against SARS-CoV-2. For example, molnupiravir is a nucleoside analogue targeting viral RNA-dependent RNA polymerase<sup>1</sup>, and paxlovid is a combination of two protease inhibitors, one previously used against human immunodeficiency virus (HIV) and the other developed to target the 3C-like protease of coronaviruses<sup>2</sup>. However, to be prepared for future epidemics and pandemics, it is important to continue developing new broad-spectrum antivirals such that emerging viral threats can be treated directly and immediately with approved drugs, preferably administered orally. This is particularly important for SARS-like coronaviruses, given the efficiency of human-to-human transmission, and has been demonstrated by recurrent outbreaks: severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003, Middle East respiratory syndrome-related coronavirus (MERS-CoV) in 2012 and SARS-CoV-2 in 2019<sup>3</sup>.

While most drugs work by inhibiting enzymes, the pharmaceutical industry is turning its attention to the more challenging but largely untouched area of protein-protein interaction interfaces. To this end, several approaches have been developed to identify novel drug targets by mapping the virus-host protein-protein interactome<sup>4-7</sup>. In a particular class of protein-protein interactions, an intrinsically disordered region (IDR) of a protein interacts with a folded domain of the binding partner. The most common interaction modules in IDRs are short linear motifs (SLiMs). SLiMs are usually less than 10 amino acid residues long and degenerate, with only 3-4 residues accounting for most of the specificity and affinity. Consequently, SLiMs can evolve

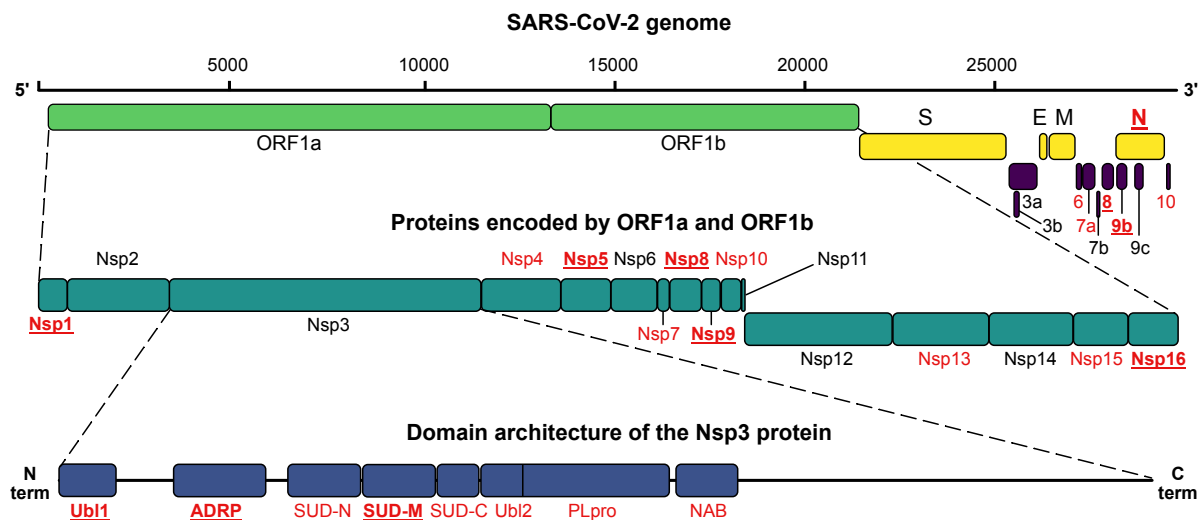
*ex nihilo* through accumulation of a few mutations. Because viruses evolve relatively rapidly, viral mimicry of host protein SLiMs has been proposed as a commonly used strategy, and several examples have been described across most viral clades<sup>6-8</sup>. Proteins expressed from viral genomes therefore often contain SLiMs which compete with native host protein-protein interactions to hijacking cellular machinery. Conversely, folded domains of the viral proteome can interact with both human and viral SLiMs<sup>8-11</sup>. Each scenario provides a drug target, *i.e.*, the folded domain of human or viral protein can be targeted with a peptide or peptidomimetic.

The SARS-CoV-2 proteome consists of 29 proteins, encoded by 14 open reading frames (ORF)<sup>5</sup>. The first two reading frames, ORF1a and ORF1b, encode 16 non-structural proteins (Nsp1 to Nsp16), formed after post-translational proteolytic cleavage by viral proteases. In addition, the SARS-CoV-2 genome codes four structural proteins: spike (S), nucleocapsid (N), membrane (M), and envelope (E), as well as nine accessory proteins (ORFs 3a, 3b, 6, 7a, 7b, 8, 9b, 9c, and 10) (**Figure 1**). Non-structural proteins facilitate viral mRNA replication and translation. Nsp1 is responsible for suppressing host translation while simultaneously promoting viral mRNA translation<sup>12</sup>. Nsp2 promotes viral evasion of the innate immune response<sup>13</sup>. Nsp3 serves as a central hub coordinating different steps of viral replication<sup>14,15</sup>. It consists of several stably folded globular domains, namely two ubiquitin-like domains (Ubl1, Ubl2), an ADP-ribose-1''-(phosphatase) hydrolase (ADRP), three SARS-unique domains (SUD-N, SUD-M and SUD-C), a papain-like protease domain (PLpro) and a nucleic acid binding domain (NAB). In addition to Nsp3, Nsp4 and Nsp6 are also involved in double membrane vesicle formation and organization<sup>16,17</sup>. Nsp5 serves as the main protease (MPro), that proteolytically processes the ORF1a and ORF1b into final proteins<sup>18</sup>. Nsp7, Nsp8, and Nsp12 form the core replication complex with Nsp12 being the RNA dependent RNA polymerase (RdRp) and Nsp7/Nsp8 forming a hexadecameric complex, that enhances processivity<sup>19-21</sup>. In addition, Nsp9 also associates with the replication complex to promote 5'-



capping of viral RNA<sup>22</sup>. The addition of a 5' cap to the viral RNA is crucial for RNA stability as well as efficient translation and requires several steps, with the Nsp10/Nsp14 and Nsp10/Nsp16 complexes also contributing to the final steps of the process<sup>22</sup>. In both cases, Nsp10 activates the catalytic activity of Nsp14 and Nsp16. Apart from its function in the 5'-capping process, Nsp14 also functions as an exoribonuclease facilitating the proofreading function of the replication complex. Finally, Nsp13 is a helicase, that unwinds the RNA and promotes efficient transcription<sup>23</sup>, while Nsp15 serves as a uridine-specific endoribonuclease that facilitates the evasion of the host immune response<sup>24</sup>.

To investigate the virus-human protein-protein interactome of folded protein domains encoded by the SARS-CoV-2 genome and SLiMs in the human proteome, we created a collection of protein domain expression constructs from SARS-CoV-2 proteins. These protein domains were used as baits in proteomic peptide phage display (ProP-PD) screens against a Human Disorderome (HD2) peptide library that displays 1 million peptides from the human proteome<sup>25</sup>. The screen identified peptide-binding SARS-CoV-2 proteins and human peptide ligands that were validated with binding assays and tested in a SARS-CoV-2 infection assay. Our results suggest that protein domains of SARS-CoV-2 may serve as targets for the design and development of peptidomimetic antivirals.



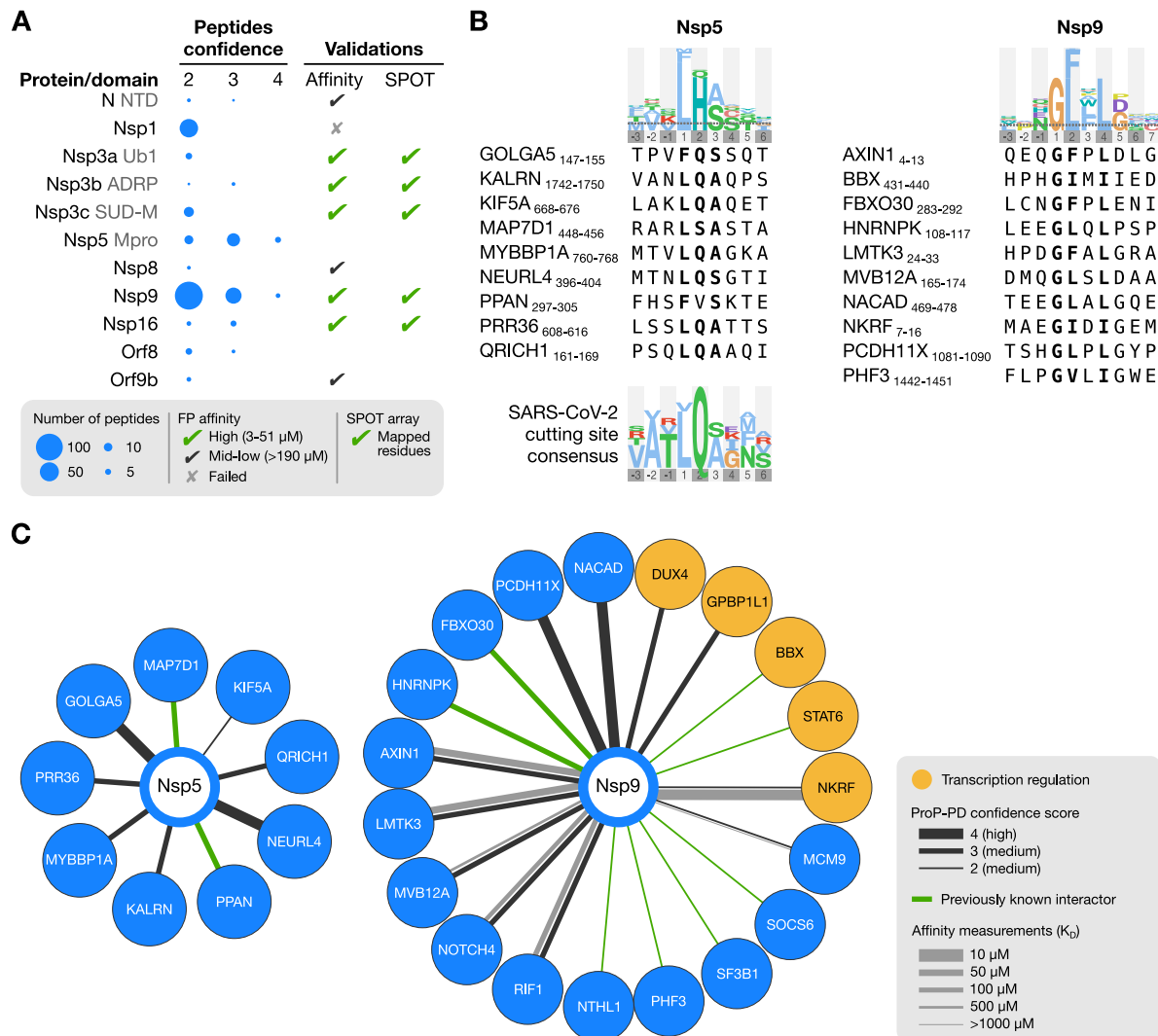
**Figure 1: Organization of the SARS-CoV-2 genome and proteome.** Proteins and protein domains that were successfully expressed in this study are shown in red. Proteins and protein domains that enriched ligands in ProP-PD selections are shown in bold and underlined. The figure is adapted from Gordon et. al.<sup>5</sup>

## Results

### Large-scale identification of SLiM-based interactions and interaction motifs of SARS-CoV-2 proteins

We generated a collection of 31 expression constructs of known or predicted modular domains from 22 SARS-CoV-2 proteins, including two catalytically inactivated protease variants (**Figure 1; Table S1**). The domains were expressed as GST-tagged fusion proteins, and 26 of the 31 protein constructs were produced in sufficient quantities to be used as baits in phage display selections. The GST-fusion proteins were immobilized for triplicate selections in a 96-well plate and used in ProP-PD selections using our HD2 peptide phage-display library. This library displays 16 amino acid overlapping peptides that tile the intrinsically disordered regions of the human proteome on the surface of M13 phage<sup>25</sup>. The peptide-coding regions of the binding-enriched phage pools were analyzed by next-generation sequencing (NGS). Using previously established quality metrics (peptides found in replicate selections, overlapping peptides identified, high NGS score and/or containing motifs) we found 281 high/medium

(8/273) confidence peptides from 239 proteins interacting with 11 SARS-CoV-2 domains of nine viral proteins (Nsp1, Nsp3 (Ubl1, ADRP and SUD-M), Nsp5, Nsp8, Nsp9, Nsp16, Orf8, Orf9b and N NTD) (**Figure 1, Table S2**). The vast majority (118) of identified interactions involve Nsp9, followed by Nsp1 (47) and the catalytically inactivated Nsp5 Mpro (32). Based on the peptides identified by ProP-PD selections, consensus motifs could be established for two proteins (Nsp5 [FLM][HQ][AS] and Nsp9 G[FL]xL[GDP]; **Figure 2**). For Nsp5 (Mpro), the ligands may serve as potential substrates because the recognition motif resembles the preferred proteolytic site of the protease (LQ↓[GAS])<sup>26</sup>.



**Figure 2: Outline of the results.** (A) Overview of the results for the 11 SARS-CoV-2 protein domains that enriched peptides with medium (2, 3) and high (4) confidence scores in ProP-PD selections. The number of peptides is proportional to the area of the blue circles. Overall validation of affinity with fluorescence polarization

and SPOT array is indicated. **(B)** The motifs obtained from the selections with the SARS-CoV-2 proteins Nsp5 and Nsp9 are highlighted in bold in the alignments of the representative peptides. The consensus motif for the Nsp5 (Mpro) protease cleavage site is shown below the alignment<sup>26</sup>. **(C)** Subset of identified interactions of SARS-CoV-2 Nsp5 and Nsp9. The thickness of dark lines shows ProP-PD confidence scores for the interactions, and the thickness of grey lines the  $K_D$  values obtained from the fluorescence polarization assays. Human proteins that contain the enriched GO terms associated with the transcriptional regulation (GO:0003700, GO:0000981, GO:0000976, GO:0001228, GO:0000978, GO:0000977, GO:0006355) are highlighted in orange.

The enrichment analysis of the Gene Ontology (GO) terms for the combined dataset revealed enrichment of ligands associated with biological processes related to transcriptional regulation (**Table S3**;  $q < 0.001$ ; PepTools). By comparing with the information available in protein interaction databases (collected August 2022, See Materials and Methods for details) we found that eighteen of the protein-protein interactions identified here are supported by reported interactions from previous studies (**Figure 2, Table S2**). The low overlap with other host-SARS-CoV-2 protein-protein interaction studies likely reflects technical differences between the approaches<sup>27</sup>. Here, we specifically searched for SLiM-based interactions, which are often missed by other methods<sup>25</sup>.

We selected 27 interactions for validation in a fluorescence polarization (FP) based binding assay, in which we first establish a valid fluorescein (FITC) labelled probe peptide, which was subsequently displaced in a second experiment with unlabelled peptide(s). The following SARS-CoV-2 protein domains were included in these binding experiments: Nsp1, Nsp3 Ubl1, Nsp3 ADRP, Nsp3 SUD-M, Nsp3 NAB, Nsp8, Nsp9, Nsp16, Orf9b, and N-NTD (**Table S4**). For Nsp1 and Nsp3 NAB we did not detect binding with the probe peptides tested whereas weak, non-saturating binding was observed for Nsp8, Orf9b and N NTD, indicating that the tested interactions are of low affinity (**Figure S1, Table S4**). In contrast, we conclusively confirmed the peptide binding capacity of five SARS-CoV-2 domains: Nsp3 Ubl1, Nsp3 ADRP, Nsp3 SUD-M, Nsp9 and Nsp16, as described below.

### Characterization of SLiMs binding to distinct domains of Nsp3

Coronavirus replication occurs primarily in double-membrane vesicles, derived from the host endoplasmic reticulum membrane, that provide protection from the host immune response<sup>28</sup>. Recently, the hexameric Nsp3 assembly was shown to form pores that span the double membrane and serve as a vital connection between the site of viral RNA synthesis inside the vesicle and the site of viral RNA translation in the cytoplasm<sup>29</sup>. Despite great recent efforts to map and characterize the SARS-CoV-2 interactome, the Nsp3 has been largely neglected due to its size and complexity, and therefore the functions of the various domains are poorly understood<sup>5,30</sup>. We expressed and purified all globular domains of Nsp3 and subjected them to Pro-PD selections (**Figure 1**). Phage selections yielded a set of medium confidence ligands for Nsp3 Ub11 (6 peptides), ADRP (3 peptides) and Nsp3 SUD-M (14 peptides) but without a clear consensus motif for any of the domains, in part due to the small number of retrieved peptides (**Table S2**). We determined the equilibrium dissociation constant ( $K_D$ ) using the FP assay for at least two human peptides for each of the three globular Nsp3 domains (**Figure 3A, Figure S1**), which confirmed the binding of six of the tested peptides with affinities in the range of 20 to 300  $\mu$ M (**Table S4**).

Three peptides derived from the protein NYNRIN (NYNRIN<sub>1031-1046</sub>; CPSLSEEILRCLSLHD), the Nuclear receptor coactivator 2 (NCOA2) (NCOA2<sub>1074-1089</sub>; SDEGALLDQLYLALRN), and from the Chromatin complexes subunit BAP18 (BAP18<sub>40-55</sub>; AKWTETEIEMLRAAVK) were selected for affinity measurements with the Nsp3 Ub11. The measurements confirmed the binding of Nsp3 Ub11 to NYNRIN<sub>1031-1046</sub> and NCOA2<sub>1074-1089</sub> (**Figure 3A, Table S4**), but not to BAP18<sub>40-55</sub>. The two binding peptides shared some sequence similarity, and possessed a putative LxxLxL motif, which is known to adopt alpha helical conformation upon binding<sup>31,32</sup>. One such interaction exists between the NCBD domain of

CBP/p300 and the CID domains of the NCOA2 protein family<sup>33</sup>. We therefore tested the binding of Nsp3 Ubl1 to the entire CBP interaction domain (residues 1071-1110) of NCOA2 as well as to the two human paralogs NCOA1 and NCOA3 (residues 924-965 and 1045-1086 respectively). We did not detect any displacement for the two paralogues suggesting that the conserved LxxLxL motif is not the main driver of the interaction (**Figure S1, Table S4**). To clarify which residues of the NCOA2 peptide are involved in the interaction, we performed a peptide SPOT array alanine scan, which revealed that the leucine residues in position 1 (P1), together with a tyrosine in position P5, was critical for binding (**Figure 3C**), indicating that the Nsp3 Ubl1 binding motif in the NCOA2<sub>1074-1089</sub> is LxxxY. This finding clarified the lack of binding of NCOA1 and NCOA2 paralogues in which valines are located at the corresponding tyrosine position. Similarly spaced hydrophobic residues are also found in the other Nsp3 Ubl1 binding peptides (**Figure S2**).

To test how the peptides might bind to Nsp3 Ubl1 we performed an *in silico* prediction using ColabFold, which is based on AlphaFold2<sup>34</sup>. Both the NCOA2<sub>1074-1089</sub> and the NYNRIN<sub>1031-1046</sub> peptides were predicted with a high per-residue confidence score (pLDDT > 80) to form an alpha helix bound in the hydrophobic binding pocket between alpha helix 1 ( $\alpha$ 1) and helix 3 ( $\alpha$ 3) of Nsp3 Ubl1 (**Figure 3B; Figure S3**). Intriguingly, a recently reported high-affinity interaction between the disordered region of the SARS-CoV-2 nucleoprotein (N; 219-LALLL-224) and Nsp3 Ubl1 also engages the same hydrophobic binding pocket of Nsp3 Ubl1<sup>9</sup>. In this case, the interaction is enhanced by an additional polar motif of N, that binds to Nsp3 Ubl1 in a distinct site. Therefore, to investigate if the ligands identified in this study directly compete with N for binding to Nsp3a Ubl1, we attempted to displace the FITC-labelled NCOA2<sub>1074-1089</sub> peptide probe from the Nsp3 Ubl1-probe complex with full length N. However, the FP signal increased upon addition of N to the premixed Nsp3 Ubl1-probe sample as well as

upon addition of N to the sample containing only the probe peptide, rendering the displacement experiment inconclusive (**Figure S1**).

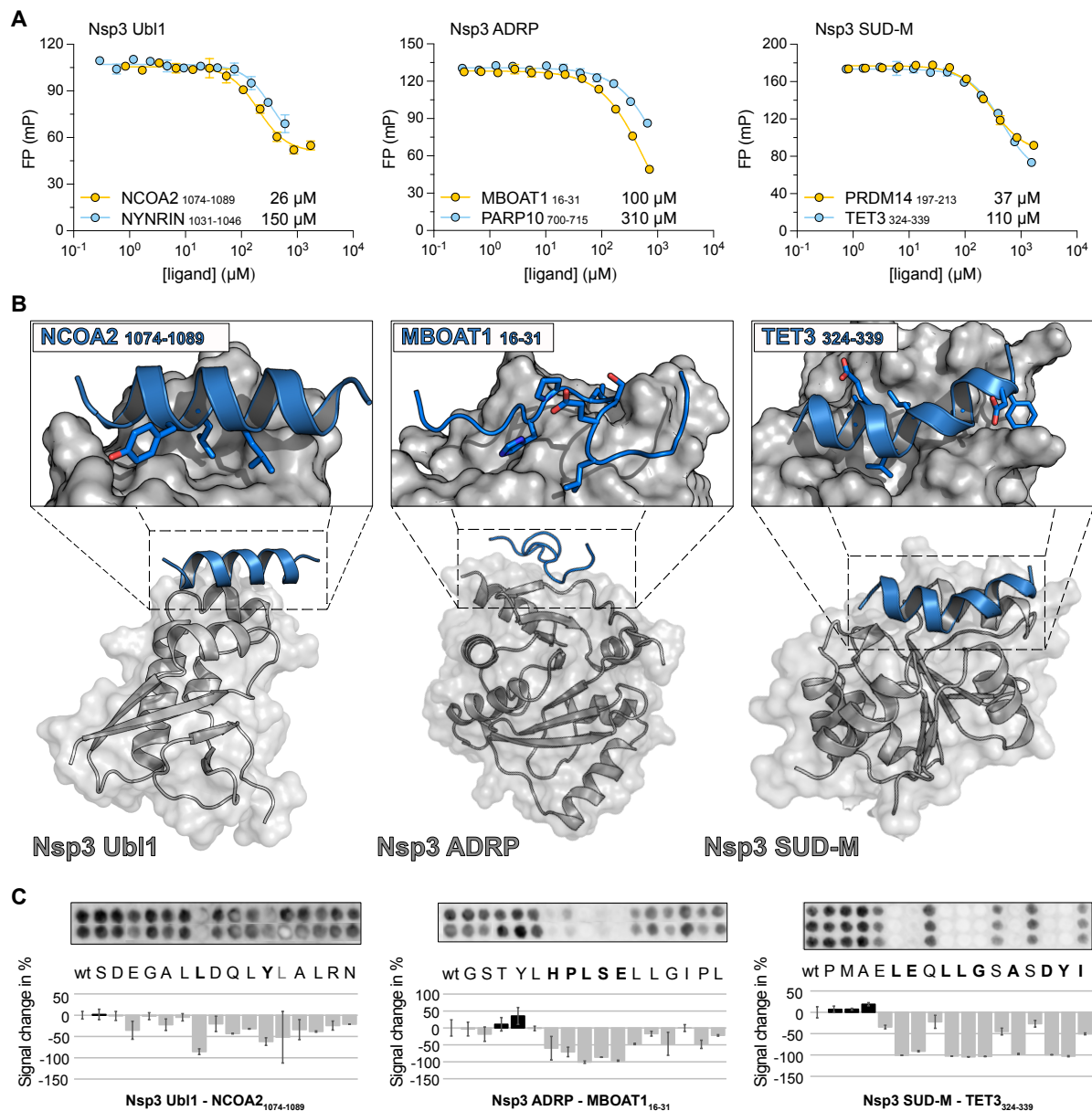
Next, we focused on the binding of peptides to the other two Nsp3 domains, ADRP and SUD-M. The peptides that bind to Nsp3 ADRP were derived from Lysophospholipid acyltransferase 1 (MBOAT1<sub>16-31</sub>; GSTYLHPLSELLGIPL) and Protein mono-ADP-ribosyltransferase PARP10 (PARP10<sub>700-715</sub>; DGGTDGKAQLVVHSA), with the MBOAT1 peptide being the ligand with higher affinity (**Figure 3A, Table S2 and Table S4**). The peptide SPOT array showed a strong signal for binding of the MBOAT1<sub>16-31</sub> peptide to Nsp3 ADRP, suggesting that the central **<sub>21</sub>-HPLSE-<sub>25</sub>** residues are critical for binding (**Figure 3C**, residues in bold). A similar **<sub>4</sub>-YLSE-<sub>7</sub>** segment was also identified in AZIN2, an additional Nsp3 ADRP ligand found in the ProP-PD selection (**Figure S2**). The ColabFold predictions of the Nsp3 ADRP binding peptides did not converge with high confidence (pLDDT of ~30; **Figure S3**). However, manual inspection indicated that the peptide binding was restricted to the surface between the N-terminal beta sheet ( $\beta$ 1) and the C-terminal alpha helix ( $\alpha$ 6) (**Figure 3B, Figure S3**). Finally, we note that PARP10 is upregulated upon SARS-CoV-2 infection<sup>35</sup> and is known to mono-ADP-ribosylate amino acid residues as a part of host response to viral infections<sup>36</sup>. The Nsp3 SUD-N (also called Mac1) domain of SARS-CoV-2 is in turn a mono-ADP-ribosylhydrolase<sup>37</sup>. Thus, although the affinity of the Nsp3 ADRP-PARP10 interaction is relatively low ( $K_D \approx 300 \mu\text{M}$ ), the interaction might be of relevance in the context of the virus counteracting the host's response to viral infection.

The peptides used to study binding to Nsp3 SUD-M were from the PR domain zinc finger protein 14 (PRDM14<sub>197-213</sub>; FTEEDLHFVLYGVTPS) and Methylcytosine dioxygenase TET3 (TET3<sub>324-339</sub>; PMAELEQLLGSASDYI). These were the most enriched peptides in the selection against Nsp3 SUD-M, and of these two, PRDM14<sub>197-213</sub> was the ligand with higher affinity (**Figure 3A, Table S2 and Table S4**). The ColabFold analysis of Nsp3 SUD-M in

complex with either PRDM14<sub>197-213</sub> or TET3<sub>324-339</sub> exhibited moderate pLDDT scores (between 50 and 70, **Figure S3**). While the two peptides did not share a specific binding motif, the SPOT array alanine scan showed distinct patterns of alternating amino acid residues interacting with Nsp3 SUD-M for both peptides (**Figure 3C, Figure S4**). In the case of Nsp3 SUD-M - TET3<sub>324-339</sub> the ColabFold predictions and SPOT arrays converged allowing us to map a plausible binding interface, where TET3<sub>324-339</sub> forms an alpha helix exposing the underlined motif 324-PMAELEQLLGSASDYI-339 for interaction with a hydrophobic pocket of SUD-M (**Figure 3B and 3C**).

Taken together we demonstrate that the Ubl1, ADRP, and SUD-M domains of Nsp3 possess the capacity to interact with SLiMs. The identified peptides bound with intermediate micromolar affinity and the binding sites on the peptides and the proteins were defined by alanine scans and docking predictions.





**Figure 3: Biophysical analysis of the interactions between Nsp3 Ubl1, Nsp3 ADRP, and Nsp3 SUD-M, with peptide ligands from human proteins.** (A) Fluorescence polarization-monitored displacement experiments measuring the affinity between globular domains of Nsp3 and peptide ligands from disordered regions of the human proteome identified by phage display. The  $K_D$  values are shown next to each peptide and in **Table S4**. (B) ColabFold structural predictions for the interaction between the globular domains of Nsp3 and the peptide ligands. The globular domains of Nsp3 are shown in grey, whereas the peptides are shown as blue ribbons. The ColabFold pLDDT confidence scores were high for the globular domains of Nsp3 (>90) but varied widely for peptide predictions (30-80) and are shown in **Figure S3**. (C) SPOT array alanine scans for the indicated peptides. Residues involved in binding are shown in bold. Signal intensities were normalized to wild type (wt) and presented as average percent signal change. Error bars indicate one standard deviation from the average.

## **Nsp9 binds to peptide ligands with consensus motifs, leading to the rearrangement of hydrophobic core and the disruption of dimerization.**

Nsp9 from SARS-CoV-2 is a 113 amino acid residue long RNA-binding protein with a stable fold that shares 98% sequence identity with its SARS-CoV homolog<sup>38,39</sup>. Recent reports have established that Nsp9 is an essential component of the replication and transcription complex where it interacts with the Nidovirus RdRp-Associated Nucleotidyltransferase (NiRAN) domain of the RNA-dependent RNA polymerase Nsp12<sup>40</sup>. In this model, the monomeric Nsp9 binds to NiRAN via a C-terminal alpha helix and facilitates the addition of the GpppA-RNA cap to the 5'-end of the newly synthesized mRNA<sup>22</sup>. The same helix has been suggested to facilitate the formation of the Nsp9 homodimer with the crucial <sub>100</sub>-GxxxG-<sub>104</sub> motif at the dimer interface<sup>41-45</sup>.

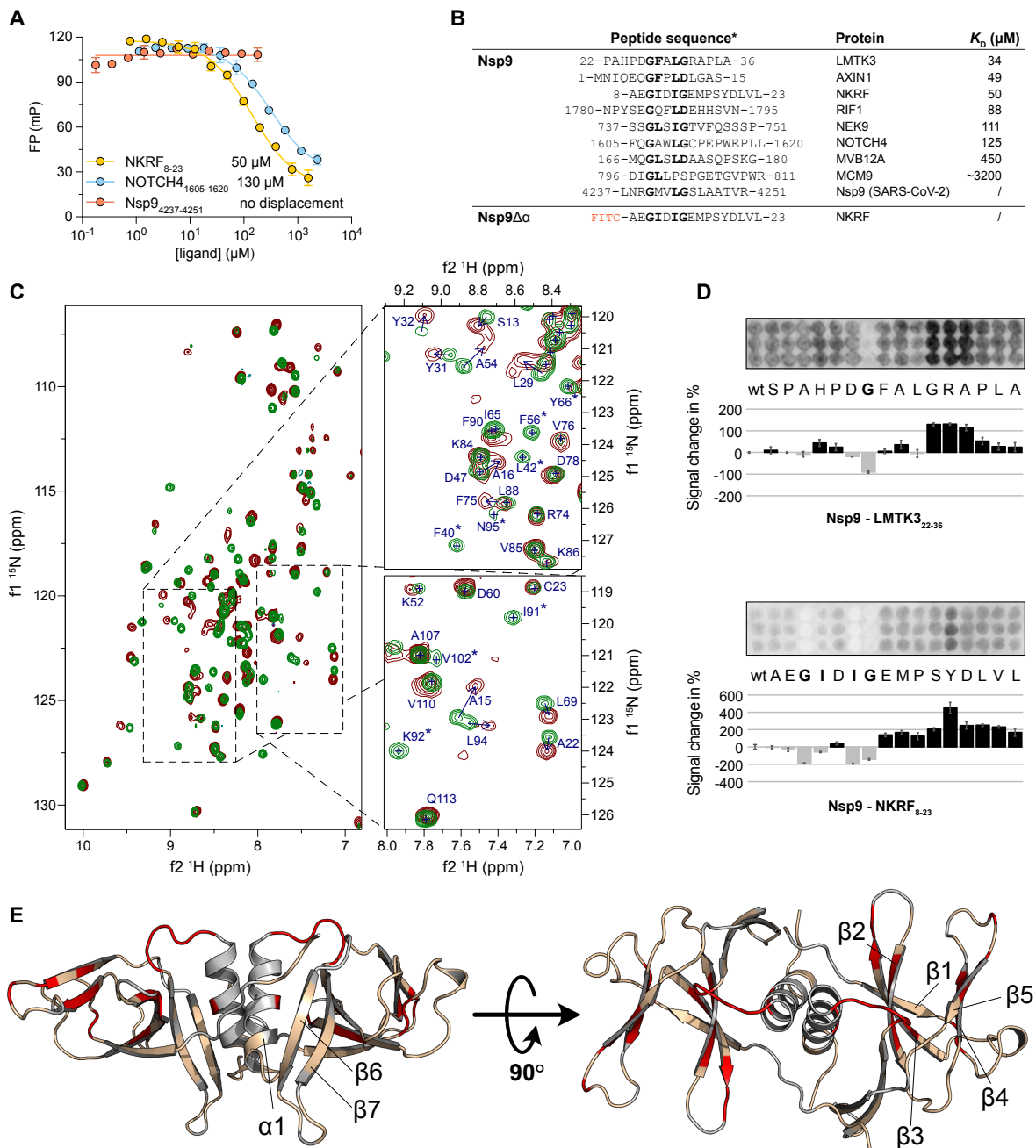
In our ProP-PD selections, we identified 147 human Nsp9 binding peptides in which a G[FL]xL[GDP] motif consensus was enriched (**Figure 2; Table S2**). We confirmed binding and determined the affinities for eight of the ligands, with the  $K_D$  values in the mid-micromolar range (**Figure 4A-B, Figure S1, Table S4**). The ligands with the highest affinity were peptides derived from the NF-kappa-B-repressing factor (NKRF<sub>8-23</sub>; AEGIDIGEMPSYDLVL), the protein kinase LMTK3 (LMTK3<sub>22-36</sub>; PAHPDGFALGRAPLA), and the Axin-1 (AXIN1<sub>1-15</sub>; MNIQEQQGFPLDLGAS), which bound to Nsp9 with  $K_D$  values of 30-50  $\mu$ M (**Table S4**). Of these, NKRF has been found to interact with other SARS-CoV-2 proteins including Nsp1<sup>46</sup> and Nsp10<sup>47</sup> with the latter interaction thought to regulate interleukin-8 production. A peptide from the Neurogenic locus notch homolog protein 4 (NOTCH4<sub>1605-1620</sub>) displayed lower affinity, and bound to Nsp9 with a  $K_D \sim 130 \mu$ M. We further tested binding to the C-terminal peptide of Nsp9 (Nsp9<sub>4237-4251</sub>; LNRGMVLGSLAATVR), as it shares the GxxxG motif, but did not detect any measurable binding within the concentration range used (**Figure 4A**). To further dissect which

residues of the peptides are essential for binding we performed a SPOT array alanine scan of the NKRF<sub>8-23</sub> and LMTK3<sub>22-36</sub> peptides. For the NKRF<sub>8-23</sub> peptide the analysis confirmed the consensus motif, with a substantial decrease in binding intensity when either of the two glycines at the positions P1 and P5 of the motif or the isoleucine at the position P4 were mutated to alanine. A minor effect was also observed upon mutation of the isoleucine at the position P2 (**Figure 4D**). Based on these results, we adjust the motif to GΦxΦ[GDP], where Φ is a hydrophobic residue. To define the binding pocket of the motif, we performed ColabFold prediction for peptide binding to either monomeric or dimeric Nsp9. This analysis however, showed low confidence scores (pLDDT < 30; **Figure S5**) and did not converge on a similar binding mode for the three peptides.

Because the peptide ligands have a GΦxΦ[GDP] motif, that is also found in the C-terminal helix of Nsp9 and facilitates dimerization, we hypothesized that Nsp9-binding peptides might interact with the dimerization interface and therefore interfere with the dimer formation<sup>41,48</sup>. To address this question and map the residues involved in the interaction between Nsp9 and the peptides, we recorded nuclear magnetic resonance (NMR) <sup>1</sup>H<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectra of <sup>15</sup>N labeled Nsp9 at increasing concentrations of NKRF<sub>8-23</sub>, LMTK3<sub>22-36</sub>, and NOTCH4<sub>1605-1620</sub>, respectively. Well-resolved spectra, which were in excellent agreement with the previously described NMR assignments<sup>44,48</sup>, confirmed a folded protein and allowed us to monitor the changes in chemical shift perturbations upon addition of peptides (**Figure 4C, Figure S6, Table S5**). The substantial overlap of chemical shift perturbation changes upon binding of the three peptides suggested an overall conserved binding interface. We focused on the residues that exhibited chemical shift changes of at least one standard deviation above average for all three peptides. The data suggested significant rearrangements of residues in the hydrophobic core of Nsp9 in the β1-β5 region while β6 and β7 remained unperturbed (**Figure 4E, Figure S6H-I**). Importantly,

chemical shifts for the C-terminal alpha helix were not observed as discussed previously<sup>44,48</sup>, leaving open the possibility that the peptides could interact with the helix without showing any perturbations in the NMR HSQC spectra. To determine whether the Nsp9 is a dimer in solution, we measured the  $T_1$  and  $T_2$  relaxation times, which report on global motion of individual residues of the protein (**Table S6**). The ratio of the two relaxation times, gives the rotational correlation time  $\tau_c$ , which was 15.3 ns for the free Nsp9 protein and 9.9 ns for the protein in complex with NOTCH4<sub>1605-1620</sub> peptide corresponding to estimated molecular weights of 29 and 18.8 kDa respectively. Because the molecular weight calculation from  $\tau_c$  is highly dependent on the shape of the molecule, the experiments provide an estimate of the size, which nonetheless correlates well with the predicted molecular weight for a Nsp9 homodimer (25.7 kDa) and a Nsp9 monomer bound to the NOTCH4 peptide (14.7 kDa) implying that the addition of the peptide disrupts formation of the Nsp9 homodimer. To further test whether the peptides bind via the C-terminal helix we constructed a Nsp9 $\Delta\alpha$  construct, which is missing 20 residues of the C-terminus (**Table S1**). As expected, the Nsp9 $\Delta\alpha$  did not interact with the FITC-NKRF<sub>8-23</sub> probe (**Figure S1, Table S4**), corroborating our hypothesis that the identified peptides bind to the C-terminal helix of Nsp9 and interfere with Nsp9 homodimer formation.

In summary, we identified short peptide binders that interact with Nsp9 via a G $\Phi$ x $\Phi$ [GDP] motif, which also encompasses the GxxxG motif found in the C-terminal protein-protein interaction region of Nsp9. This motif is present in many human proteins (**Figure 2B, Figure 4B**), and the interaction with these proteins could affect viral replication. Using NMR, we located the peptide binding interface and showed that peptide binding perturbs packing of the Nsp9 hydrophobic core, consequently interfering with dimer formation. Since Nsp9 interacts with NiRAN through the same C-terminal helix, this interaction, which is critical for viral replication, should also be disrupted<sup>22</sup>.



**Figure 4: Interactions between Nsp9 and human peptide ligands are mediated by a  $G\Phi x\Phi[GDP]$  motif and lead to rearrangement of the hydrophobic core of Nsp9. (A)** Representative FP-monitored displacement experiments measuring the affinity between Nsp9 and three human peptide ligands. Affinities are shown next to each peptide. **(B)** Alignment of peptides identified by ProP-PD for which the affinities were measured. Residues, corresponding to the identified  $G\Phi x\Phi[GDP]$  motif are shown in bold. For Nsp9Δα we did not observe saturation with the FITC-NKRF<sub>8-23</sub> peptide. **(C)** HSQC experiments showing the  $^1H$ - $^{15}N$  peak shifts. The left panel shows the superposition of Nsp9 protein (green) and Nsp9 mixed with NKRF peptide (wine red). The right panels show representative examples of the observed chemical peak shift perturbations. The arrows indicate the directions of

the chemical shift change and the asterisk indicates that the peaks disappeared after addition of the peptide, indicating a large perturbation of the chemical environment. The full spectra are shown in **Figure S6. (D)** SPOT array alanine scans for the indicated peptides. Residues involved in binding are shown in bold. Signal intensities were normalized to wild type (wt) and presented as percentage signal change. **(E)** Representation of the residues which displayed large change in chemical shift after addition of the NKRF<sub>8-23</sub> peptide, as observed by NMR experiments. Residues that changed more than one standard deviation above the average are colored in red, the residues below this threshold are beige, and the residues whose peak shifts could not be unambiguously assigned are gray. The PDBid: 6wxd model of Nsp9 dimer was used for visualization.

### **Identified peptide ligands directly interfere with the Nsp10-Nsp16 complex.**

Since the SARS-CoV-2 virus does not have access to the cellular mRNA capping machinery, its genome encodes a full set of enzymes that facilitate complete 5' capping of viral RNA, thereby promoting translation and evasion of the immune response<sup>22,49,50</sup>. In the final step of this process the methyltransferase Nsp16 catalyzes the C2'-O methylation of the first nucleotide of the RNA, forming a fully functional <sup>7</sup>MeGpppA<sub>2'-O-Me</sub>-RNA<sup>22,51,52</sup>. To catalyse this reaction Nsp16 forms a complex with Nsp10, which in turn activates Nsp16 by stabilizing the vital S-adenosylmethionine binding pocket<sup>49,51</sup>.

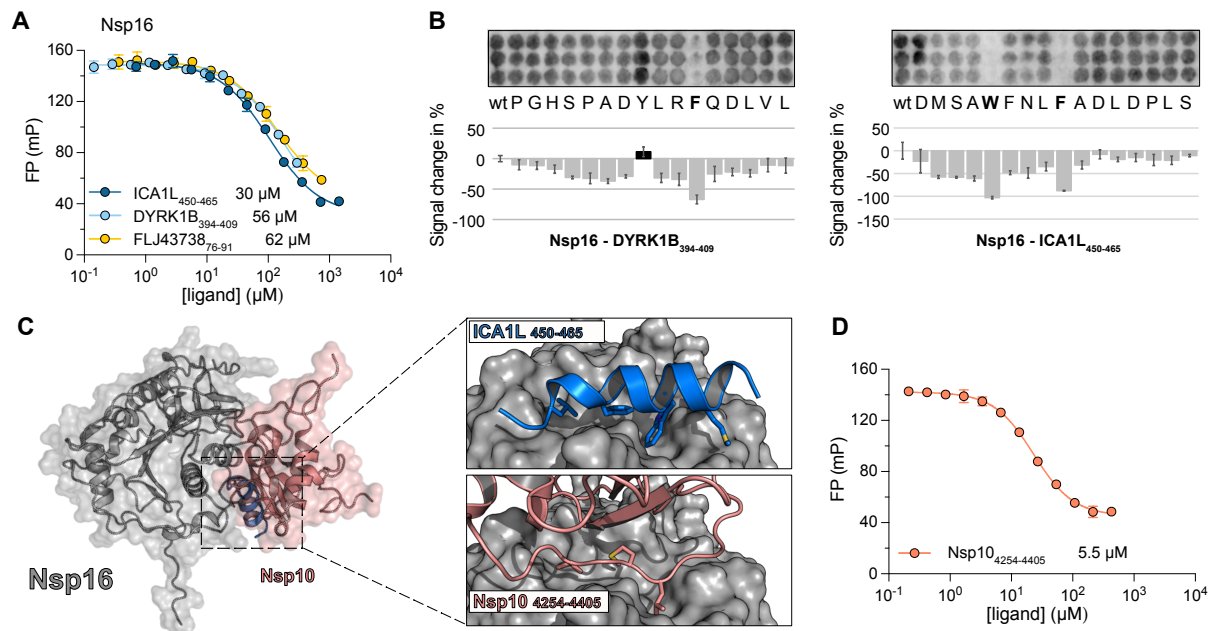
While ProP-PD selections against Nsp10 did not result in specific enrichment of peptides, we identified peptides from six proteins that bind to Nsp16 (**Table S2, Figure S2**) of which we selected three peptides for affinity measurements. These peptides were derived from Dual specificity tyrosine-phosphorylation-regulated kinase 1B (DYRK1B<sub>395-410</sub>; PGHSPADYLRFDLVL), Islet cell autoantigen 1-like protein (ICA1L<sub>450-465</sub>; NQDMSAWFNLFDLDP), and uncharacterized protein FLJ43738 (FLJ43738<sub>76-91</sub>; EDPLDSYLNFDLISP). Using the FP assay, we confirmed the interactions and determined that the peptides interacted with Nsp16 with  $K_D$  values of 56  $\mu$ M, 30  $\mu$ M and 62  $\mu$ M, respectively (**Figure 5A**). Interestingly, the tightest binder, FITC-labelled DYRK1B (FITC-DYRK1B<sub>394-409</sub>), bound with a  $K_D$  of  $\sim$ 3  $\mu$ M, exhibiting a 10-fold higher affinity compared to



the unlabeled peptide with the same sequence (DYRK1B<sub>394-409</sub>) (**Figure S1, Table S4**) implying that the FITC label itself contributes to binding. Alanine scan SPOT array analysis of ICA1L<sub>450-465</sub> and DYRK1B<sub>395-410</sub> peptides revealed the presence of a hydrophobic motif (WxxxF) in the ICA1L<sub>450-465</sub> peptide, which is also present in the peptide from the homologous ICA1 (**Figure 5B, Figure S2**). For the DYRK1B<sub>395-410</sub> peptide the SPOT array analysis established the importance of a phenylalanine in P1 but did not confirm the second position of the hydrophobic motif (P5, which in this case would be a valine (**Figure 5B, Figure S2**)). A putative [WF]xxxΦ motif is also present in the validated binding peptide from FLJ43738<sub>76-91</sub> (**Figure S2**). ColabFold predicted that the three validated peptides form an alpha helix and bind to the hydrophobic groove formed by alpha helices α3, α4 and α10 of Nsp16 (**Figure 5C, Figure S7**). This result correlated with the SPOT array data and corroborated the importance of hydrophobic residues at the P1 and P5 positions of the Nsp16 binding motif.

Since the same hydrophobic interface of Nsp16 facilitates the interaction with Nsp10<sup>52</sup> we hypothesized that our newly characterized peptide ligands could compete with Nsp10 for binding to Nsp16 (**Figure 5C**). Indeed, we demonstrated that full-length Nsp10 directly competes with the labelled FITC-DYRK1B<sub>394-409</sub> peptide in a displacement experiment (**Figure 5D**). This experiment also allowed us to determine the  $K_D$  of the interaction between Nsp10 and Nsp16 to be 5.5 μM.

Overall, we identified several peptide binders that interact with Nsp16, measured their affinity, and determined their binding interface with Nsp16. The peptide ligands compete with Nsp10 for binding to Nsp16 and could therefore act as inhibitors of the essential Nsp10-Nsp16 interaction, thereby hampering viral replication.



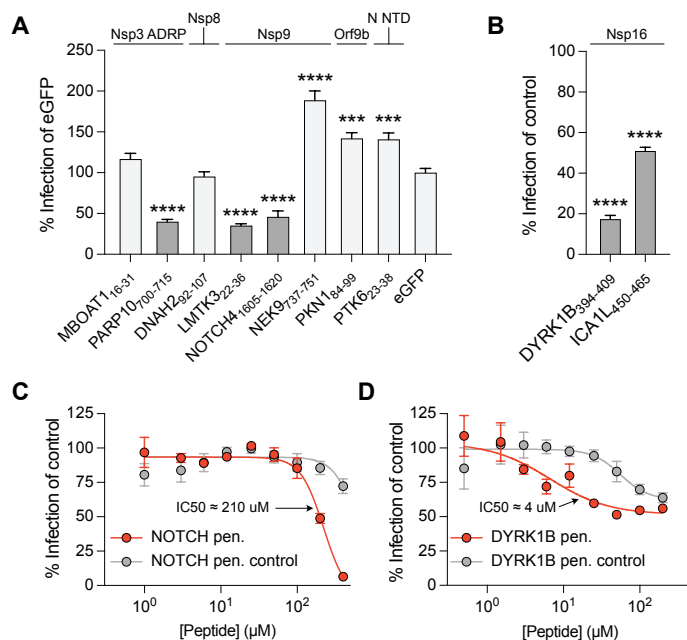
**Figure 5: Peptides identified with ProP-PD directly compete with Nsp10 for binding to Nsp16.** (A) Displacement experiments measuring affinity between Nsp16, and three peptide ligands derived from human proteins.  $K_D$  values are shown next to each peptide. (B) SPOT array alanine scans for the DYRK1B<sub>394-410</sub> and ICA1L<sub>450-465</sub> peptides. Residues involved in binding are shown in bold. Signal intensities were normalized to wild type (wt) and presented as percent signal change. (C) ColabFold structural predictions for the interaction between Nsp16 (grey), and the ICA1L<sub>450-465</sub> peptide (blue). The predicted structure is superimposed onto the crystal structure of the Nsp16-Nsp10 complex (PDBid: 7jyy). The right panels show a magnified view of the peptide binding pocket highlighting the similarities of the Nsp16 binding to the peptide or Nsp10, respectively. (D) FP-monitored displacement by Nsp10 of a complex between Nsp16 and FITC-DYRK1B<sub>394-409</sub> shows that Nsp10 and the peptide binds to the same surface of Nsp16. The  $K_D$  value for the Nsp10-Nsp16 interaction is indicated.

### Peptide ligands targeting Nsp3 ADRP, Nsp9 and Nsp16 inhibit viral proliferation.

To investigate the role of the identified protein-protein interactions in viral replication, we selected 11 of the peptide ligands for further experiments in virus infection assays. We designed lentiviral expression constructs, expressing four repeats of each peptide interspaced by flexible GlySerThr linkers and conjugated C-terminally to an enhanced green fluorescent protein (Table S7). VeroE6 cells were first transduced with the lentiviruses and, after 72 hours, infected



with SARS-CoV-2 for 16 hours, after which the number of infected cells was measured. We found that five of the lentiviral constructs inhibited the production of viral particles, namely the Nsp3 ADRP binding EGFP-PARP10 construct, the Nsp9 binding EGFP-NOTCH4 and EGFP-LMTK3 constructs, and the Nsp16 binding EGFP-DYRK1B and EGFP-ICA1L constructs (**Figure 6A-B**). Additionally, we confirmed the interference of the ligands with viral replication by treating the infected cells with the NOTCH4 and DYRK1B-derived peptides fused to the cell-penetrating HIV Tat-derived sequence (YGRKKRRQRRRGSG) (**Table S8**). These experiments confirmed the inhibitory effects of the Nsp9 binding NOTCH4, as well as the Nsp16 binding DYRK1B peptides (**Figure 6C**), establishing these ligands as potential starting points for the development of peptidomimetic inhibitors of SARS-CoV-2 infection. To investigate if the peptides specifically inhibit SARS-CoV-2 replication, we also treated human coronavirus 229E (HCoV-229E) infected MRC5 cells with the cell penetrating constructs. Interestingly, the peptides failed to inhibit HCoV-229E infection (**Figure S8**), suggesting a beta-coronavirus specific inhibitory effect.



**Figure 6. Inhibition of SARS-CoV-2 viral infection propagation by lentiviral constructs or cell penetrating peptides.** (A) Inhibition is shown as percent infection of eGFP. Significance was determined using an unpaired t-

test; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ . Highlighted columns represent constructs, that showed significant inhibition. The target proteins of SARS-CoV-2 are stated above. **(B)** Inhibitions by lentiviral constructs expressing Nsp16-targeting peptides are shown as percent infection of control. In the control construct the key interacting residues are mutated to alanine. A complete list of the lentiviral constructs can be found in **Table S7**. Significance was determined using an unpaired t- test; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ . **(C)** A serial dilution of either the cell penetrating NOTCH4 peptide (red dots) or the cell penetrating NOTCH4 control peptide (gray dots) was added to SARS-CoV-2 infected VeroE6 cells. The IC<sub>50</sub> for the cell penetrating NOTCH4 peptide is indicated. **(D)** The same experiment was repeated but with either the cell penetrating DYRK1B peptide (red dots) or the cell penetrating DYRK1B control peptide (gray dots). The IC<sub>50</sub> values shown were calculated using a sigmoidal dose response fitting equation.

## Discussion

In the present work, we performed a large-scale exploration of SLiM-mediated interactions between globular domains of SARS-CoV-2 proteins and peptides found in the intrinsically disordered regions of the human proteome. We aimed to answer three main questions: (1) How widespread are the SLiM-based interactions of viral proteins; (2) What are the human SLiM-containing binding partners of SARS-CoV-2 proteins; and (3) Can we exploit the newly identified peptide binders as antiviral agents.

Out of 26 SARS-CoV-2 protein domain constructs that we were able to express and purify, 11 showed phage enrichment from ProP-PD screens, and in total we identified 281 high/medium confidence human proteome-derived peptides that interact with SARS-CoV-2 protein domains. Several of the interactions between SARS-CoV-2 proteins are mediated through SLiMs and facilitate correct localization (Nsp3-N)<sup>9</sup>, complex formation (Nsp9-Nsp12)<sup>22</sup>, or regulate the activity of other viral proteins (Nsp10-Nsp16)<sup>53,54</sup>. Our results show that the same viral protein domains, which facilitate SLiM-mediated interactions between viral proteins, can also engage in interactions with human SLiMs. This may result in competition between host and viral SLiMs for binding to Nsp3, Nsp9 and Nsp16, potentially impacting the rate of infection processes such as viral RNA capping.

We found that Nsp9 can bind to several human proteins that contain GΦxΦ[GDP] motifs, many of which are associated with transcriptional regulation, suggesting a possible biological function of these interactions during the viral life cycle. While these findings are intriguing, it is not clear which human globular domain(s), if any, bind to the native, human GΦxΦ[GDP] motifs. Similarly, the C-terminal domain of the Ebola VP30 protein has been shown to bind to PPxPxY containing peptides found in both viral (N) and human proteins<sup>10,55</sup>. As with the Nsp9 binding GΦxΦ[GDP] motif, it is not yet known which human protein(s) binds to the PPxPxY motif in a functional context. The discovery of such human SLiMs binding to

viral proteins can provide an alternative starting point in the search for novel SLiM binding domains and thus contribute to our understanding of human SLiM based interactions.

Since the start of the pandemic, there has been a rush to characterize SARS-CoV-2 protein interactions. Many of the large-scale studies have employed low-resolution techniques such as mass spectroscopy<sup>5,27,30,47,56-58</sup> that provide information on binary interactions as well as on larger complexes. Furthermore, several studies aimed to identify potential inhibitors of viral infection, mainly targeting RNA-dependent-RNA polymerase<sup>59-62</sup> and the proteases<sup>63-65</sup>, or producing antibodies against spike protein<sup>66-70</sup>. However, to increase the likelihood of successful drug development, efforts need to be extended to include other SARS-CoV-2 proteins. The Nsp9 dimer interface in the C-terminal helix with the GxxxG motif at its core has been proposed as a valid target for inhibitor development<sup>71,72</sup>. Our results confirmed that the Nsp9 binding GxxxG-containing peptides from LMTK3 and NOTCH4 indeed have antiviral effects, confirming the validity of the approach. The coronavirus Nsp10/Nsp16 interaction has also been shown to be a valid target for antiviral peptides, using Nsp10-derived peptides<sup>53,54,73</sup>. Similarly, we found that our NSP16 binding peptides had an antiviral effect. Other strategies targeting the Nsp10/Nsp16 complex include methyl donor site<sup>51,71</sup> and RNA binding site targeting, as reviewed recently<sup>74</sup>.

In summary, we have demonstrated that ProP-PD screening is a viable strategy for identification of human peptides that bind to the globular domains of the SARS-CoV-2 proteome. We also showed that a subset of identified ligands inhibited viral replication in cell culture, and that these peptides could be successfully converted into cell-penetrating anti-viral inhibitors. Thus, our study expands the available peptide repertoire that may be used as starting point for future drug discovery targeting coronaviruses.

## Experimental procedures

### Protein expression and purification

The cDNAs encoding SARS-CoV-2 protein domains were ordered from Genescript in pETM33 expression plasmids (Table S1). The plasmid encoded an N-terminal His-tagged GST, a PreScission protease cleavage site and the SARS-CoV-2 protein of interest. The proteins were expressed in BL21(DE3) gold *E. coli*. Bacteria was grown in 2YT medium (16 mg/ml peptone, 10 mg/ml yeast extract, 5 mg/ml NaCl) supplemented with 50 µg/ml kanamycin until OD<sub>600</sub> reached 0.6, when the protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Proteins were expressed overnight at 18°C, bacterial cultures were harvested by centrifugation (4,500 g, 10 minutes) at 4°C and resuspended in lysis buffer A (50 mM Tris/HCl pH 7.8, 300 mM NaCl, 10 µg/ml DNase I and RNase, 4 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub> and cComplete EDTA-free Protease Inhibitor Cocktail). For ProP-PD selections, protein domains were purified on a GSH affinity column (Pierce glutathione agarose) according to the manufacturer's instructions. Following elution with 10 mM GSH in buffer A, the sample was used in selections where the His-GST moiety was immobilized on the plate according to the protocol described previously<sup>25</sup>. For fluorescence polarization experiments, after the initial GSH affinity purification step, the His-GST was cleaved off by PreScission protease (1:100 dilution; produced in-house) overnight at 4 °C. Following the cleavage, the samples were applied on a nickel Sepharose excel resin column to remove the His-GST tag. The SARS-CoV-2 protein was recovered in unbound fraction from the nickel column. The purity of the samples was inspected by SDS-PAGE and if needed an additional purification step was introduced, where the SARS-CoV-2 protein sample was applied to a size-exclusion chromatography column (Superdex 75, Cytiva). The identity of pure protein samples was confirmed with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS).

Finally, protein samples were dialyzed against 50 mM potassium phosphate buffer pH 7.4 and flash frozen until further use.

For NMR experiments the Nsp9 expression construct was expressed in M9 minimal medium containing 1 g/l  $^{15}\text{NH}_4\text{Cl}$ . After  $\text{OD}_{600}$  reached 0.6 the protein expression was induced with 1 mM IPTG, and expressed overnight at 18 °C. After expression, the purification protocol was the same as described above.

### **Proteomic peptide phage display**

We recently published the design of a Pro-PD library expressing disordered regions found in the human proteome and a pipeline to quickly analyze data from deep sequencing of enriched phages<sup>25,75</sup>. Briefly, GST-tagged bait proteins were immobilized in a 96-well Flat bottom Nunc MaxiSorp plates (Nunc, Roskilde, Denmark) at 4°C for 18 hours (10 µg per protein in 100 µl PBS). GST was immobilized in adjacent well as control. After immobilization the wells were blocked with blocking buffer (0.5% (w/v) BSA in PBS) for 1 hour at 4°C and washed 4 times with PT buffer (PBS supplemented with 0.05% Tween20). The phage library (prepared according to Benz et. al.<sup>25</sup>) was added (100 µl in PBS supplemented with 1 mM DTT) to the wells with immobilized GST for 1 hour, to remove nonspecifically binding phages, before being transferred to the wells with bait proteins. After 2 hours the unbound phages were removed by washing the wells 5 times with 200 µl of PT buffer, and the bound phages were eluted by the addition of 100 µl of *E. coli* OmiMAX in log phase with subsequent incubation for 30 minutes at 37°C. Bacteria were hyperinfected by the addition of M13KO7 helper phage and further incubated for 45 minutes at 37 °C. After successful infection, 100 µl of bacteria were transferred into 1 ml of 2xYT media supplemented with 100 µg/ml carbenicillin, 30 µg/ml kanamycin and 0.3 mM IPTG, followed by incubation for 18 hours at 37 °C under agitation. Next, bacteria were pelleted at 2000 g for 10 minutes and phage supernatant was transferred to a fresh 96-well

plate, where the pH was adjusted by the addition of 1/10 volume of 10 x PBS and 1 mM DTT (final concentration), and heat inactivated at 65°C for 10 minutes. The resulting phages were used next day as in-phages in the second round of selection, where the whole procedure was repeated. To ensure high enrichment of binding phages the selection procedure was repeated four times. After the final day of selection, the phage enrichment was evaluated by pooled phage ELISA in 384-well Flat bottom Nunc MaxiSorp plates (Nunc, Roskilde, Denmark). Bait proteins and GST control were immobilized (5 µg in 50 µl PBS per well) at 4°C for 18 hours followed by blocking of the remaining well surface with 100 µl of 0.5% BSA in PBS, at 4 °C for 1 hour. Enriched phages from third and fourth rounds of selections (50 µl) were incubated with the corresponding bait protein for 1 hour and the unbound phages were washed away 4 times with 100 µl PT buffer. Bound phages were incubated for 1 hour with 50 µl of M13 HRP-conjugated M13 bacteriophage antibody (Sino Biological Inc; 1:5,000 diluted in PT, 0.05% BSA). The wells were again washed 4 times with PT buffer followed by one wash with PBS. Substrate was added (40 µl of TMB substrate, Seracare KPL) and the enzymatic reaction was quenched by the addition of 40 µl of 0.6 M sulfuric acid. Finally, the absorbance at 450 nm was measured to quantify the phage enrichment with an iD5 plate reader (Molecular Devices).

### **NGS analysis and ProP-PD data analysis**

Peptide-coding regions were PCR-amplified and barcoded from the binding-enriched phage pools (5 µL) using Phusion High-Fidelity polymerase (Thermo Scientific). PCR products were normalized using Mag-bind Total Pure NGS, purified from a 2% agarose gel (QIAquick Gel extraction Kit), and analyzed using Illumina MiSeq v3 (1x150 bp read setup, 20% PhiX). Reads were demultiplexed, adapter and barcode regions were trimmed, and sequences were translated into peptide sequences. Peptides were annotated using PepTools. Confidence levels were assigned based on four different criteria: occurrence in replicate selections, identification of

overlapping peptide sequences, high counts, and occurrence of sequences matching consensus motifs, as previously outlined<sup>25</sup>.

Protein interaction networks were built with Cytoscape 3.9.1<sup>76</sup>. Inkscape 1.2.1 (<https://inkscape.org>) and Python Matplotlib 3.5.1 library (<https://matplotlib.org>) were used to build the figures. For the comparison of ProP-PD results with other protein-protein interaction datasets the previously reported evidence of protein-protein interactions was obtained from Biogrid (release 4.4.212)<sup>77</sup>, VirHostNet (release 3.0)<sup>78</sup>, IMEx (release 1.4.11)<sup>79</sup>, IntAct (release 4.2.18)<sup>80</sup> and MINT (August 2022)<sup>81</sup>

### **Fluorescence polarization**

Fluorescence polarization experiments were performed as previously described in detail<sup>7</sup>. Briefly, peptides were ordered either unlabeled or as FITC-labelled constructs from GeneCust. First,  $K_D$  of the FITC labelled peptide for a certain interaction was determined by varying the protein concentration at a constant FITC-peptide concentration and fitting a hyperbolic function to the data. Next, a displacement experiment was performed to determine  $K_D$  for the unlabelled peptide. Labelled peptide (10 nM) was pre-incubated with protein such that approximately 50-60% of the labelled peptide was bound to protein (Nsp3 Ubl1: 150  $\mu$ M, Nsp3 ADRP: 75  $\mu$ M, Nsp3 SUD-M: 100  $\mu$ M, Nsp9: 17.4-25  $\mu$ M, and Nsp16: 7  $\mu$ M). Unlabelled peptide was then added at increasing concentration to compete out the binding of the labelled peptide. Results were analyzed in GrafPad Prism (version 9.4.1). From this data set, the sigmoidal dose-response fitting function was used to obtain IC50 values. These IC50 values were further converted to  $K_i$  ( $=K_D$ ) of unlabelled peptide as described by Nikolovska-Coleska et. al.<sup>82</sup>

### **Nuclear magnetic resonance spectroscopy**



Before NMR experiments the  $^{15}\text{N}$  Nsp9 sample was dialyzed into 50 mM potassium phosphate buffer pH 6,7 supplemented with and supplemented with 1mM TCEP and 10%  $\text{D}_2\text{O}$ . The final concentration of the sample was between 175-225  $\mu\text{M}$ . All NMR experiments were performed on a 600 MHz Avance Neo HD NMR spectrometer (Bruker) equipped with a 5 mm TCI cryogenic probe. All  $^1\text{H}$ - $^{15}\text{N}$  TROSY HSQC spectra were recorded at  $25^\circ\text{C}$  making use of the BEST pulse sequence with 512 points in the direct and 256 points in the indirect dimension. Two or four scans per datapoint were taken with the relaxation delay of 200  $\mu\text{s}$ . Similar experiments were performed upon addition of the Nsp9 binding peptide ligands. Final concentrations of NKRF<sub>8-23</sub>, LMTK3<sub>22-36</sub>, and NOTCH4<sub>1605-1620</sub> were 416  $\mu\text{M}$ , 230  $\mu\text{M}$  and 323  $\mu\text{M}$ , respectively. All Spectra were processed with TopSpin 3.2 and subsequently analyzed by MestReNova 14.1.0. The chemical peak shifts of Nsp9 residues were assigned based on comparison with previous assignments<sup>44,48</sup>. The perturbation of the chemical peak shifts upon addition of the peptides were calculated using equation 1:

$$\sqrt{(\Delta\delta_H)^2 + \left(\frac{\Delta\delta_N}{R_{scale}}\right)^2}$$

Where  $\Delta\delta_H$  is chemical shift change in the hydrogen chemical shift dimension and  $\Delta\delta_N$  is the chemical peak shift change in the nitrogen chemical shift dimension upon addition of peptide.  $R_{scale}$  is a scaling factor set to 6.5 as described before<sup>83</sup>.

To determine T1 and T2 relaxation times, TROSY HSQC based experiments were employed<sup>84,85</sup> (PMID: 10729271 Zhu, PMID: 22689066). The overall rotational correlation time  $\tau_c$  was estimated from the ratio of T1/T2 times (**Table S6**). The same Nsp9 - NOTCH4<sub>1605-1620</sub> sample was used as for the previous experiments. For Nsp9 T1/T2 measurements fresh sample of Nsp9 was used at the concentration of 221  $\mu\text{M}$ . The molecular weight of the species present in the sample was further estimated according to the equation 2:  $M_w \approx \frac{3.8}{2} \times \tau_c$  as described previously<sup>86-88</sup>.

### **ColabFold predictions**

ColabFold<sup>34</sup> was used to predict binding of peptides to globular protein domains from SARS-CoV-2. ColabFold is based on AlphaFold2<sup>89</sup> and AlphaFold2-multimer<sup>90</sup>. While the confidence scores varied for the predicted peptide conformations, all predictions for the structures of globular domains were in excellent agreement with solved crystal structure models (backbone alignment root mean square deviation between 0.4 and 1 Å).

### **Alanine scanning SPOT arrays**

20-mer, N-acetylated peptides on cellulose membranes were ordered from JPT (PepSpots). The membranes were activated with 5 ml methanol for 5 min at room temperature and washed with 10 ml TBST (50 mM Tris, 137 mM NaCl, 2.7 mM KCl, pH adjusted to 8.0 with HCl, 0.05% Tween-20) three times for 3 min at room temperature. The membranes were then incubated with 10 ml blocking buffer (5% skim milk powder (Merck Millipore, 115363) in TBST) for 2 hours at room temperature while rotated. The blocked membranes were incubated with concentrated GST-HA-tagged proteins of interest in blocking buffer overnight, at 4 °C, while rotated. After three quick rinses with ice-cold TBST, the membranes were incubated with HRP-conjugated anti-GST antibody (Cytiva, RPN1236) in blocking buffer for 1 hour at 4 °C, while rotated. Following three quick washes with ice-cold TBST, the chemiluminescent readout was carried out using ECL reagent (Clarity Max Western ECL substrate, 1705062, Bio-Rad) and ChemiDoc Imaging system (Bio-Rad). The acquired raw tiff images were analyzed using Image Studio Lite Ver. 5.2., and all values were normalized to the wild type results.

### **SARS-CoV-2 infection assay.**

Vero E6 (*Cercopithecus aethiops*) cells (ATCC, CRL-1586) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma) supplemented with 10% fetal bovine serum (FBS)

(HyClone) and 100 units/ml penicillin G with 100 µg/ml streptomycin solution (Gibco) at 37°C, 5% CO<sub>2</sub>, humidified chamber. MRC5, human lung fibroblast (ATCC CCL-171) cells were cultured in MEM medium (Gibco) supplemented with 10% FBS, and 100 units/ml penicillin G with 100 µg/ml streptomycin solution (Gibco) at 37 °C, 5 % CO<sub>2</sub>, humidified chamber.

SARS CoV-2 (SARS-CoV-2/01/human2020/SWE accession no/GeneBank no MT093571.1, provided by the Public Health Agency of Sweden), was grown in VeroE6 cells, and used at passage number four. Human coronavirus 229E (HCoV-229E, ATCC CCL-171) was grown and titrated in MRC5 cells and used at passage one.

VeroE6 cells were transduced using the indicated lentiviruses as previously described<sup>6</sup>. After 72 h of transduction cells were infected with SARS CoV-2 for 16 h with MOI: 0.05 at 37 °C, 5% CO<sub>2</sub>. For peptide treatments cells were first infected with SARS CoV-2 MOI: 0.05 at 37 °C, 5% CO<sub>2</sub>, after 1 h of incubation the inoculum was replaced with medium containing the indicated concentration of peptide and cells were incubated at 37 °C, 5% CO<sub>2</sub> for 16 h. After the infections, cells were fixed in 4% formaldehyde, permeabilized in 0.5% triton X-100 in PBS. Viral infected cells were revealed by staining using primary monoclonal antibodies directed against SARS CoV-2 nucleocapsid (SARS CoV-2, Sino Biological Inc., 40143-R001) or primary monoclonal mouse antibodies directed against J2 (HCoV-229E, Scicons 10010500), and secondary antibodies either donkey anti-mouse or donkey anti-rabbit IgG Alexa Fluor 555 secondary antibody (Invitrogen). Nuclei were counterstained with DAPI. Total cell number and number of infected cells/well were determined using a TROPHOS Plate RUNNER HD® (Dioscure, Marseille, France). Number of infected cells were normalized to DAPI count and presented as percentage infection of mutated control peptide for the lentivirus transduced cells or as percentage of mock treated cells in the case of peptide treatment. Results were analyzed

in GraphPad Prism (version 9.4.1) and the sigmoidal dose-response fitting function was used to obtain IC<sub>50</sub> values.

## **Acknowledgements**

This work was supported by the grants from the Swedish Foundation for Strategic research (Y.I., P.J.: SB16-0039), the Swedish Research Council (Y.I.: 2020-03380; PJ: 2020-04395; A.Ö.: 2018-05851), the Knut and Alice Wallenberg Foundation (Y.I., P.J., and A.Ö. via Science for Life Laboratory, KAW 2020.0241, V-2020-0699) and a Cancer Research UK Senior Cancer Research Fellowship (N.D.: C68484/A28159). We thank the medical faculty Umeå University strategic research resource and the Laboratory for Molecular Infection Medicine Sweden for generous support (A.Ö.), and the Biochemical Imaging Center at Umeå University and the National Microscopy Infrastructure, NMI (VR-RFI 2016-00968) for assistance in microscopy. Sequencing was performed by the SNP&SEQ Technology Platform in Stockholm. The facility is part of the National Genomic Infrastructure (NGI) Sweden and Science for Life Laboratory and is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. We used the NMR Uppsala infrastructure, which is funded by the Department of Chemistry - BMC and the Disciplinary Domain of Medicine and Pharmacy, Uppsala University.

## **Competing interests**

The authors declare no competing interests.

## **Author contributions**

FM, CB, ND, PJ and YI conceived the study. FM performed FP experiments. CB performed phage selections. EK performed SPOT arrays, RL performed viral assays, RI produced lentiviruses, HA, EA, CB and FM produced proteins, FM and CNC conducted NMR experiments and analyzed data. FM, LS, CB, ND, AKW, PJ and YI analyzed data. FM and PJ wrote the first draft. PJ and YI coordinated the study.

## **References**

- 1 Kabinger, F. *et al.* Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* **28**, 740-746 (2021). <https://doi.org/10.1038/s41594-021-00651-0>
- 2 Owen, D. R. *et al.* An oral SARS-CoV-2 M<sup>(PRO)</sup> inhibitor clinical candidate for the treatment of COVID-19. *Science* **374**, 1586-1593 (2021). <https://doi.org/10.1126/science.abl4784>
- 3 Tian, J. *et al.* Emerging viruses: Cross-species transmission of coronaviruses, filoviruses, henipaviruses, and rotaviruses from bats. *Cell Rep* **39**, 110969 (2022). <https://doi.org/10.1016/j.celrep.2022.110969>
- 4 Lasso, G. *et al.* A Structure-Informed Atlas of Human-Virus Interactions. *Cell* **178**, 1526-1541.e1516 (2019). <https://doi.org/10.1016/j.cell.2019.08.005>
- 5 Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459-468 (2020). <https://doi.org/10.1038/s41586-020-2286-9>
- 6 Kruse, T. *et al.* Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nat Commun* **12**, 6761 (2021). <https://doi.org/10.1038/s41467-021-26498-z>
- 7 Mihalic, F. *et al.* Large-scale phage-based screening reveals extensive pan-viral mimicry of host short linear motifs. *bioRxiv*, 2022.2006.2019.496705 (2022). <https://doi.org/10.1101/2022.06.19.496705>
- 8 Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem Sci* **36**, 159-169 (2011). <https://doi.org/10.1016/j.tibs.2010.10.002>
- 9 Bessa, L. M. *et al.* The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci Adv* **8**, eabm4034 (2022). <https://doi.org/10.1126/sciadv.abm4034>
- 10 Kirchdoerfer, R. N., Moyer, C. L., Abelson, D. M. & Saphire, E. O. The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis. *PLoS Pathog* **12**, e1005937 (2016). <https://doi.org/10.1371/journal.ppat.1005937>
- 11 Batra, J. *et al.* Non-canonical proline-tyrosine interactions with multiple host proteins regulate Ebola virus infection. *EMBO J* **40**, e105658 (2021). <https://doi.org/10.15252/emboj.2020105658>
- 12 Thoms, M. *et al.* Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255 (2020). <https://doi.org/10.1126/science.abc8665>
- 13 Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc Natl Acad Sci U S A* **119**, e2204539119 (2022). <https://doi.org/10.1073/pnas.2204539119>
- 14 Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M. & Bergeron, J. R. C. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Front Mol Biosci* **7**, 605236 (2020). <https://doi.org/10.3389/fmolb.2020.605236>
- 15 Arya, R. *et al.* Structural insights into SARS-CoV-2 proteins. *J Mol Biol* **433**, 166725 (2021). <https://doi.org/10.1016/j.jmb.2020.11.024>
- 16 Ricciardi, S. *et al.* The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606**, 761-768 (2022). <https://doi.org/10.1038/s41586-022-04835-6>
- 17 Oudshoorn, D. *et al.* Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *mBio* **8** (2017). <https://doi.org/10.1128/mBio.01658-17>

- 18 Lee, J. *et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat Commun* **11**, 5877 (2020). <https://doi.org/10.1038/s41467-020-19662-4>
- 19 Wang, Q. *et al.* Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **182**, 417-428.e413 (2020). <https://doi.org/10.1016/j.cell.2020.05.034>
- 20 Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782 (2020). <https://doi.org/10.1126/science.abb7498>
- 21 te Velthuis, A. J., van den Worm, S. H. & Snijder, E. J. 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 (2012). <https://doi.org/10.1093/nar/gkr893>
- 22 Park, G. J. *et al.* The mechanism of RNA capping by SARS-CoV-2. *Nature* (2022). <https://doi.org/10.1038/s41586-022-05185-z>
- 23 Newman, J. A. *et al.* Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat Commun* **12**, 4848 (2021). <https://doi.org/10.1038/s41467-021-25166-6>
- 24 Kim, Y. *et al.* Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci* **29**, 1596-1605 (2020). <https://doi.org/10.1002/pro.3873>
- 25 Benz, C. *et al.* Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol Syst Biol* **18**, e10584 (2022). <https://doi.org/10.15252/msb.202110584>
- 26 Scott, B. M., Lacasse, V., Blom, D. G., Tonner, P. D. & Blom, N. S. Predicted coronavirus Nsp5 protease cleavage sites in the human proteome. *BMC Genom Data* **23**, 25 (2022). <https://doi.org/10.1186/s12863-022-01044-y>
- 27 Chen, Z. *et al.* Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. *EMBO J* **40**, e107776 (2021). <https://doi.org/10.15252/emboj.2021107776>
- 28 Snijder, E. J. *et al.* A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS Biol* **18**, e3000715 (2020). <https://doi.org/10.1371/journal.pbio.3000715>
- 29 Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395-1398 (2020). <https://doi.org/10.1126/science.abd3629>
- 30 Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370** (2020). <https://doi.org/10.1126/science.abe9403>
- 31 Voegel, J. J. *et al.* The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *EMBO J* **17**, 507-519 (1998). <https://doi.org/10.1093/emboj/17.2.507>
- 32 Demarest, S. J. *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549-553 (2002). <https://doi.org/10.1038/415549a>
- 33 Karlsson, E., Lindberg, A., Andersson, E. & Jemth, P. High affinity between CREBBP/p300 and NCOA evolved in vertebrates. *Protein Sci* **29**, 1687-1691 (2020). <https://doi.org/10.1002/pro.3868>
- 34 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022). <https://doi.org/10.1038/s41592-022-01488-1>
- 35 Heer, C. D. *et al.* Coronavirus infection and PARP expression dysregulate the NAD metabolome: An actionable component of innate immunity. *J Biol Chem* **295**, 17986-17996 (2020). <https://doi.org/10.1074/jbc.RA120.015138>



- 36 Hoch, N. C. Host ADP-ribosylation and the SARS-CoV-2 macrodomain. *Biochem Soc Trans* **49**, 1711-1721 (2021). <https://doi.org/10.1042/BST20201212>
- 37 Alhammad, Y. M. O. *et al.* The SARS-CoV-2 Conserved Macrodomain Is a Mono-ADP-Ribosylhydrolase. *J Virol* **95** (2021). <https://doi.org/10.1128/JVI.01969-20>
- 38 Littler, D. R. *et al.* A natural product compound inhibits coronaviral replication in vitro by binding to the conserved Nsp9 SARS-CoV-2 protein. *J Biol Chem* **297**, 101362 (2021). <https://doi.org/10.1016/j.jbc.2021.101362>
- 39 El-Kamand, S. *et al.* A distinct ssDNA/RNA binding interface in the Nsp9 protein from SARS-CoV-2. *Proteins* **90**, 176-185 (2022). <https://doi.org/10.1002/prot.26205>
- 40 Yan, L. *et al.* Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **184**, 184-193.e110 (2021). <https://doi.org/10.1016/j.cell.2020.11.016>
- 41 Sutton, G. *et al.* The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* **12**, 341-353 (2004). <https://doi.org/10.1016/j.str.2004.01.016>
- 42 Miknis, Z. J. *et al.* Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol* **83**, 3007-3018 (2009). <https://doi.org/10.1128/JVI.01505-08>
- 43 Littler, D. R., Gully, B. S., Colson, R. N. & Rossjohn, J. Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. *iScience* **23**, 101258 (2020). <https://doi.org/10.1016/j.isci.2020.101258>
- 44 F Dudás, E. *et al.* Backbone chemical shift spectral assignments of SARS coronavirus-2 non-structural protein nsp9. *Biomol NMR Assign* **15**, 235-241 (2021). <https://doi.org/10.1007/s12104-021-10011-0>
- 45 Frieman, M. *et al.* Molecular determinants of severe acute respiratory syndrome coronavirus pathogenesis and virulence in young and aged mouse models of human disease. *J Virol* **86**, 884-897 (2012). <https://doi.org/10.1128/JVI.05957-11>
- 46 Bujanic, L. *et al.* The key features of SARS-CoV-2 leader and NSP1 required for viral escape of NSP1-mediated repression. *RNA* **28**, 766-779 (2022). <https://doi.org/10.1261/rna.079086.121>
- 47 Li, J. *et al.* Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis. *Med (N Y)* **2**, 99-112.e117 (2021). <https://doi.org/10.1016/j.medj.2020.07.002>
- 48 Buchko, G. W., Zhou, M., Craig, J. K., Van Voorhis, W. C. & Myler, P. J. Backbone chemical shift assignments for the SARS-CoV-2 non-structural protein Nsp9: intermediate (ms -  $\mu$ s) dynamics in the C-terminal helix at the dimer interface. *Biomol NMR Assign* **15**, 107-116 (2021). <https://doi.org/10.1007/s12104-020-09992-1>
- 49 Chen, Y. & Guo, D. Molecular mechanisms of coronavirus RNA capping and methylation. *Virol Sin* **31**, 3-11 (2016). <https://doi.org/10.1007/s12250-016-3726-4>
- 50 Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511-7526 (2016). <https://doi.org/10.1093/nar/gkw551>
- 51 Rosas-Lemus, M. *et al.* High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for structure-based inhibitor design. *Sci Signal* **13** (2020). <https://doi.org/10.1126/scisignal.abe1202>
- 52 Krafcikova, P., Silhan, J., Nencka, R. & Boura, E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun* **11**, 3717 (2020). <https://doi.org/10.1038/s41467-020-17495-9>
- 53 Ke, M. *et al.* Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2'-O-methyltransferase activity of



- nsp10/nsp16 complex. *Virus Res* **167**, 322-328 (2012). <https://doi.org/10.1016/j.virusres.2012.05.017>
- 54 Wang, Y. *et al.* Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis. *J Virol* **89**, 8416-8427 (2015). <https://doi.org/10.1128/JVI.00948-15>
- 55 Xu, W. *et al.* Ebola virus VP30 and nucleoprotein interactions modulate viral RNA synthesis. *Nat Commun* **8**, 15576 (2017). <https://doi.org/10.1038/ncomms15576>
- 56 Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *ACS Infect Dis* **6**, 3174-3189 (2020). <https://doi.org/10.1021/acsinfecdis.0c00500>
- 57 Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**, 246-252 (2021). <https://doi.org/10.1038/s41586-021-03493-4>
- 58 Schmidt, N. *et al.* The SARS-CoV-2 RNA-protein interactome in infected human cells. *Nat Microbiol* **6**, 339-353 (2021). <https://doi.org/10.1038/s41564-020-00846-z>
- 59 Tian, L. *et al.* RNA-dependent RNA polymerase (RdRp) inhibitors: The current landscape and repurposing for the COVID-19 pandemic. *Eur J Med Chem* **213**, 113201 (2021). <https://doi.org/10.1016/j.ejmech.2021.113201>
- 60 Min, J. S., Kwon, S. & Jin, Y. H. SARS-CoV-2 RdRp Inhibitors Selected from a Cell-Based SARS-CoV-2 RdRp Activity Assay System. *Biomedicines* **9** (2021). <https://doi.org/10.3390/biomedicines9080996>
- 61 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504 (2020). <https://doi.org/10.1126/science.abc1560>
- 62 Yin, W. *et al.* Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. *Nat Struct Mol Biol* **28**, 319-325 (2021). <https://doi.org/10.1038/s41594-021-00570-0>
- 63 Jin, Z. *et al.* Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293 (2020). <https://doi.org/10.1038/s41586-020-2223-y>
- 64 Günther, S. *et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **372**, 642-646 (2021). <https://doi.org/10.1126/science.abf7945>
- 65 Lutgens, A. *et al.* Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc* **144**, 2905-2920 (2022). <https://doi.org/10.1021/jacs.1c08402>
- 66 Hussain, A. *et al.* Targeting SARS-CoV2 Spike Protein Receptor Binding Domain by Therapeutic Antibodies. *Biomed Pharmacother* **130**, 110559 (2020). <https://doi.org/10.1016/j.biopha.2020.110559>
- 67 Cao, Y. *et al.* Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 73-84.e16 (2020). <https://doi.org/10.1016/j.cell.2020.05.025>
- 68 Corti, D., Purcell, L. A., Snell, G. & Veisler, D. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* **184**, 3086-3108 (2021). <https://doi.org/10.1016/j.cell.2021.05.005>
- 69 Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650-655 (2020). <https://doi.org/10.1126/science.abc6952>

- 70 Ghosh, A. K., Mishevich, J. L., Mesecar, A. & Mitsuya, H. Recent Drug Development and Medicinal Chemistry Approaches for the Treatment of SARS-CoV-2 and Covid-19. *ChemMedChem* (2022). [https://doi.org:10.1002/cmdc.202200440](https://doi.org/10.1002/cmdc.202200440)
- 71 Gorgulla, C. *et al.* A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening. *iScience* **24**, 102021 (2021). [https://doi.org:10.1016/j.isci.2020.102021](https://doi.org/10.1016/j.isci.2020.102021)
- 72 Hu, T. *et al.* Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9. *Protein Sci* **26**, 1037-1048 (2017). [https://doi.org:10.1002/pro.3150](https://doi.org/10.1002/pro.3150)
- 73 Decroly, E. *et al.* Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog* **7**, e1002059 (2011). [https://doi.org:10.1371/journal.ppat.1002059](https://doi.org/10.1371/journal.ppat.1002059)
- 74 Nencka, R. *et al.* Coronaviral RNA-methyltransferases: function, structure and inhibition. *Nucleic Acids Res* **50**, 635-650 (2022). [https://doi.org:10.1093/nar/gkab1279](https://doi.org/10.1093/nar/gkab1279)
- 75 Ali, M., Simonetti, L. & Ivarsson, Y. Screening Intrinsically Disordered Regions for Short Linear Binding Motifs. *Methods Mol Biol* **2141**, 529-552 (2020). [https://doi.org:10.1007/978-1-0716-0524-0\\_27](https://doi.org/10.1007/978-1-0716-0524-0_27)
- 76 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003). [https://doi.org:10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- 77 Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res* **47**, D529-D541 (2019). [https://doi.org:10.1093/nar/gky1079](https://doi.org/10.1093/nar/gky1079)
- 78 Guirimand, T., Delmotte, S. & Navratil, V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* **43**, D583-587 (2015). [https://doi.org:10.1093/nar/gku1121](https://doi.org/10.1093/nar/gku1121)
- 79 Orchard, S. *et al.* Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* **9**, 345-350 (2012). [https://doi.org:10.1038/nmeth.1931](https://doi.org/10.1038/nmeth.1931)
- 80 Orchard, S. *et al.* The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358-363 (2014). [https://doi.org:10.1093/nar/gkt1115](https://doi.org/10.1093/nar/gkt1115)
- 81 Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* **40**, D857-861 (2012). [https://doi.org:10.1093/nar/gkr930](https://doi.org/10.1093/nar/gkr930)
- 82 Nikolovska-Coleska, Z. *et al.* Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **332**, 261-273 (2004). [https://doi.org:10.1016/j.ab.2004.05.055](https://doi.org/10.1016/j.ab.2004.05.055)
- 83 Mulder, F. A., Schipper, D., Bott, R. & Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilis*ins. *J Mol Biol* **292**, 111-123 (1999). [https://doi.org:10.1006/jmbi.1999.3034](https://doi.org/10.1006/jmbi.1999.3034)
- 84 Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* **143**, 423-426 (2000). [https://doi.org:10.1006/jmre.2000.2022](https://doi.org/10.1006/jmre.2000.2022)
- 85 Lakomek, N. A., Ying, J. & Bax, A. Measurement of <sup>15</sup>N relaxation rates in perdeuterated proteins by TROSY-based methods. *J Biomol NMR* **53**, 209-221 (2012). [https://doi.org:10.1007/s10858-012-9626-5](https://doi.org/10.1007/s10858-012-9626-5)
- 86 García de la Torre, J., Huertas, M. L. & Carrasco, B. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* **147**, 138-146 (2000). [https://doi.org:10.1006/jmre.2000.2170](https://doi.org/10.1006/jmre.2000.2170)

- 87 Aragon, S. & Hahn, D. K. Precise boundary element computation of protein transport properties: Diffusion tensors, specific volume, and hydration. *Biophys J* **91**, 1591-1603 (2006). <https://doi.org/10.1529/biophysj.105.078188>
- 88 Bae, S. H., Dyson, H. J. & Wright, P. E. Prediction of the rotational tumbling time for proteins with disordered segments. *J Am Chem Soc* **131**, 6814-6821 (2009). <https://doi.org/10.1021/ja809687r>
- 89 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 90 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034 (2021). <https://doi.org/10.1101/2021.10.04.463034>

## **Supplementary information**

Supplementary figures

**Figure S1. All affinity measurements performed in this study.**

**Figure S2. Alignments of peptides**

**Figure S3. ColabFold for Nsp3 globular domains**

**Figure S4. SPOT array for Nsp3 SUD-M with PRDM14 peptide**

**Figure S5. ColabFold for Nsp9**

**Figure S6. Raw HSQC spectra and Nsp9-peptide model .**

**Figure S7. ColabFold for Nsp16**

**Figure S8. Cell penetrating peptides fail to inhibit HCoV 229E proliferation**

Supplementary tables; see separate files

**Table S1. Overview of all SARS-CoV-2 protein domain constructs used in this study.**

**Table S2. ProP-PD selection results**

**Table S3. GO term analysis**

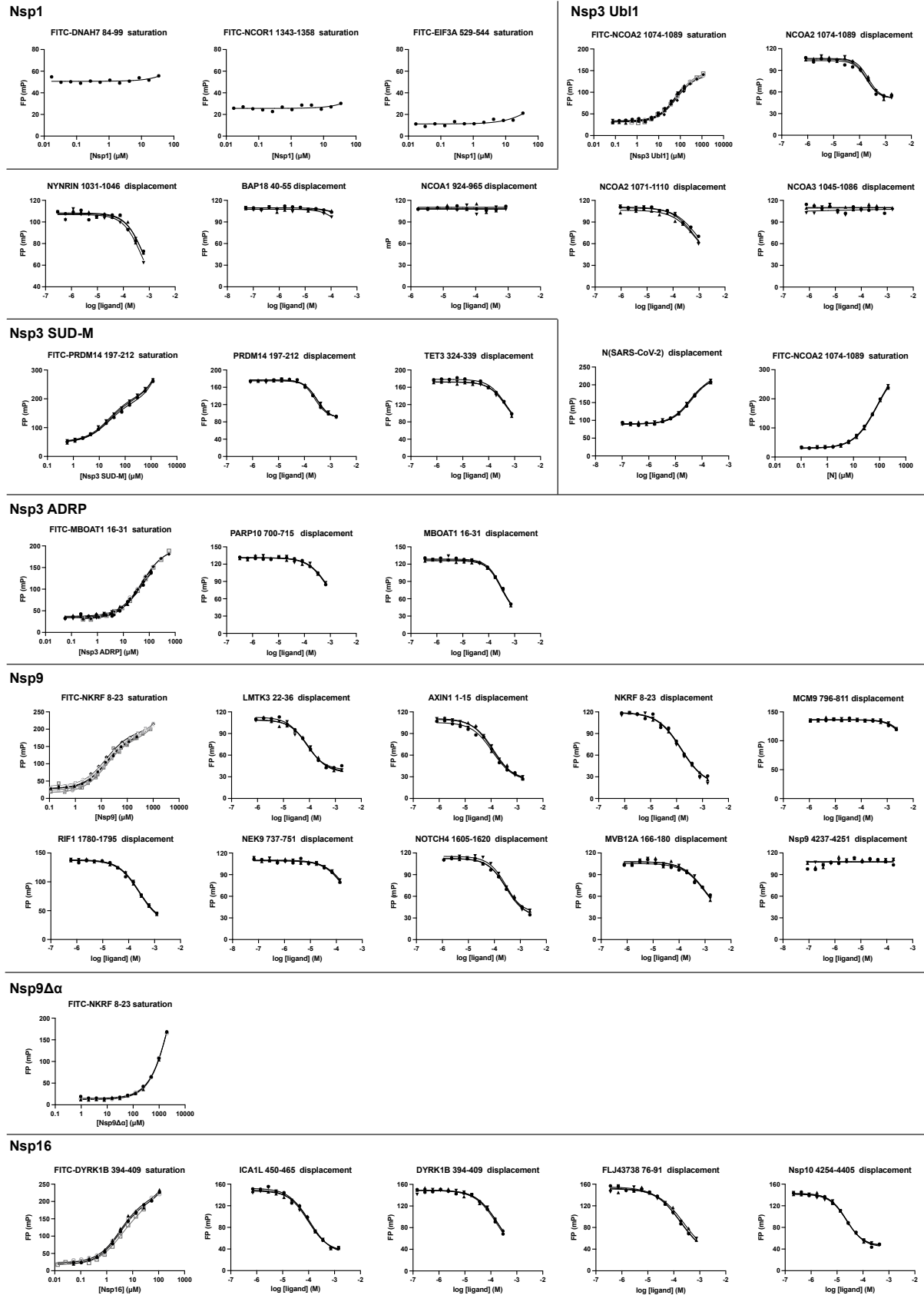
**Table S4. Overview of the peptides and affinity measurements performed in present work.**

**Table S5. NMR peak shift perturbation calculation**

**Table S6. NMR  $T_1$ ,  $T_2$  calculation**

**Table S7. Lentiviral constructs used in this study**

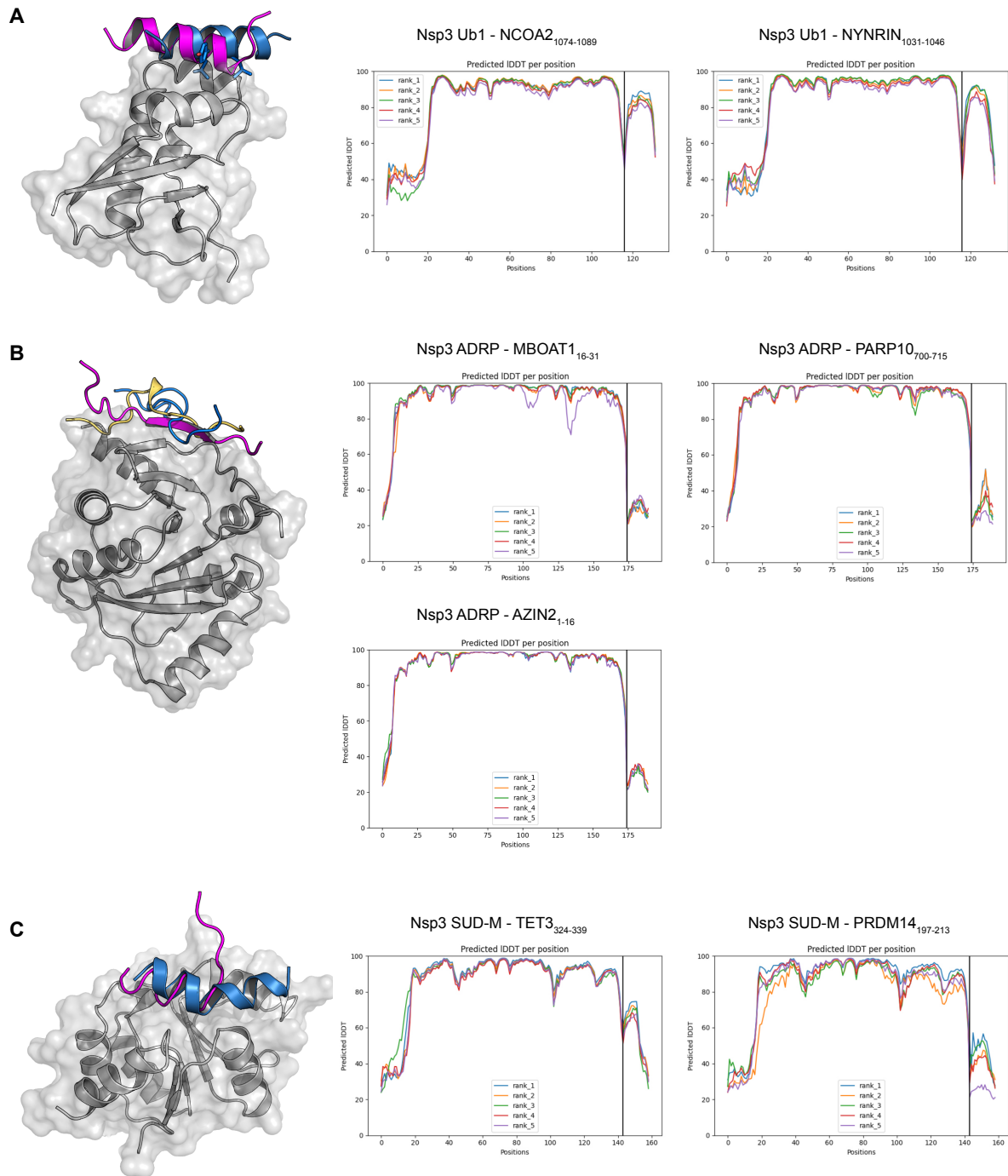
**Table S8. Cell penetrating peptides**



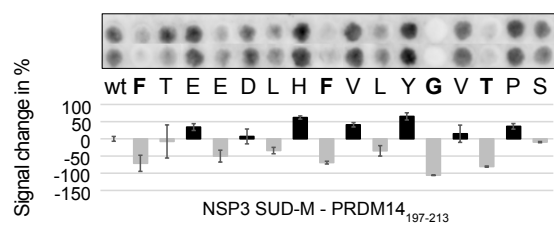
**Figure S1. All affinity measurements performed in this study. Peptide sequences and calculated  $K_D$  values are presented in Table S4.**

<b>A</b>			<b>B</b>		
	Peptide sequence	Protein		Peptide sequence	Protein
<b>Nsp3 Ubl1</b>	631-QEDE <b>W</b> WNVQY <b>P</b> DQP <b>E</b> E-646	PLEKHM1	<b>Nsp3 ADRP</b>	18-GSTYL <b>HPLSE</b> LLGIPL-33	MBOAT1
	973-NTFCTY <b>LLEQ</b> ID <b>ML</b> FF-988	PCNX2		1-MAG <b>YLSE</b> SDFVMVEEG-16	AZIN2
	1031-CPSLS <b>EEILR</b> CLSLHD-1046	NYNRIN		700-DGGTDGKAQLVVHSAF-715	PARP10
	286-NDDL <b>VQRLMD</b> ILY <b>A</b> SE-301	MAPRE2			
	1074-SDEG <b>ALLDQL</b> YLALRN-1089	NCOA2			
	1145-WSQVDEATIR <b>AI</b> AVL-1160	YTHDC2			
	40-AKWTE <b>TEI</b> MLRAAVK-55	BAP18			
<b>C</b>					
	Peptide sequence	Protein			
<b>Nsp16</b>	452-DMSA <b>W</b> F <b>NLF</b> ADLDPLS-467	ICA1L			
	447-KAASDLT <b>AWFSL</b> FADL-462	ICA1			
	394-PGHSPADYLR <b>FQDL</b> LVL-409	DYRK1B			
	78-PLDSYLN <b>FQAL</b> ISPRE-93	FLJ43738			
	324-TDPLIRWDSYENLSAD-339	TNS3			
	228-AHHTG <b>TAS</b> SEITPAQ-243	SCAPER			

**Figure S2. Alignment of the peptides identified by ProP-PD screens against.** (A) Peptides identified by ProP-PD screens against Nsp3 Ubl1. Hydrophobic residues roughly corresponding to  $\Phi$ xx $\Phi$ x $\Phi$  are in bold. (B) Peptides identified by ProP-PD screens against Nsp3 ADRP. The common LSE motif is in bold. (C) Peptides identified by ProP-PD screens against Nsp16. The [W/F]xxx $\Phi$  motif is in bold.

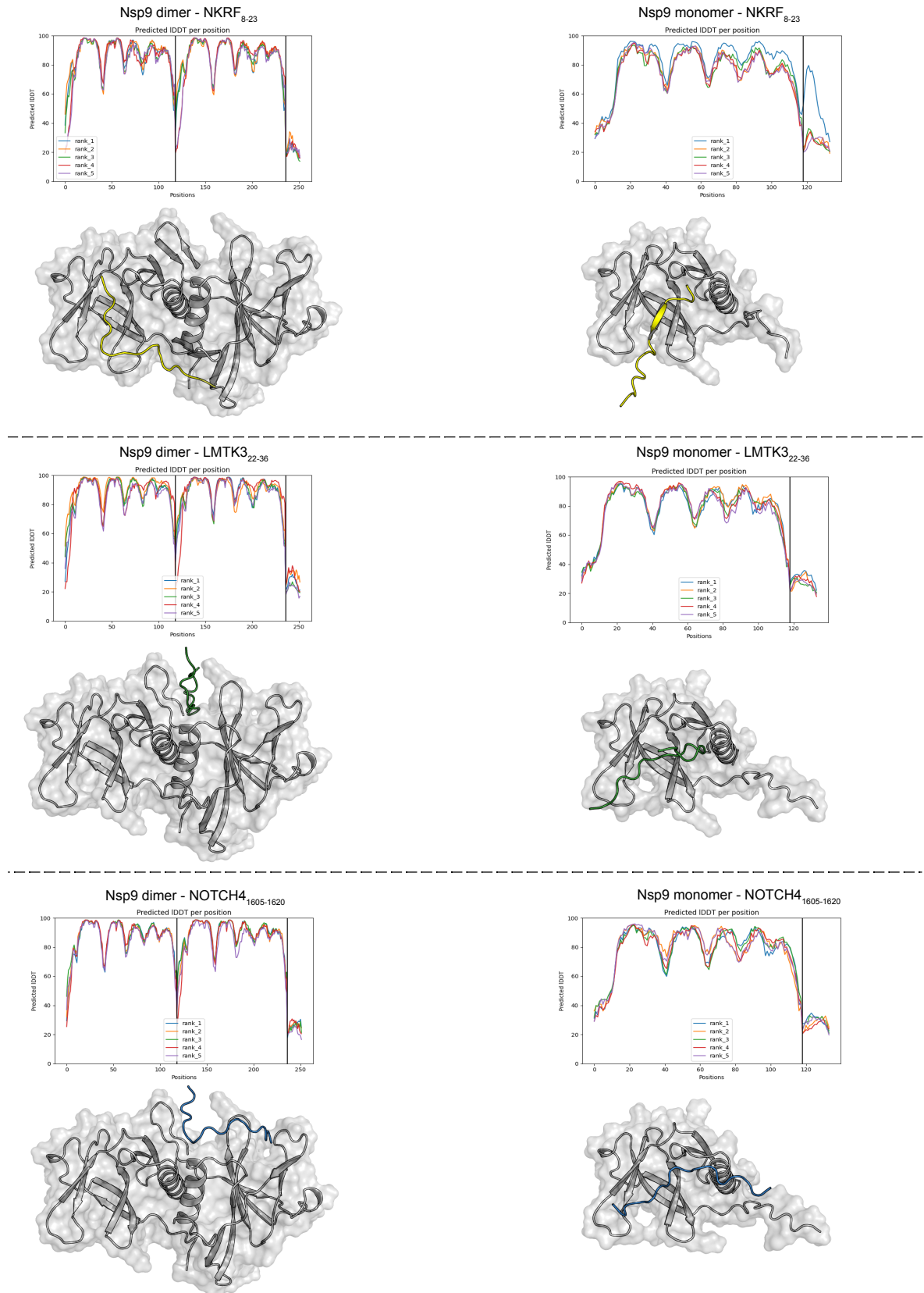


**Figure S3. Models and confidence scores from ColabFold predictions for binding between Nsp3 globular domains and peptides identified in ProP-PD selections. (A) Model and confidence scores for prediction of Nsp3 Ub1 in complex with peptides. The NCOA2<sub>1074-1089</sub> peptide is in blue, and the NYNRIN<sub>1031-1046</sub> peptide is purple. (B) Model and confidence scores for prediction of Nsp3 ADRP in complex with peptides. The MBOAT1<sub>16-31</sub> peptide is in blue, the PARP10<sub>700-715</sub> peptide is in purple, and the AZIN<sub>1-16</sub> peptide is yellow. (C) Model and confidence scores for prediction of Nsp3 SUD-M in complex with peptides. The TET<sub>324-339</sub> peptide is in blue, and the PRDM14<sub>197-213</sub> peptide is purple.**



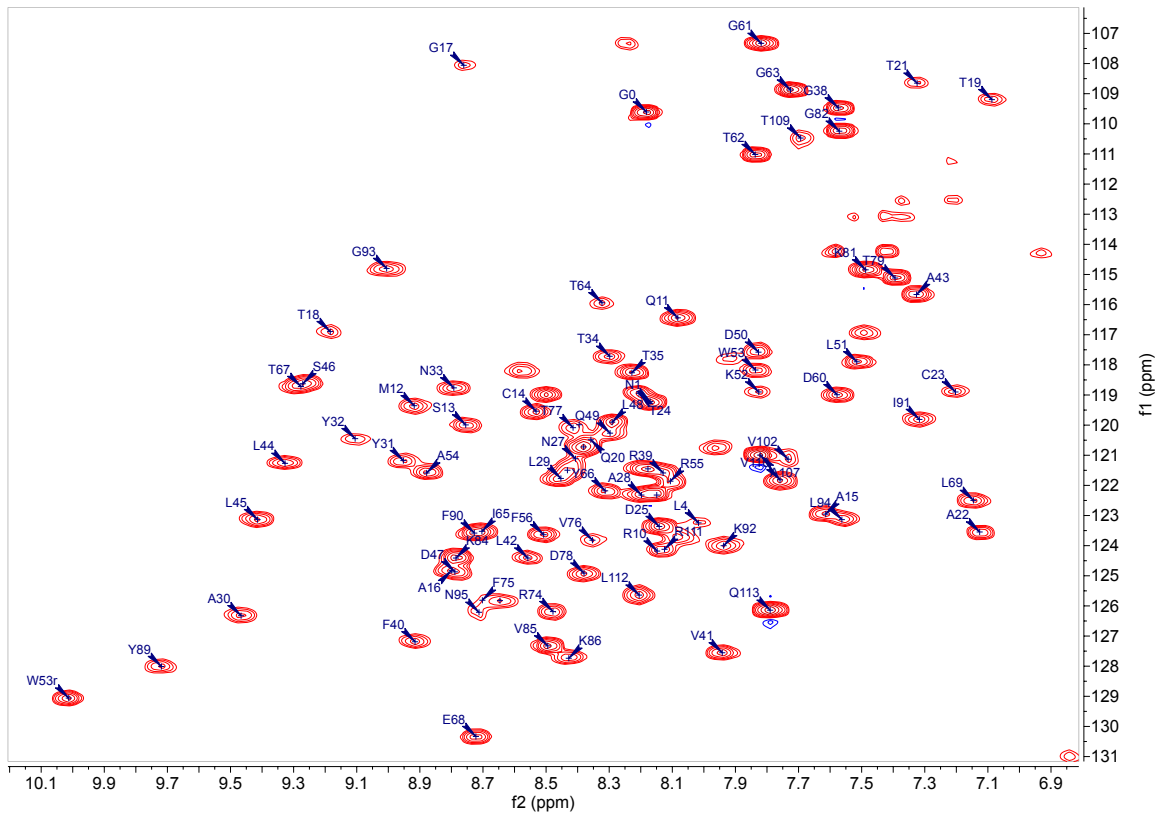
**Figure S4. SPOT array alanine scans for the PRDM14<sup>197-213</sup> peptide.** The peptide was challenged with Nsp3 SUD-M. The residues where clear reduction of binding is observed are in bold.



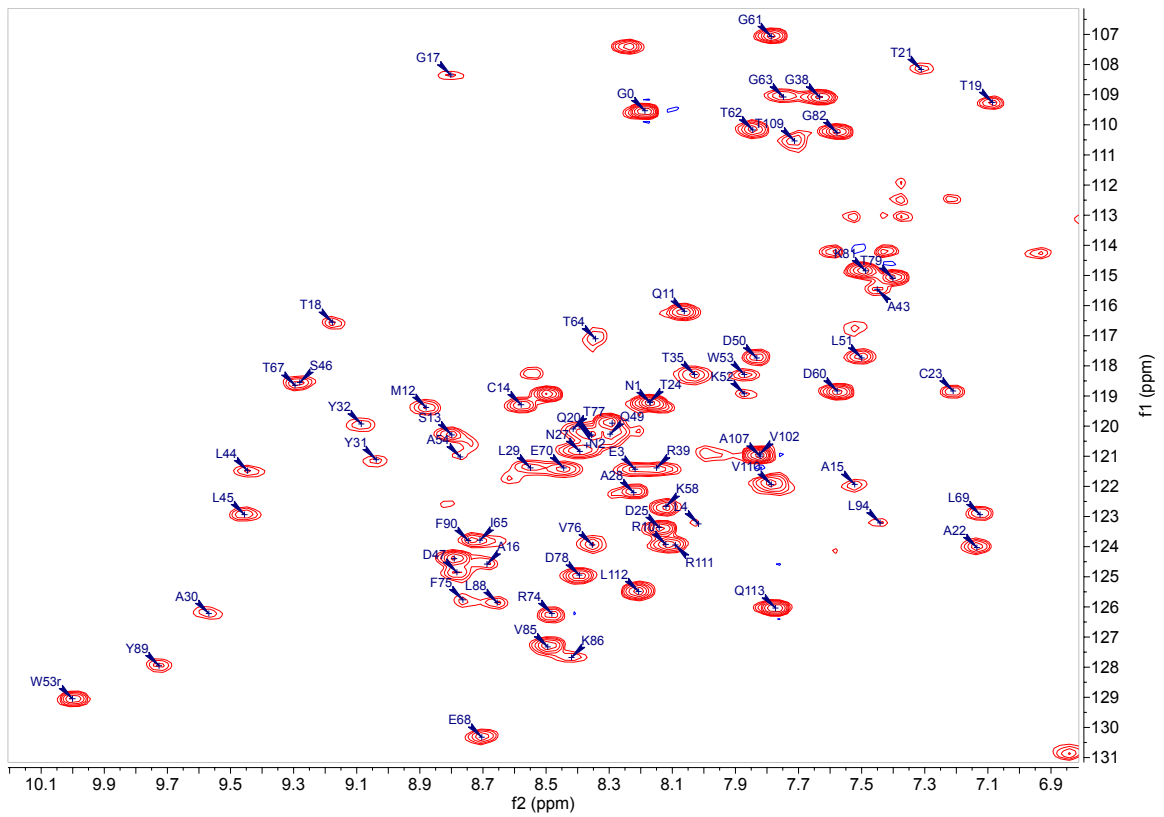


**Figure S5. Models and confidence scores of ColabFold predictions for interaction between Nsp9 with NKRF<sub>8-23</sub>, LMTK3<sub>22-36</sub>, and NOTCH4<sub>1605-1620</sub> peptides.** For all three peptides the prediction was run with either monomeric or dimeric Nsp9. The predicted peptides did not converge with either Nsp9 monomer or dimer model.

A

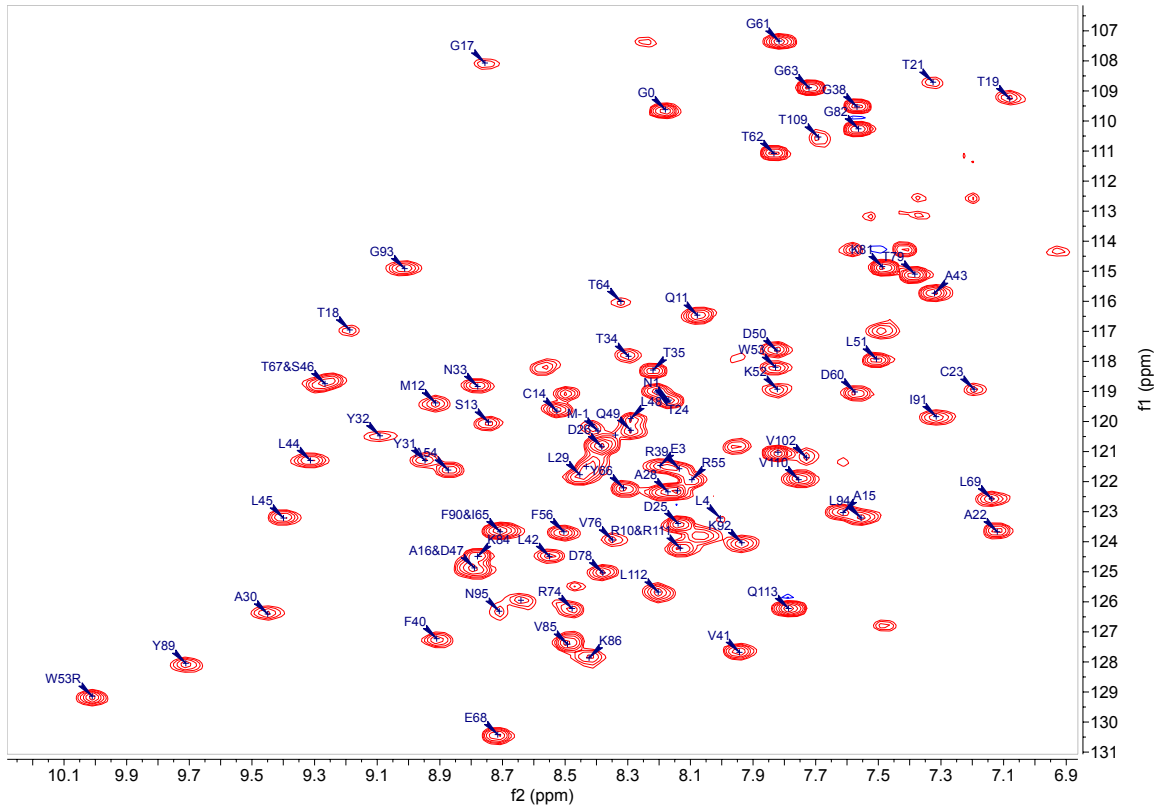


B

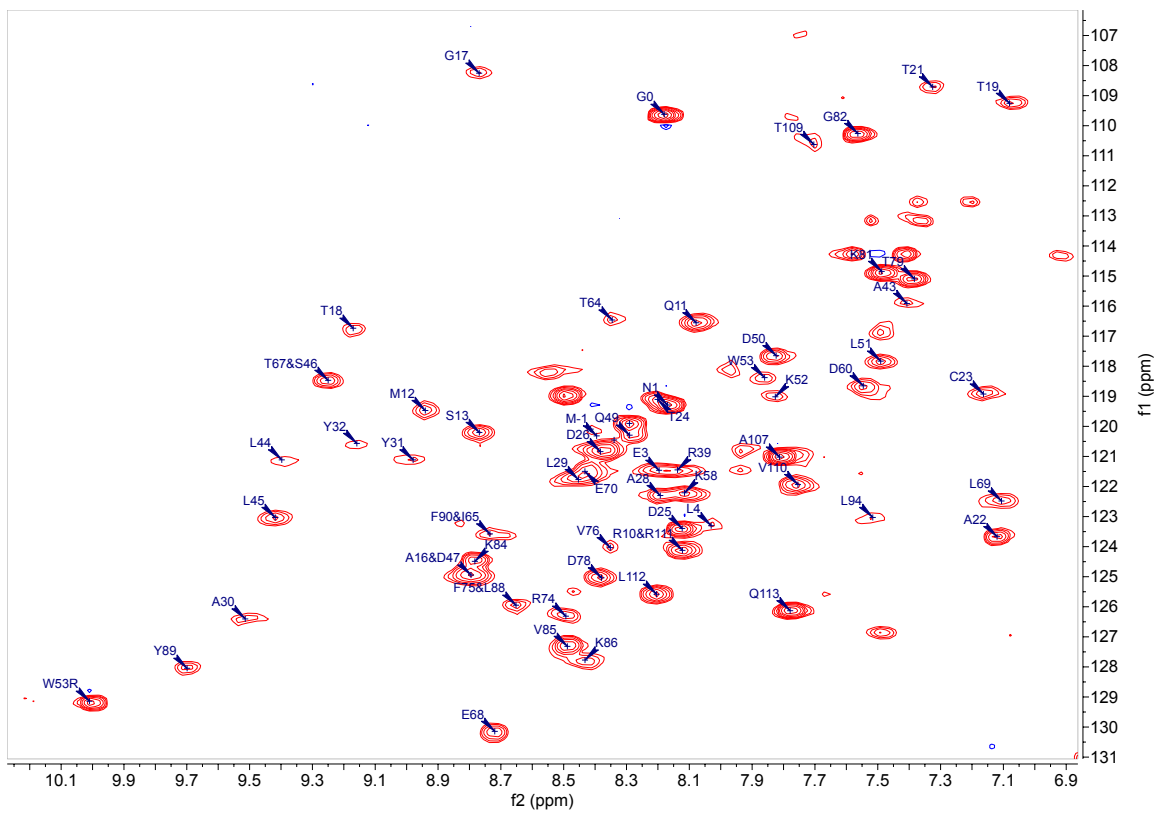


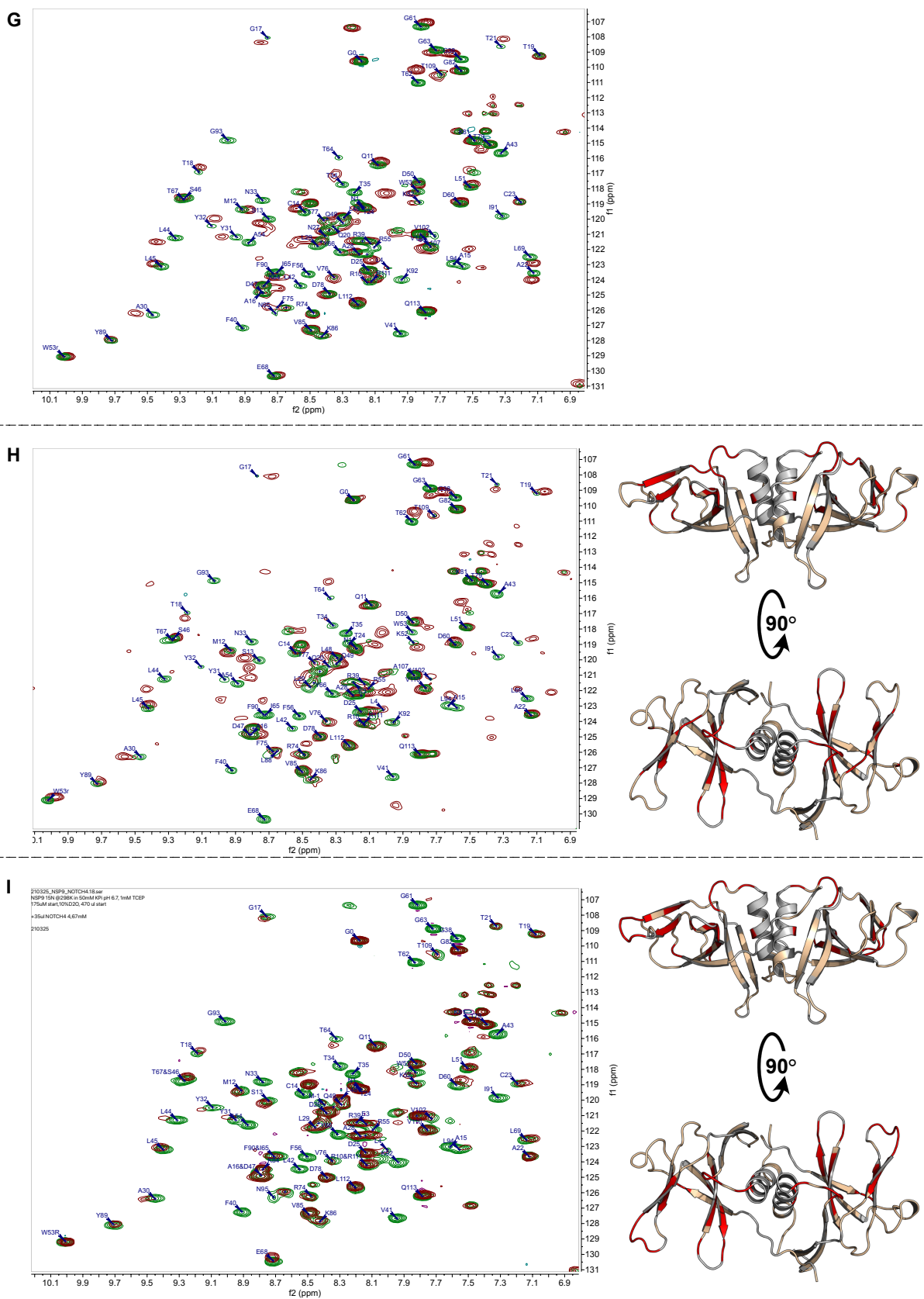


E



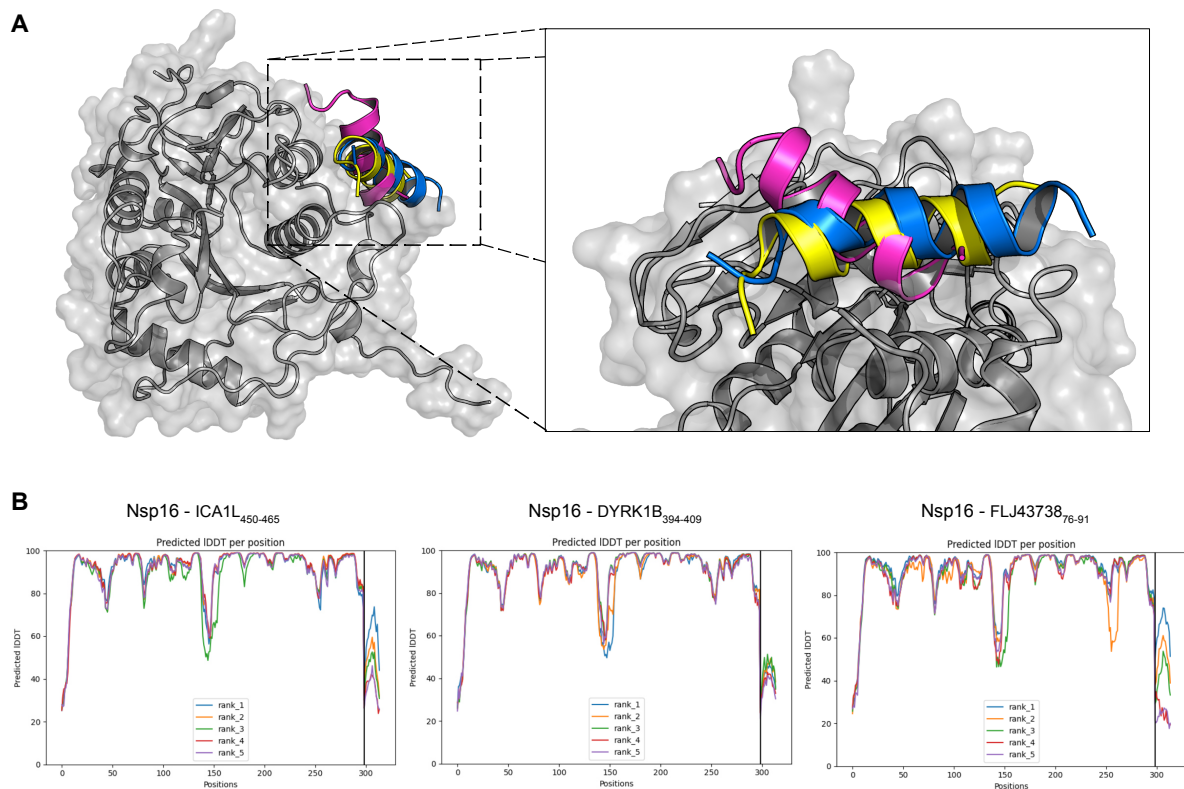
F



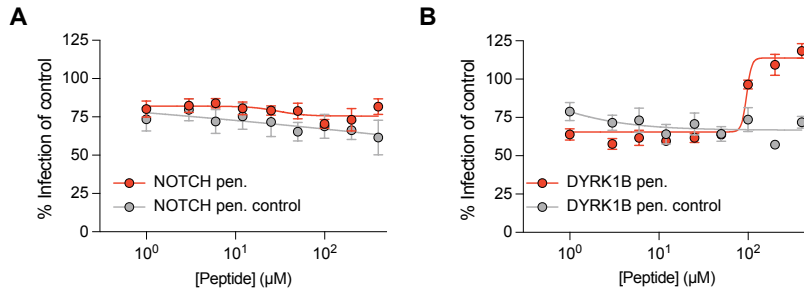


**Figure S6. Raw HSQC spectra and Nsp9 model with highlighted peak shift changes for NKRF<sub>8-23</sub>, LMTK3<sub>22-36</sub> and NOTCH4<sub>1605-1620</sub> peptides. (A) Chemical shift annotation for Nsp9 without added peptide for the**

experiment where NKRF<sub>8-23</sub> was added to the Nsp9. **(B)** Chemical shift annotation for Nsp9 with NKRF<sub>8-23</sub> peptide. **(C)** Chemical shift annotation for Nsp9 without peptide for the experiment where LMTK<sub>322-36</sub> was added to the Nsp9. **(D)** Chemical shift annotation for Nsp9 with LMTK<sub>322-36</sub> peptide. **(E)** Chemical shift annotation for Nsp9 without peptide for the experiment where NOTCH<sub>41605-1620</sub> was added to the Nsp9. **(F)** Chemical shift annotation for Nsp9 with NOTCH<sub>41605-1620</sub> peptide. **(G)** Overlay of spectra from **(A)** and **(B)**. **(H)** Overlay of spectra from **(C)** and **(D)**. On the right side the residues of Nsp9 with a chemical shift change larger than one standard deviation upon addition of LMTK<sub>322-36</sub> are highlighted in red, the residues whose shift is below that threshold are in beige and the unassigned residues in gray. **(I)** Overlay of spectra from **(E)** and **(F)**. On the right side, residues of Nsp9 with a chemical shift change larger than one standard deviation upon addition of NOTCH<sub>41605-1620</sub> are highlighted in red, the residues whose shift is below that threshold are in beige and the unassigned residues in gray.



**Figure S7. Models and confidence scores of ColabFold predictions for the interaction between Nsp16 and DYRK1B<sub>395-410</sub>, ICA1L<sub>450-465</sub>, and FLJ43738<sub>76-91</sub> peptides. (A) Superimposed best predicted model for the interaction. The ICA1L<sub>450-465</sub> peptide is in blue the DYRK1B<sub>395-410</sub> peptide in purple and the FLJ43738<sub>76-91</sub> peptide is in yellow. (B) Confidence scores for the predictions.**



**Figure S8. Cell penetrating peptides fail to inhibit HCoV 229E proliferation.** (A) The MRC5 cells were infected with HCoV 229E and subsequently treated with either NOTCH4 cell penetrating peptide (red dots) or the NOTCH4 control cell penetrating peptide (gray dots). (B) The same experiments were performed with the exception that the cells were treated with either DYRK1B cell penetrating peptide (red dots) or DYRK1B control cell penetrating peptide (gray dots).