ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Revealing the Structural Plasticity of SARS-CoV-2 nsp7 and nsp8 Using Structural Proteomics

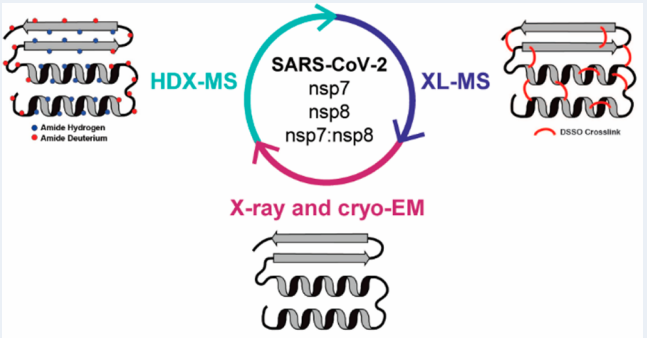
Valentine V. Courouble,[⊥] Sanjay Kumar Dey,[⊥] Ruchi Yadav,[⊥] Jennifer Timm, Jerry Joe E. K. Harrison, Francesc X. Ruiz, Eddy Arnold,* and Patrick R. Griffin*

Cite This: *J. Am. Soc. Mass Spectrom.* 2021, 32, 1618−1630

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Coronavirus (CoV) nonstructural proteins (nsps) assemble to form the replication−transcription complex (RTC) responsible for viral RNA synthesis. nsp7 and nsp8 are important cofactors of the RTC, as they interact and regulate the activity of RNA-dependent RNA polymerase and other nsps. To date, no structure of the full-length SARS-CoV-2 nsp7:nsp8 complex has been published. The current understanding of this complex is based on structures from truncated constructs, with missing electron densities, or from related CoV species where SARS-CoV-2 nsp7 and nsp8 share upward of 90% sequence identity. Despite available structures solved using crystallography and cryo-EM representing detailed static snapshots of the nsp7:nsp8 complex, it is evident that the complex has a high degree of structural plasticity. However, relatively little is known about the conformational dynamics of the individual proteins and how they complex to interact with other nsps. Here, the solution-based structural proteomic techniques, hydrogen−deuterium exchange mass spectrometry (HDX-MS) and cross-linking mass spectrometry (XL-MS), illuminate the dynamics of SARS-CoV-2 full-length nsp7 and nsp8 proteins and the nsp7:nsp8 protein complex. Results presented from the two techniques are complementary and validate the interaction surfaces identified from the published three-dimensional heterotetrameric crystal structure of the SARS-CoV-2 truncated nsp7:nsp8 complex. Furthermore, mapping of XL-MS data onto higher-order complexes suggests that SARS-CoV-2 nsp7 and nsp8 do not assemble into a hexadecameric structure as implied by the SARS-CoV full-length nsp7:nsp8 crystal structure. Instead, our results suggest that the nsp7:nsp8 heterotetramer can dissociate into a stable dimeric unit that might bind to nsp12 in the RTC without significantly altering nsp7−nsp8 interactions.



INTRODUCTION

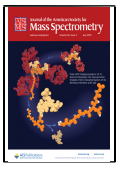
The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; CoV-2) is the agent responsible for the coronavirus disease 2019 (COVID-19) that has infected over 150 million people and claimed over 3.3 million lives worldwide.[1,2] As part of the betacoronavirus genus, CoV-2 is an enveloped positive single-stranded RNA virus.[3] The large ∼30 kb genome encodes six functional open reading frames (ORFs): replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N), ordered from the 5′ to 3′ end. An additional seven putative ORFs encoding accessory proteins are interspersed between the structural proteins.[4]

Replication of the very large genome is mediated by a highly dynamic protein−RNA complex known as the replication−transcription complex (RTC).[5] The nonstructural proteins (nsps) that make up this complex are encoded by the replicase genes. ORF1a and ORF1b, which take up approximately two-thirds of the viral genome, are translated to produce two large polyproteins, pp1a and pp1ab, which then undergo proteolytic processing to release 16 individual nsps. This process is mediated by two self-encoded proteases: the papain-like protease (PLpro; nsp3) responsible for cleaving nsp1−nsp4 and the main chymotrypsin-like protease (Mpro; 3CLpro; nsp5) responsible for cleaving nsp4−nsp16.

Over the past year, many studies have focused on solving the structure of the RNA−polymerase complex to aid in the development of antiviral drugs to treat CoV-2 infections.[6−12] These structures suggest that the minimal RNA−polymerase complex is composed of the RNA-dependent RNA−polymerase (RdRp) catalytic domain encoded in the C-terminus of nsp12 bound to two cofactor proteins nsp7 and nsp8. Cryo-electron microscopy (cryo-EM) structures have suggested the stoichiometry of the nsp12:nsp7:nsp8 complex to be 1:1:2,

Special Issue: Focus: Protein Footprinting

Received: March 8, 2021
Revised: May 24, 2021
Accepted: May 28, 2021
Published: June 14, 2021



respectively. *In vitro* studies have shown that the polymerase activity of nsp12 requires binding of at least nsp7 and nsp8.[7,9] Structural and biochemical analyses of the nsp7 and nsp8 complex in the absence of nsp12 have been less studied.

There are currently no available crystal structures of CoV-2 nsp7 or nsp8 in isolation and seven available crystal structures for the CoV-2 nsp7:nsp8 complex in the Protein Data Bank (PDB).[13−19] Only one of the seven structures has been peer-reviewed and published.[13] Overall, the seven structures can be classified into three representative conformations: (A) dimer (PDB: 6WIQ and 6M5I), (B) linear heterotetramer (PDB: 6YHU, 6WQD, and 7JLT), and (C) cubic heterotetramer (PDB: 6WTC and 6XIP). A comparison of the heterotetrameric structures reveals that the linear heterotetramer structures are arranged around an nsp7 dimer core while the cubic heterotetramer structures are composed of two adjacent nsp7:nsp8 dimers. Interestingly, X-ray crystal structures of the full-length nsp7:nsp8 complex from SARS-CoV and from feline coronavirus reveal alternative conformations. The SARS-CoV complex forms a hexadecameric structure primarily organized around N-terminal nsp8 interactions.[20] On the contrary, the feline coronavirus structure revealed an nsp7:nsp8 heterotrimer with one nsp8 molecule associated with two nsp7 molecules that self-interact.[21] Furthermore, a study of the SARS-CoV nsp7:nsp8 complex using native mass spectrometry (MS) reported a heterotetrameric structure with no nsp7 self-interaction.[22] Nevertheless, to date, the structures of neither nsp7 and nsp8 as individual proteins nor the intact full-length nsp7:nsp8 complex from CoV-2 have been reported. To address this lack of knowledge, we applied a combination of hydrogen−deuterium exchange MS (HDX-MS) and cross-linking MS (XL-MS) to probe the in-solution dynamics of the intact full-length nsp7, nsp8, and nsp7:nsp8 complex from CoV-2. Our studies demonstrate that the nsp7:nsp8 complex contains a high degree of architectural plasticity that is not apparent from X-ray crystallographic or cryo-EM studies. Results from the structural proteomic studies are used to propose a new model of assembly to expand our understanding of the nsp7 and nsp8 biological function.

EXPERIMENTAL SECTION

Reagents and Plasmids. Unless otherwise specified, all chemical and reagents were purchased from Sigma-Aldrich (St. Louis, MO). Formic acid, trifluoroacetic acid, and UHPLC-grade solvents were purchased from ThermoFisher. pGBW-m4046979 (coding for full-length nsp7, NCBI Reference Sequence: YP_009725303.1, codon optimized, with an initial Met and a cleavable C-terminal TEV His_6 tag) was a gift from Ginkgo Bioworks (Addgene plasmid 145611; http://n2t.net/addgene:145611; RRID: Addgene_145611). pGBW-m4046852 (coding for full-length nsp8, NCBI Reference Sequence: YP_009725304.1, codon optimized, with an initial Met and a cleavable C-terminal TEV His_6 tag) was a gift from Ginkgo Bioworks (Addgene plasmid 145584; http://n2t.net/addgene:145584; RRID: Addgene_145584).

Protein Expression and Purification. nsp7 was expressed in T7 express competent cells (previously transformed with the aforementioned plasmid) at 37 °C, and then, when the OD_{600} was ∼0.6, the culture was brought to 20 °C for 1 h, followed by overnight induction with 1 mM IPTG at 20 °C. Cells were pelleted by centrifugation at 7200*g* and, after a freeze−thaw cycle, were resuspended with lysis buffer containing 50 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 10 mM CHAPS, 5% glycerol, 1 mM TCEP, 1 μM leupeptin, 1 μM pepstatin A, and 1 mM PMSF and sonicated and centrifuged at 30 000*g* for 60 min. Ni-NTA affinity purification was performed, and after an extensive wash with lysis buffer, an imidazole-based step elution was done, followed by overnight digestion with TEV protease and a reverse Ni-NTA affinity purification to recover the cleaved protein. Subsequent heparin affinity and Mono Q anion exchange steps further improved the purity of nsp7. nsp8 was similarly expressed, with an additional size exclusion chromatography step that was done after the heparin affinity and Mono Q steps. The final buffer for both proteins was 50 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol, and 1 mM TCEP.

Cross-Linking Mass Spectrometry (XL-MS). *Sample Preparation.* For DSSO (disuccinimidyl sulfoxide) (ThermoFisher) cross-linking reactions, individual protein and protein−protein complexes were diluted to 10 μM in cross-linking buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 1 mM TCEP) and incubated for 30 min at room temperature prior to initiating the cross-linking reaction. DSSO cross-linker was freshly dissolved in DMSO to a final concentration of 75 mM before being added to the protein solution at a final concentration of 1.5 mM. The reaction was incubated at 25 °C for 45 min and then quenched by adding 1 μL of 1.0 M Tris pH 8.0 and incubating an additional 10 min at 25 °C. Control reactions were performed in parallel without adding the DSSO cross-linker. All cross-linking reactions were carried out in three replicates. The presence of cross-linked proteins was confirmed by comparing to the no cross-link negative control samples using SDS-PAGE and Coomassie staining. The remaining cross-linked and non-cross-linked samples were separately pooled and then precipitated using methanol and chloroform. Dried protein pellets were resuspended in 12.5 μL of resuspension buffer (50 mM ammonium bicarbonate, 8 M urea, pH 8.0). ProteaseMAX (Promega, V5111) was added to 0.02%, and the solutions were mixed on an orbital shaker operating at 400 rpm for 5 min. After resuspension, 87.5 μL of digestion buffer (50 mM ammonium bicarbonate, pH 8.0) was added. Protein samples were reduced by adding 1 μL of 500 mM DTT followed by incubation of the protein solutions in an orbital shaker operating at 400 rpm at 56 °C for 20 min. After reduction, 2.7 μL of 550 mM iodoacetamide was added, and the solutions were incubated at room temperature in the dark for 15 min. Reduced and alkylated protein solutions were digested overnight using trypsin at a ratio of 1:150 (w/w trypsin:protein) at 37 °C. Peptides were acidified to 1% trifluoroacetic acid (TFA) and then desalted using C_{18}ZipTip (Millipore cat no. ZTC18 5096). Dried peptides were resuspended in 10 μL of 0.1% TFA in water. Samples were then frozen and stored at −20 °C until LC-MS analysis.

Liquid Chromatography and Mass Spectrometry. 500 ng of sample was injected (triplicate injections for cross-linked samples and duplicate injections for control samples) onto an UltiMate 3000 UHP liquid chromatography system (Dionex, ThermoFisher). Peptides were trapped using a μPAC C18 trapping column (PharmaFluidics) using a load pump operating at 20 μL/min. Peptides were separated on a 200 cm μPAC C18 column (PharmaFluidics) using a linear gradient (1% Solvent B for 4 min, 1−30% Solvent B from 4 to 70 min, 30−55% Solvent B from 70 to 90 min, 55−97% Solvent B from 90 to 112 min, and isocratic at 97% Solvent B from 112 to 120 min) at a flow rate of 800 nL/min. Gradient Solvent A contained 0.1% formic acid, and Solvent B contained

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

80% acetonitrile and 0.1% formic acid. Liquid chromatography eluate was interfaced to an Orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher) with a Nanospray Flex ion source (ThermoFisher). The source voltage was set to 2.5 kV, and the S-Lens RF level was set to 30%. Cross-links were identified using a previously described MS2-MS3 method[23] with slight modifications. Full scans were recorded from m/z 150 to 1500 at a resolution of 60 000 in the Orbitrap mass analyzer. The AGC target value was set to 4×10^5, and the maximum injection time was set to 50 ms in the Orbitrap. MS2 scans were recorded at a resolution of 30 000 in the Orbitrap mass analyzer. Only precursors with a charge state between 4 and 8 were selected for MS2 scans. The AGC target was set to 5×10^4, a maximum injection time of 150 ms, and an isolation width of 1.6 m/z. The CID fragmentation energy was set to 25%. The two most abundant reporter doublets from the MS2 scans with a charge state of 2−6, a 31.9721 Da mass difference (Kao et al. 2011), and a mass tolerance of ±10 ppm were selected for MS3. The MS3 scans were recorded in the ion trap in rapid mode using HCD fragmentation with 35% collision energy. The AGC target was set to 2×10^4, and the maximum injection time was set for 200 ms and the isolation width to 2.0 m/z.

Data Analysis. To identify cross-linked peptides, Thermo.Raw files were imported into Proteome Discoverer 2.5 (ThermoFisher) and analyzed via the XlinkX algorithm[24] using the MS2_MS3 workflow with the following parameters: MS1 mass tolerance, 10 ppm; MS2 mass tolerance, 20 ppm; MS3 mass tolerance, 0.5 Da; digestion, trypsin with four missed cleavages allowed; minimum peptide length of 4 amino acids, fixed modification, carbamidomethylation (C); variable modification, oxidation (M); and DSSO (K, S, T, Y). The XlinkX/PD Validator node was used for cross-linked peptide validation with a 1% false discovery rate (FDR). Identified cross-links were further validated and quantified using Skyline (version 19.1)[25] using a previously described protocol.[26] Cross-link spectral matches found in Proteome Discoverer were exported and converted to the sequence spectrum list format using Excel (Microsoft). Cross-link peak areas were assessed using the MS1 full-scan filtering protocol for peaks within 8 min of the cross-link spectral match identification. Peak areas were assigned to the specified cross-linked peptide identification if the mass error was within 10 ppm of the theoretical mass, if the isotope dot product was greater than 0.95, and if the peak was not found in the non-cross-linked negative control samples. The isotope dot product compares the distribution of the measured MS1 signals against the theoretical isotope abundance distribution calculated based on the peptide sequence. Its value ranges between 0 and 1, where 1 indicates a perfect match.[27] Pairwise comparisons were made using the "MSstats" package[28] implemented in Skyline to calculate relative fold changes and significance. Significant change thresholds were defined as a $\log_2$ fold change less than −2 or greater than 2 and $-\log_{10}$ p-value greater than 1.3 (p-value less than 0.05). The visualization of proteins and cross-links was generated using xiNET.[29]

The data have been deposited to the ProteomeXchange Consortium via the PRIDE[30] partner repository with the data set identifier PXD026191.

Hydrogen−Deuterium Exchange Mass Spectrometry (HDX-MS). Solution-phase amide HDX experiments were carried out with a fully automated system (CTC HTS PAL, LEAP Technologies, Carrboro, NC; housed inside a 4 °C cabinet) as previously described[31] with the following modifications. Peptides were identified using tandem MS (MS/MS) experiments performed on a QExactive (Thermo Fisher Scientific, San Jose, CA) over a 70 min gradient. Product ion spectra were acquired in a data-dependent mode, and the five most abundant ions were selected for the product ion analysis per scan event. The MS/MS *.raw data files were converted to *.mgf files and then submitted to MASCOT (version 2.3 Matrix Science, London, UK) for peptide identification. The maximum number of missed cleavages was set at 4 with the mass tolerance for precursor ions ±0.6 Da and for fragment ions ±8 ppm. Oxidation to methionine was selected for variable modification. Pepsin was used for digestion, and no specific enzyme was selected in MASCOT during the search. Peptides included in the peptide set used for HDX detection had a MASCOT score of 20 or greater. The MS/MS MASCOT search was also performed against a decoy (reverse) sequence, and false positives were ruled out if they did not pass a 1% false discovery rate.

For differential HDX, protein−protein complexes were formed by incubating nsp7 and nsp8 at 1:1, 3:1, or 1:3 molar ratios for 30 min at room temperature. The reactions (5 μL) were mixed with 20 μL of D_2O-containing HDX buffer (50 mM HEPES, 500 mM NaCl, 1 mM TCEP, pD 8.4) and incubated at 4 °C for 0, 10, 30, 60, 900, or 3600 s. Following on-exchange, unwanted forward- or back-exchange was minimized, and the protein was denatured by the addition of 25 μL of a quench solution (5 M urea, 1% TFA, pH 2). Samples were then immediately passed through an immobilized pepsin column (prepared in-house) at 50 μL min^{-1} (0.1% v/v TFA, 4 °C), and the resulting peptides were trapped and desalted on a 2 mm × 10 mm C_8 trap column (Hypersil Gold, ThermoFisher). The bound peptides were then gradient-eluted (4−40% CH_3CN v/v and 0.3% v/v formic acid) across a 2.1 mm × 50 mm C_{18} separation column (Hypersil Gold, ThermoFisher) for 5 min. Sample handling and peptide separation were conducted at 4 °C. The eluted peptides were then subjected to electrospray ionization directly coupled to a high-resolution Orbitrap mass spectrometer (QExactive, ThermoFisher).

The HDX experiments were performed in triplicates with single preparations of each protein/protein complex. The intensity weighted mean m/z centroid value of each peptide envelope was calculated and subsequently converted into a percentage of deuterium incorporation. This is accomplished by determining the observed averages of the undeuterated and fully deuterated spectra using the conventional formula described elsewhere.[32] The fully deuterated control, 100% deuterium incorporation, was calculated theoretically, and corrections for back-exchange were made on the basis of an estimated 70% deuterium recovery and accounting for 80% final deuterium concentration in the sample (1:5 dilution in D_2O HDX buffer). The statistical significance for the differential HDX data is determined by an unpaired t test for each time point, a procedure that is integrated into the HDX Workbench software.[33]

The HDX data from all overlapping peptides were consolidated to individual amino acid values using a residue averaging approach. Briefly, for each residue, the deuterium incorporation values and peptide lengths from all overlapping peptides were assembled. A weighting function was applied in which shorter peptides were weighted more heavily, and longer peptides were weighted less. Each of the weighted deuterium

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

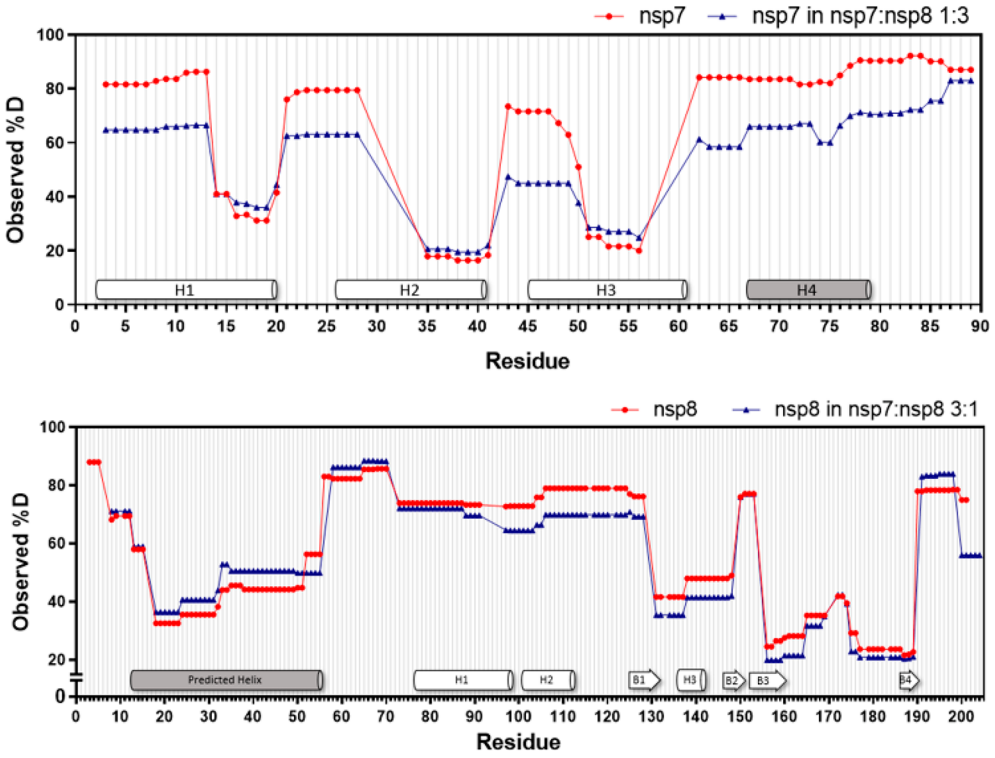


Figure 1. HDX-MS analysis of nsp7, nsp8, and nsp7:nsp8 complexes. Observed percent deuterium (%D) uptake of nsp7 (top) and nsp8 (bottom) in isolation and in complex during the 1 h experiment time course. Secondary structure from PDB 6YHU annotated in white and predicted secondary structure based on observed percent deuterium in gray.

incorporation values were then averaged incorporating this weighting function to produce a single value for each amino acid. The initial two residues of each peptide, as well as prolines, were omitted from the calculations. This approach is similar to that previously described.[34]

Deuterium uptake for each peptide is calculated as the average of the observed percent deuterium (%D) for all on-exchange time points, and the difference in average %D values between the unbound and bound samples is presented as a heat map with a color code given at the bottom of the figures (warm colors for deprotection and cool colors for protection). Peptides are colored by the software automatically to display significant differences, determined either by a >5% difference (less or more protection) in average deuterium uptake between the two states or by using the results of unpaired *t* tests at each time point (*p*-value <0.05 for any two time points or a *p*-value <0.01 for any single time point). Peptides with nonsignificant changes between the two states are colored gray. The exchange at the first two residues for any given peptide is not colored. Each peptide bar in the heat map view displays the average Δ %D values, associated standard deviation, and the charge state. Additionally, overlapping peptides with a similar protection trend covering the same region are used to rule out data ambiguity.

The data have been deposited to the ProteomeXchange Consortium via the PRIDE[30] partner repository with the data set identifier PXD026066.

RESULTS AND DISCUSSION

HDX-MS and XL-MS are Complementary Techniques to Investigate Structural Dynamics. While X-ray crystallography and cryo-EM are the gold-standard techniques for determining atomic resolution structures of macromolecular complexes, they represent static snapshots of protein structures. While they may reveal singular aspects of the conformational ensemble, they do not fully capture the structural plasticity inherent to proteins in solution. Characterizing this structural plasticity is essential to understand the biologically relevant structure-function relationships of proteins and protein complexes. Hydrogen−deuterium exchange mass spectrometry (HDX-MS) and cross-linking MS (XL-MS) are complementary techniques that can be used to garner information on the solution-phase dynamics of proteins and protein complexes. Specifically, HDX-MS reports on protein backbone dynamics while XL-MS reports on side-chain residency and reactivity.

HDX-MS Provides a Footprint of the nsp7 and nsp8 Secondary Structure. HDX-MS measures changes in mass due to isotopic exchange of the amide hydrogens of the protein backbone with the solvent. The rate of isotopic exchange is influenced by the intrinsic properties of the amino acid sequence, the folded state of the protein, and the dynamics of the hydrogen-bonding network. Given that protein secondary structures such as α-helices and β-sheets are characterized and stabilized via hydrogen bonds, HDX-MS can be used to predict or assess the dynamics of proteins' secondary structure elements. As shown in Figure 1, HDX-MS analyses of nsp7 and nsp8 proteins in isolation and in complex reveal similar patterns of solvent exchange behaviors. Furthermore, this behavior is largely in agreement with the secondary structure observed in the published nsp7:nsp8 heterotetrameric crystal structure (PDB: 6YHU).[13] For nsp7, the three α-helical bundles within the protein ($H1^{nsp7}$Lys2-Leu20, $H2^{nsp7}$Ser26-Leu41, $H3^{nsp7}$Thr45-Ser61) are characterized by lower deuterium uptake relative to other regions of the protein. For nsp8, the C-terminal subdomain is predicted to be

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

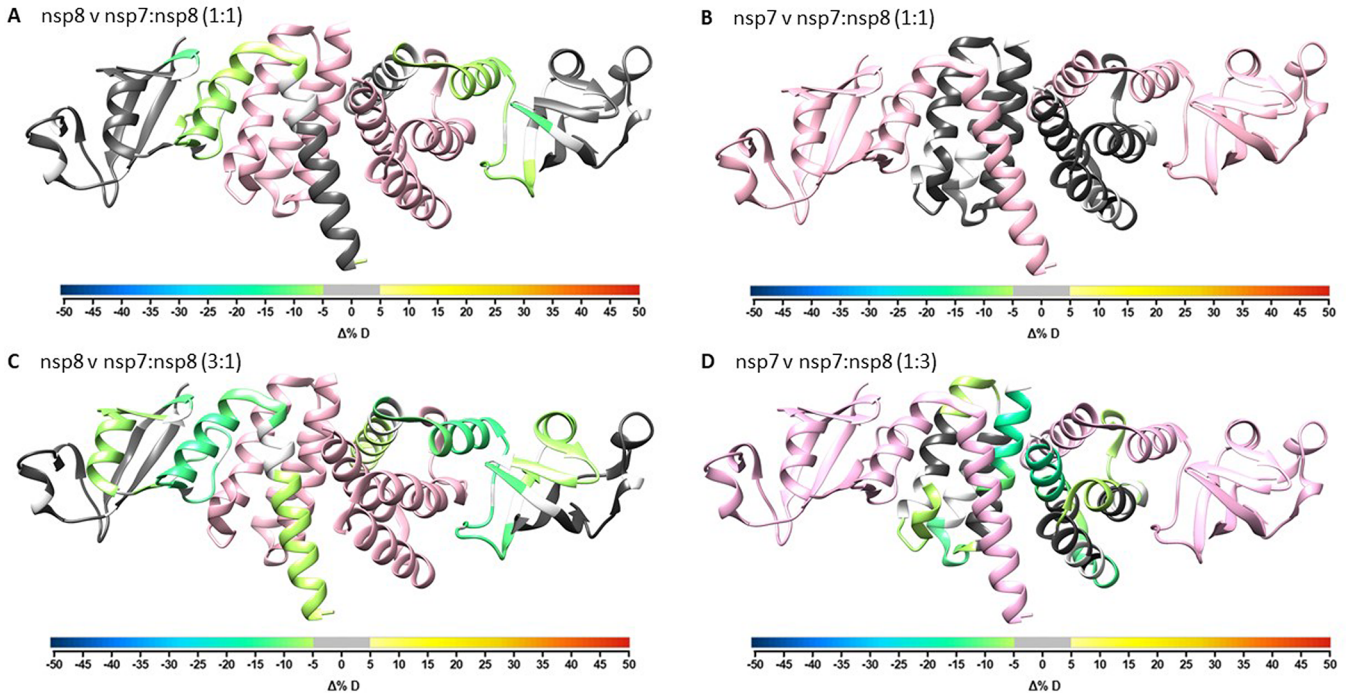


Figure 2. Overlay of differential HDX-MS perturbation values onto PDB 6YHU. Perturbation values for nsp8 vs nsp7:nsp8 1:1 (A), nsp7 vs nsp7:nsp8 1:1 (B), nsp8 vs nsp7:nsp8 3:1 (C), and nsp7 vs nsp7:nsp8 1:3 (D) colored according to the change in percent deuterium levels shown in the color bar and respective partner protein colored in pink. Residues not observed by HDX-MS are colored in white.

composed of four antiparallel β-strands (B1^{nsp8} Ala125-Ile132, B2^{nsp8} Thr146-Tyr151, B3^{nsp8} Ala152-Val160, B4^{nsp8} Leu186-Arg190) with an α-helix inserted between the first two β-strands (H3^{nsp8} Tyr135-Thr142). The α-helical region of this domain has the lowest deuterium uptake. However, the other two predicted α-helices (H1^{nsp8} Glu77-Leu98, H2^{nsp8} Asp101-Asp112) show a high percent deuterium uptake suggesting greater conformational dynamics such that these regions are not ordered when nsp8 is in isolation and thus form into helices upon interaction with nsp7.

As the X-ray structure (PDB: 6YHU) used truncated forms of both nsp7 and nsp8 to facilitate crystallization, the HDX footprint of the full-length protein can predict the secondary structure of the missing regions. The truncated nsp8 N-terminus (residues 1−75; note that residues 193−198 in the C-terminus are truncated too) shows low deuterium uptake, 30−60% observed deuterium, which suggests the presence of helical structure in this region. This prediction is supported by available cryo-EM structures of nsp8 in complex with nsp7, nsp12, and RNA that reveal the nsp8 N-terminus to be primarily helical in nature.[7−10] This region is not always fully resolved in these complexes suggesting that it may be more mobile and sampling multiple orientations, which is supported by the HDX results showing that this region has relatively higher percent deuterium uptake (32−45% deuterium uptake) compared to the other helices in nsp8 and nsp7 (20−25% deuterium uptake). The truncated C-terminus of nps7 (residues 72−83) shows high levels of deuterium uptake suggesting greater dynamics in this region and potentially lacking a defined secondary structure (Figure 1).

Differential HDX Validates nsp7:nsp8 Binding Surfaces and Reveals Additional Long-Range Conformational Dynamics. Differential HDX (comparison of the exchange kinetics of individual proteins in isolation to that within the protein complex) of the nsp7:nsp8 complex was initially conducted at a 1:1 molar ratio in line with the reported stoichiometry.[13] Reduced deuterium uptake was only observed in nsp8 in H2^{nsp8}, B1^{nsp8}, and B3^{nsp8} (Figure 2A, Figures S1B and S2B and Tables S1 and S3). Based on the X-ray structure (PDB: 6YHU), H2^{nsp8} is located at the nsp7:nsp8 dimerization interface. This interface is stabilized by a leucine zipper motif composed of nsp7 Leu56, Leu60, and Leu71 and nsp8 Leu95 and Leu102. Regions within nsp8 containing both Leu residues are significantly protected to exchange when nsp8 is in complex with nsp7. Furthermore, peptides containing nsp8 Phe92 and adjacent hydrophobic residues implicated in stabilizing the complex are significantly protected from exchange.[13] Surprisingly, there was no detectable protection to solvent exchange in nsp7 under the experimental conditions employed at a 1:1 molar ratio (Figure 2B, Figures S1A and S2A and Tables S2 and S3). Therefore, we repeated the analysis with 3-fold molar excess of nsp8 to saturate all nsp7 binding sites (Figure 2D, Figures S2C and S2A and Tables S2 and S3). In this experiment, with excess nsp8, protection to exchange was observed in all three α-helices of nsp7 as well as in its C-terminal region. H1^{nsp7} and H3^{nsp7} are predicted to be involved in forming the nsp7:nsp8 dimer interface while H1^{nsp7} and H2^{nsp7} are predicted to be involved with the nsp7:nsp8 2:2 heterotetrameric interface.[13] Significant protection from solvent exchange in H1^{nsp7} as well as the N-termini of H2^{nsp7} and H3^{nsp7} is consistent with this speculation. Regions of nsp7 that contain a disulfide bridge formed between the symmetric nsp7 Cys8 residues to stabilize the heterotetramer demonstrated the highest magnitude of solvent protection. Interestingly, the C-terminal region of nsp7 showed levels of protection similar to that of the nsp7 Cys8 region. In PDB 6YHU, this region is not observed and thus was not reported to be involved in the interaction with nsp8. This observed protection outside of the binding interface suggests that complex formation may induce allosteric conformational

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

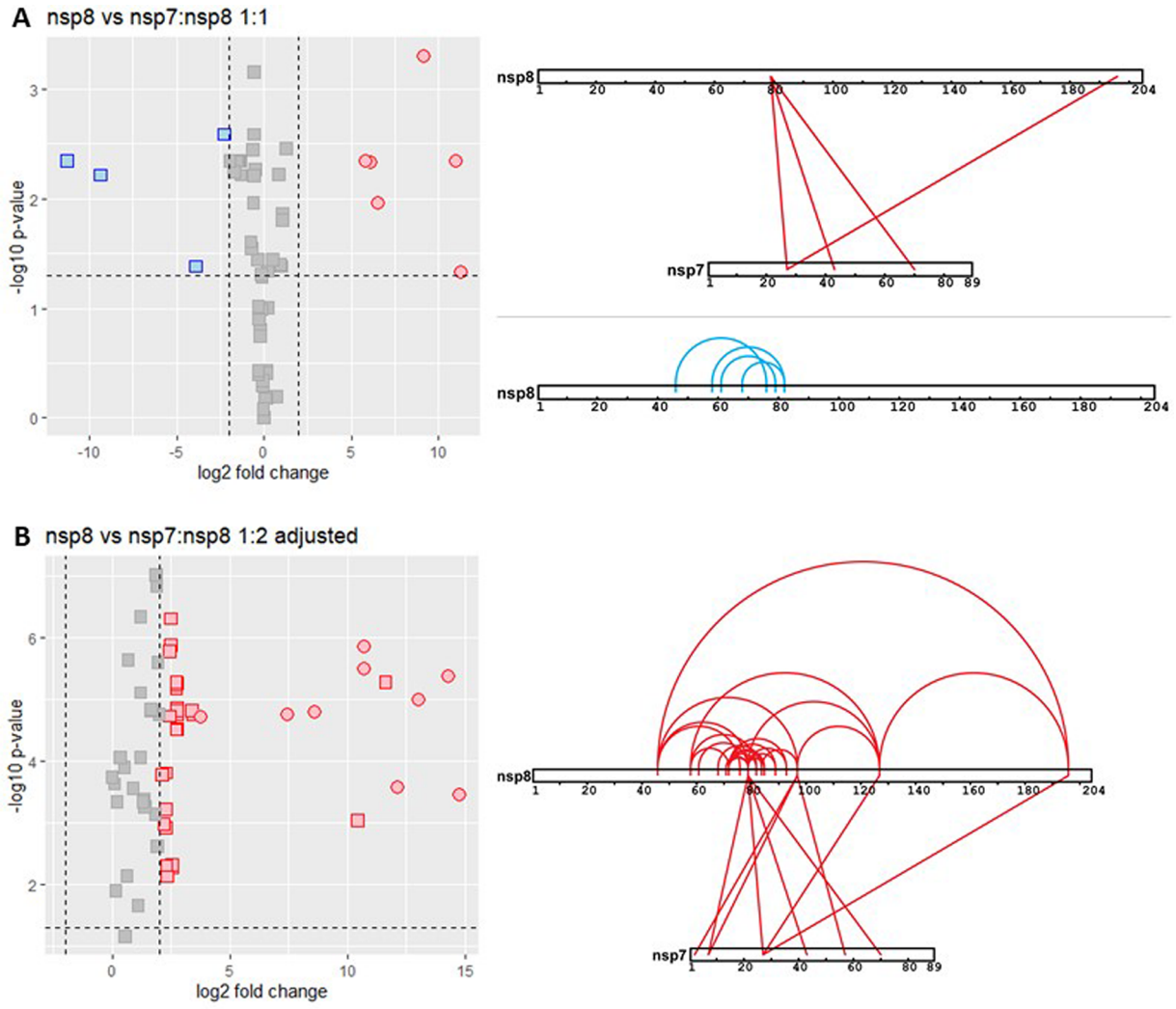


Figure 3. XL-MS relative quantification results of nsp7, nsp8, and the nsp7:nsp8 complex. (A) Volcano plot summarizing the relative quantification results of cross-links in nsp8 versus nsp7:nsp8 1:1. Cross-links enriched in the nsp7:nsp8 1:1 complex shown in the red box and cross-links enriched in the nsp8 alone sample shown in the blue box. (B) Volcano plot summarizing the relative quantification results of nsp8 versus nsp7:nsp8 1:2 (adjusted) cross-linking reactions. Intra nsp8 cross-links shown as squares and inter nsp7-nsp8 cross-links shown as circles in volcano plot. Cross-links enriched in nsp8 alone and in the nsp7:nsp8 1:1 complex mapped to sequences in blue andred, respectively.

changes that may influence the function of the complex. The other structures in the linear heterotetramer group (PDB: 6WQD and 7JLT) were able to resolve more of the C-termini of nsp7, and these structures show that this region is organized into a fourth α-helix (Asp67−Arg79). The significant protection to solvent exchange observed in this region supports the formation of $H4^{nsp7}$ upon binding to nsp8 (Table S2).

To complement the above experiment, we repeated the differential HDX analysis of nsp8 with a 3-fold molar excess of nsp7 to saturate nsp8 binding sites (Figure 2C, Figures S1D and S2A and Tables S1 and S3). In the presence of excess nsp7, an increased magnitude of protection to solvent exchange was observed in regions that showed protection from exchange at the 1:1 molar ratio. Additional protection to exchange was observed in new regions including $H1^{nsp8}$, $H3^{nps8}$, and $B2^{nsp8}$. The HDX analyses of nsp8 in complex with nsp7 suggest that observed $H1^{nsp8}$ and $H2^{nps8}$ in the nsp7:nsp8 crystal structure are only formed upon binding to nsp7 as this region shows high percent deuterium levels in isolation and significantly lower percent deuterium levels in the presence of nsp7.

XL-MS Validates nsp7:nsp8 Direct Interaction. To complement the analysis of backbone dynamics, we conducted differential XL-MS to observe changes in side-chain residency and reactivity. Figure S3 shows the XL reaction efficiency assessed by SDS-PAGE. To minimize the false assignment of cross-link products, we utilized the MS cleavable cross-linker disuccinimidyl sulfoxide (DSSO). Cross-linked peptides are cleaved using collision-induced dissociation (CID) at the MS2 level to produce unique fragment pair ions that trigger MS3 analysis of the single peptide chain fragments allowing for unambiguous sequence identification. Cross-links were identified as described in the Experimental Section using the XlinkX software in Proteome Discoverer,[24] and cross-link spectral matches were exported for further manual validation in Skyline[25] using the MS1 full-scan filtering protocol. Peak areas were only assigned to cross-linked peptides if the isotope dot product was greater than 0.95 and if the peak was not found in the non-cross-linked negative control sample. This allowed us to perform differential XL-MS via a relative pairwise comparison using the MSstats package in Skyline.[28] The thresholds for significant changes in cross-links were set to a

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

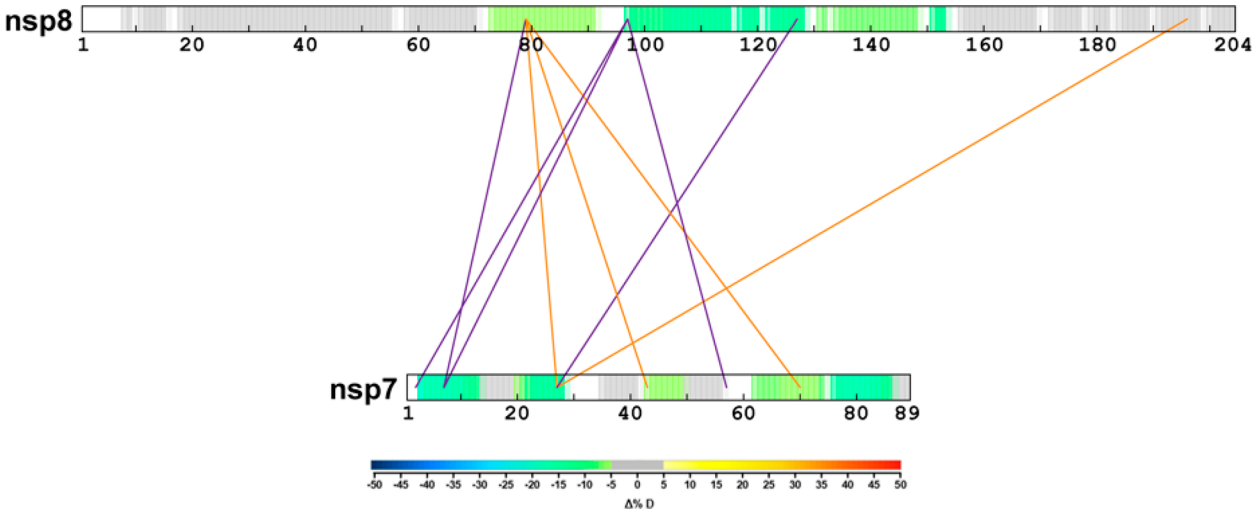


Figure 4. Overlay of HDX-MS and XL-MS results on nsp7 and nsp8 sequences. Observed inter nsp7-nsp8 cross-links only identified in the nsp7:nsp8 1:1 complex are colored in orange, and inter nsp7-nsp8 cross-links also identified in the nsp7:nsp8 1:2 complex are colored in purple. Consolidated changes in percent deuterium uptake from Figure 1C,D are overlaid on the nsp7 and nsp8 sequences, respectively. The initial Met residue is removed from both nsp7 and nsp8 sequences to maintain correct residue numbering.

$\log_2$ fold change less than −2 or greater than 2 and $-\log_{10}$ *p*-value greater than 1.3 (*p*-value less than 0.05).

XL-MS of nsp7 in isolation resulted in no detectable intraprotein cross-links. Several dead-end cross-links were identified (Figure S4A). The presence of only dead-end cross-links suggests nsp7 to be in a more linear conformation unable to form intraprotein cross-links. It also suggests no significant nsp7:nsp7 side-chain interactions (homomultimerization). While there have been mixed reports suggesting that nsp7 does or does not engage in self-interaction,[35,36] the results presented here support that nsp7 does not self-interact. Interestingly, in the presence of nsp8, the residues with dead-end cross-links all form inter nsp7-nsp8 cross-links with nsp8. XL-MS analysis of nsp8 in isolation resulted in the identification of 25 intraprotein cross-links (Figure S4B). As shown in Figure 3A, the XL-MS analysis of the nsp7:nsp8 complex resulted in the identification of both intra nsp8 cross-links as well as inter nsp7-nsp8 cross-links. A total of four inter nsp7-nsp8 cross-links were identified in the nsp7:nsp8 complex at a 1:1 molar ratio. Comparing cross-links detected in nsp8 to those detected in the nsp7:nsp8 complex, we observed that a majority of the intra nsp8 cross-links are not significantly different (unchanged) between the two sample groups. However, the intra nsp8 cross-links located near Lys79, which forms inter nsp7-nsp8 cross-links in the nsp7:nsp8 sample group, are significantly enriched in the nsp8 alone samples (decreased in the nsp7:nsp8 complex). These data suggest that residues adjacent to nsp8 Lys79 are involved in the interaction with nsp7 such that the formation of inter nsp7-nsp8 cross-links is favored over intra nsp8 cross-links when in the presence of nsp7.

Previous studies have demonstrated that nsp7 and nsp8 play important roles as cofactors of nsp12 to give it processive RNA-dependent RNA polymerase activity.[7] Numerous cryo-EM structures of nsp7, nsp8, and nsp12 with or without RNA report a 1:1:2 nsp12:nsp7:nsp8 stoichiometry.[6−10] Based on these studies, we repeated the XL-MS analysis of the nsp7:nsp8 complex at a molar ratio of 1:2, respectively. For this differential analysis, the intensity values of intra nsp8 cross-links identified in the nsp7:nsp8 complex were halved in order to appropriately compare intensities to the nsp8 alone sample (Figure 3B). This analysis revealed four additional inter nsp7-nsp8 cross-links. While the majority of intra nsp8 cross-links remain unchanged between the two samples, some are slightly enriched in the nsp7:nsp8 complex when at a molar ratio of 1:2. The presence of these enriched intra nsp8 cross-links suggests that nsp8 interacts with nsp7 and nsp8 simultaneously, consistent with a 1:2 nsp7:nsp8 stoichiometry. Furthermore, nsp8 self-interaction has also been characterized by a yeast two-hybrid screen and coimmunoprecipitation experiment[35] as well as a novel phase separation-based method used to identify protein−protein interactions *in vitro*.[36]

HDX-MS and XL-MS Results Overlay to Confirm nsp7:nsp8 Interaction in Solution. Overlaying the HDX-MS and XL-MS results onto the nsp7 and nsp8 sequences as shown in Figure 4 demonstrates the complementary nature of the two techniques as the inter nsp7-nsp8 cross-links map to regions protected from solvent exchange in the differential HDX-MS experiment. Interestingly, regions containing both nsp8 Lys79 and Lys97, located within $H1^{nsp8}$, both demonstrate the largest magnitude of protection from deuterium exchange and make the most interprotein cross-links with nsp7. These observations are consistent with the X-ray structure (PDB: 6YHU) showing $H1^{nsp8}$ to be involved in stabilizing both the dimer and heterotetramer interface. An additional cross-link observed between nsp8 Lys196 (truncated in the aforementioned crystal structure) and nsp7 Lys27 suggests that additional contact sites are likely outside of the proposed helical bundle interactions.

Mapping Inter-nsp7 nsp8 Cross-Links Can Be Used to Assess Three-Dimensional Structures. The XL-MS analysis can be used to assess available three-dimensional structures and their relevance to the solution-based structure. While the DSSO cross-linker measures 10.1 Å, it has a theoretical upper limit of approximately 26 Å to account for the orientation of side chains from cross-linked residues.[37] Xlink Analyzer[38] in Chimera[39] was used to map the interprotein cross-links onto all of the available nsp7:nsp8 crystal structures (Table S4). The calculated distances of the interprotein cross-links at either the dimer or heterotetramer

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

interface within each group of structures are in good agreement. Comparing the interprotein cross-links mapped to the dimer interface across the three groups, we also observe good agreement in distances. However, when comparing the interprotein cross-links mapped to the heterotetramer interface, differences are observed in which cross-links are above or below the 26 Å limit. The cubic heterotetramer structure is unable to successfully map two cross-links (nsp7 Lys7 to nsp8 Lys79 and nsp7 Lys43 to nsp8 Lys79) to either dimer or heterotetramer interfaces. On the contrary, all of the interprotein cross-links were successfully mapped to either interface in the linear heterotetramer structures. This suggests that the linear heterotetramer structure is a better representation of the structural conformation of the nsp7:nsp8 complex in solution.

Focusing on the published nsp7:nsp8 structure (PDB: 6YHU) that shows the relevant linear heterotetramer conformation, we can map 10 of the 11 interprotein cross-links onto the structure (Figure 5). The cross-link identified to

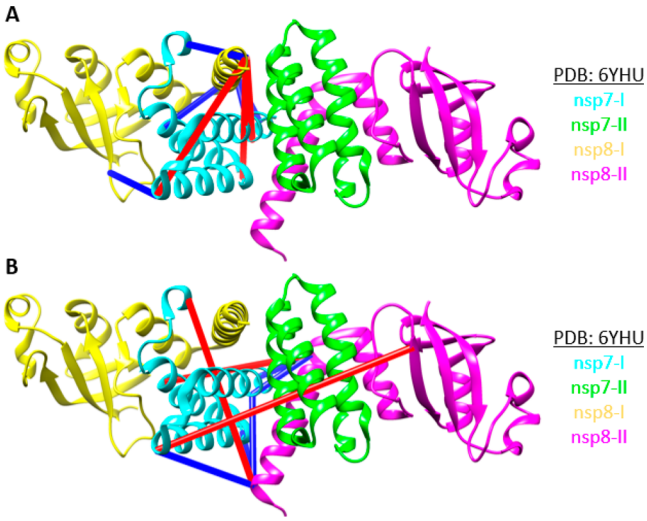


Figure 5. Mapping of nsp7 nsp8 interprotein cross-links on the dimer (A) and heterotetramer (B) interaction surfaces of the SARS-CoV-2 nsp7 nsp8 heterotetramer crystal structure (PDB: 6YHU). Cross-links greater than 26 Å distance are labeled in red, and cross-links less than 26 Å are labeled in blue.

nsp8 Lys196 could not be mapped as this residue fell outside the truncated protein sequence used to obtain the structure. All 10 cross-links were successfully mapped to either the dimer (Figure 5A, Table 1: dimer interface) or heterotetramer interface (Figure 5B, Table 1: heterotetramer interface) in PDB 6YHU. Only two cross-links were successfully mapped to both interfaces: nsp7 Lys7 to nsp8 Lys97 and nsp7 Lys2 to nsp8 Lys97. These cross-linked residues are located within $H1^{nsp7}$ and $H1^{nsp8}$ which are the two helices predicted to be involved in forming both the dimer and heterotetramer interface. The third cross-link between these two helices (nsp7 Lys7 and nsp8 Lys79) has a calculated distance of 26.8 Å which is outside the threshold for DSSO distance of 26 Å. However, this threshold is not a hard cutoff as it tries to account for the various possible orientations of the Lys side chains. As such, the nsp7 Lys7 and nsp8 Lys79 cross-link should also be considered to be in agreement with the structure. Thus, all solution-based interprotein cross-links observed between $H1^{nps7}$ and $H1^{nsp8}$ can be mapped to both

Table 1. Distances of Inter-nsp7 nsp8 Cross-Links Mapped to the SARS-CoV-2 nsp7:nsp8 Heterotetramer Structure (PDB: 6YHU)

nsp7 position	nsp8 position	Distance at Dimer Interface (Å)	Distance at Heterotetramer Interface (Å)
27	79	29.3	19.3
27	196	NF	NF
27	127	16.9	46.7
70	79	18.5	34.4
7	79	26.8	16.1
43	79	37.3	14.5
57	97	17.9	28.4
7	97	13.0	14.1
2	97	9.4	11.5

the dimer and heterotetramer surfaces, and the remaining cross-links can be mapped to either the dimer or heterotetramer interaction surfaces.

Mapping Inter-nsp7 nsp8 Cross-Links to Higher-Order Assembly Structures Provides Insight into RTC Complex Formation. The CoV-2 nsp7:nsp8 dimer was reported to superimpose well with the subunits of higher-order assemblies of SARS-CoV nsp7:nsp8 and the cryo-EM structure of the replicating polymerase nsp7, nsp8, nsp12, and RNA.[13] To validate this finding, we mapped our inter-nsp7 nsp8 cross-links onto these additional structures. Zhai et al. reported a hexadecameric circular complex for SARS-CoV nsp7:nsp8.[20] They showed that nsp8 can exist in two different conformations, with one adopting a "golf-club-like" structure (Figure 6: nsp8-II in pink) and the other adopting a "golf-club"

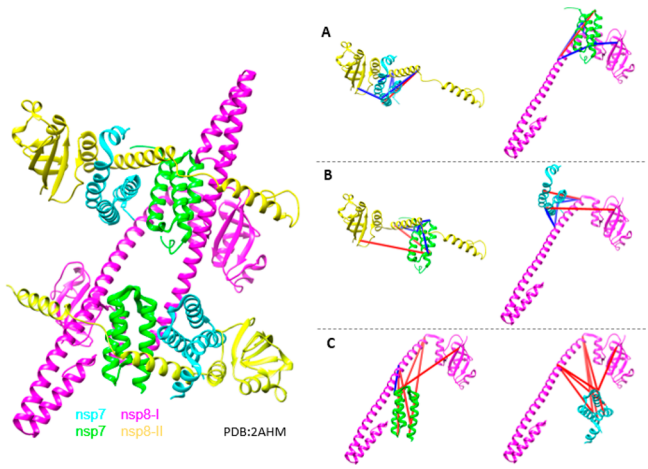


Figure 6. Mapping of nsp7-nsp8 interprotein cross-links on the dimer (A), heterotetramer (B), and hexadecamer (C) interaction surfaces of the SARS-CoV nsp7:nsp8 hexadecamer cryo-EM structure (PDB: 2AHM). Cross-links greater than 26 Å distance are labeled in red, and cross-links less than 26 Å are labeled in blue.

with a bent shaft (Figure 6: nsp8-I in yellow). Nevertheless, the contacts between either nsp8 conformations and nsp7 are roughly the same. The nsp8 head is primarily responsible for interacting with nsp7, and this region is similar in either conformations. As shown in Figure 6A, mapping the observed inter nsp7-nsp8 cross-links to either possible SARS-CoV

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

Table 2. Distances of Internsp7-nsp8 Cross-Links Mapped to the SARS-CoV nsp7:nsp8 Hexadecamer Structure (PDB: 2AHM)[a]

nsp7 position	nsp8 position	PDB: 6YHU Dimer Interface	Dimer Interface		PDB: 6YHU Heterotetramer Interface	Heterotetramer Interface		Hexadecamer Interface	
		Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)
27	79	29.3	29.2	28.6	19.3	19.7	19.8	25.6	38.6
27	196	NF	NF	NF	NF	NF	NF	NF	NF
27	127	16.9	17.9	18.6	46.7	43.3	43.9	49.5	34.6
K70R	79	18.5	---	---	34.4	---	---	---	---
7	79	26.8	27.8	27.7	16.1	16.1	16.0	39.2	38.0
43	79	37.3	38.5	38.3	14.5	14.0	14.3	46.3	39.3
57	97	17.9	17.8	18.0	28.4	31.0	31.2	49.4	61.7
7	97	13.0	13.2	13.0	14.1	16.6	17.0	60.8	52.8
2	97	9.4	9.8	9.4	11.5	16.6	17.0	65.4	58.3

[a]Cross-links to residues not resolved in the structure are labeled as not found (NF).

nsp7:nsp8 dimers shows similar cross-link distances in both conformations (Table 2: dimer interface). These interprotein cross-link distances from the SARS-CoV nsp7:nsp8 dimer interface are similar to the observed cross-link distances in the CoV-2 dimer (PDB: 6YHU) (Figure 6A, Table 2: dimer interface). Of the 10 inter nsp7-nsp8 cross-links mapped in the SARS-CoV-2 structure, only one cross-link could not be mapped in the structure, as the corresponding residue of Lys70 is Arg70 in SARS-CoV. Comparing distances from the CoV-2 heterotetramer interface to the equivalent interface in SARS-CoV, we also observe similar cross-link distances (Figure 6B, Table 2: heterotetramer interface). However, when the cross-links are mapped to the hexadecameric interface, all but one cross-link violate the 26 Å limit (Figure 6C, Table 2: hexadecamer interface). This suggests that full-length CoV-2 nsp7 and nsp8 likely do not assemble into the higher-order complex observed for CoV nsp7:nsp8 complex.

This finding is supported by other structural analyses that determine that the heterotetramer is the likely predominant species at relevant *in vivo* concentrations, such that the observed CoV nsp7:nsp8 hexadecamer may be an arrangement stabilized by crystal lattice formation and perhaps less relevant in solution or biologically.[40] These findings are, however, at odds with the proposed transition model by Wang et al. of the CoV-2 primase complex to the polymerase complex.[8] Their model suggests that RNA binds the hexadecameric nsp7:nsp8 structure to *de novo* synthesize primers before dissociating in halves to assemble with nsp12 prebound with an nsp8 molecule. The results presented here suggest that CoV-2 does not assemble into a hexadecameric structure and thus would require RNA to be able to bind the nsp7:nsp8 dimer or heterotetramer to support primer *de novo* synthesis. An electrostatic surface analysis by Konkolova et al. suggests that while the dimer does not form any highly positively charged surfaces, the heterotetramer interface is framed by positively charged residues that can serve as the putative RNA-binding site.[13] We can thus hypothesize that RNA binds the heterotetramer structure to *de novo* synthesize primers and then dissociates into the dimers to allow assembly with nsp12. Further experiments to confirm that the nsp7:nsp8 hetero-tetrameric structure can bind RNA are still needed.

To investigate the hypothesis that the nsp7:nsp8 dimer can bind nsp12 in the given conformation, we mapped the inter nsp7-nsp8 cross-links to the replicating polymerase structure (PDB: 6YYT) reported by Hillen et al. as the nsp7:nsp8 X-ray dimer was shown to superimpose well onto the nsp7:nsp8 dimer in this complex.[7] All 10 cross-links from the nsp7:nsp8 structure were mapped, and similar cross-link distances were determined (Figure 7A, Table 3). This suggests that nsp7:nsp8

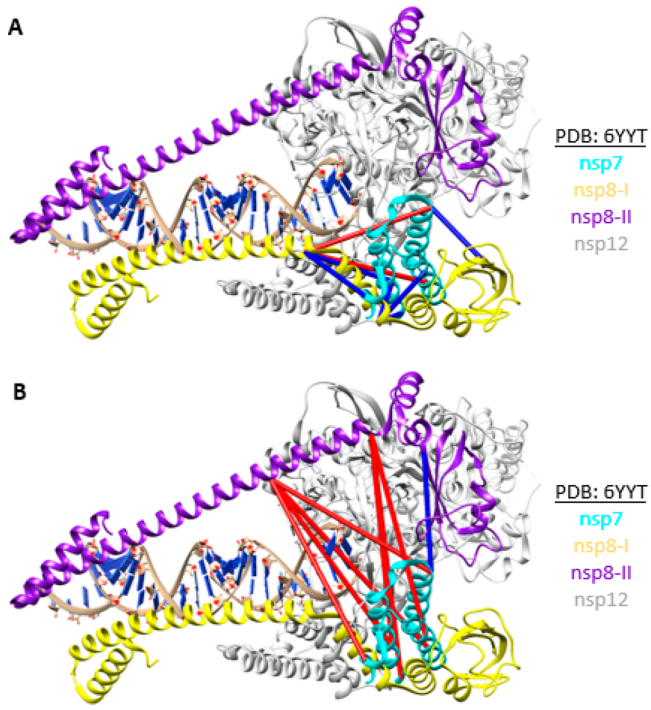


Figure 7. Mapping of nsp7-nsp8 interprotein cross-links to nsp8-I (A) or nsp8-II (B) of the SARS-CoV-2 replicating polymerase structure (PDB: 6YYT). Cross-links greater than 26 Å distance are labeled in red, and cross-links less than 26 Å are labeled in blue.

binding to nsp12 does not alter nsp7-nsp8 interaction. Mapping the cross-links between nsp7 and the additional nsp8 subunit gave cross-link distances all greater than 26 Å, consistent with the notion that the second nsp8 subunit does not make any contact with nsp7 and may be prebound to nsp12 (Figure 7B, Table 3).

Additional structures of the replicating polymerase have been reported since Hillen et al. published their structure, including structures bound to RNA (PDB: 7BZF, 6XQB) or unbound (PDB: 7BV1, 7BW4, 7BTF, and 6M71).

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

Table 3. Distances of Inter-nsp7 nsp8 Cross-Links Mapped to the SARS-CoV-2 Replicating Polymerase Complex Structure (PDB: 6YYT)[a]

nsp7 position	nsp8 position	PDB: 6YHU Dimer Interface Distance (Å)	PDB: 6YYT nsp7 to nsp8-I Distance (Å)	PDB: 6YYT nsp7 to nsp8-II Distance (Å)
27	79	29.3	30.9	43.6
27	196	NF	NF	NF
27	127	16.9	17.6	27.7
70	79	18.5	22.1	52.5
7	79	26.8	23.4	48.5
43	79	37.3	33.9	55.5
57	97	17.9	17.7	48.9
7	97	13.0	13.3	51.0
2	97	9.4	10.3	59.4

[a]Cross-links to residues not resolved in the structure are labeled as not found (NF).

Comparing the structures of the apo polymerase complex without RNA bound to the nsp7:nsp8 heterotetramer, we observe that more of the nsp8 N-terminal cannot be resolved, and thus, we are not able to map the cross-links to nsp8 Lys79. All other inter-nsp7 nsp8 cross-links were successfully mapped to the nsp7:nsp8 dimer and gave similar cross-link distances to the distances observed at the dimer interface of the nsp7:nsp8 heterotetramer (PDB: 6YHU) (Table 4). This suggests that while the nsp8 N-terminal may become more flexible, the core of the nsp7:nsp8 interface is not significantly altered by the presence of nsp12.

Analyzing the polymerase structures bound to DNA, we note that the Hillen et al. structure described above represents the complex in a pretranslocation state, while the two additional structures (PDB: 7BZF and 6XQB) are representative of the complex in a post-translocation state.[8,12] Comparing these two states we observe that more of the nsp8 N-terminal can be resolved in the pretranslocated state versus the post-translocation state as the cross-links to Lys79 can be mapped in PDB 6YYT but not PDB 7BZF and 6XQB (Table 5). These cross-link distances are similar to those observed in the nsp7:nsp8 heterotetramer. The remaining cross-links can be mapped in both the pre- and post-translocation states and show similar distances (Table 5). These observations suggest that the presence of RNA in the polymerase complex in the pretranslocation state results in transient stabilization of the nsp8 N-terminal helix, until subsequent rounds of nucleotide addition increase its mobility. This is supported by the Hillen et al. publication, which described the nsp8 N-terminal helix and its extensions as "sliding poles" that interact with the extending RNA duplex.[7] As such, the periodic movement of the RNA duplex as it undergoes additional polymerization cycles may be sensed by and increase the mobility of the nsp8 N-terminal helix. This does not appear to have any significant effect on the core of the nsp7:nsp8 interface as the other cross-links have similar distances in the pre- and post-translocation states. Additional XL-MS and HDX-MS experiments with nsp7, nsp8, nsp12, and/or RNA are required to validate this proposed dynamic movement of the nsp7:nsp8 complex as it transitions from the primase complex to the replicating polymerase complex that facilitates viral RNA synthesis.

CONCLUSIONS

HDX-MS and XL-MS are complementary structural proteomic techniques used to characterize proteins and protein complexes. While X-ray crystallography and cryo-EM may identify one or more predominant conformations, they present a static representation of the protein or protein complex. Moreover, the observed conformation may be influenced by the crystallization process and be less relevant in solution. As such, HDX-MS and XL-MS serve as alternative and/or complementary techniques to interrogate proteins in-solution to allow for a more direct analysis of protein conformational dynamics. HDX-MS specifically probes protein backbone dynamics while XL-MS reports on side-chain residency and reactivity. This information can then be used to validate or build integrated models of three-dimensional structures to ultimately inform our understanding of how the protein structure correlates with biological function.[41−44]

Herein we used HDX-MS and XL-MS to probe the solution-phase structural dynamics of CoV-2 nsp7 and nsp8 in isolation and in complex to ultimately assess the relevance of the available nsp7:nsp8 crystal structures. HDX-MS revealed that the solvent exchange kinetics of nsp7 and nsp8 in isolation mimics the behavior of the proteins in complex, suggesting an overall similar secondary structure for nsp7 and nsp8 in isolation or in complex. While the available crystal structures suggest three possible structural arrangements (dimer, linear heterotetramer, and cubic heterotetramer), the differential HDX-MS and XL-MS results obtained suggest that the linear heterotetramer structure is likely to be the most relevant

Table 4. Distances of Inter-nsp7 nsp8-I Cross-Links Mapped to Available SARS-CoV-2 Apo Polymerase Complex Structures Composed of nsp7, nsp8, and nsp12[a]

nsp7 position	nsp8 position	PDB: 6YHU Distance (Å)	PDB: 7BV1 Distance (Å)	PDB: 7BW4 Distance (Å)	PDB: 7BTF Distance (Å)	PDB: 6M71 Distance (Å)
27	79	29.3	NF	NF	NF	NF
27	196	NF	NF	NF	NF	NF
27	127	16.9	17.1	16.3	17.2	NF
70	79	18.5	NF	NF	NF	NF
7	79	26.8	NF	NF	NF	NF
43	79	37.3	NF	NF	NF	NF
57	97	17.9	17.7	18.0	17.9	18.2
7	97	13.0	13.3	13.5	13.2	13.0
2	97	9.4	10.2	10.8	10.1	10.5

[a]Cross-links to residues not resolved in the structure are labeled as not found (NF).

https://doi.org/10.1021/jasms.1c00086
J. Am. Soc. Mass Spectrom. 2021, 32, 1618−1630

Table 5. Distances of Inter nsp7-nsp8-I Cross-Links Mapped to Available SARS-CoV-2 Replicating Polymerase Complex Structures Composed of nsp7, nsp8, nsp12, and RNA[a]

nsp7 position	nsp8 position	PDB: 6YHU	*Pre-translocation State* PDB: 6YYT	*Post-translocation State* PDB: 7BZF	PDB: 6XQB
		Distance (Å)	Distance (Å)	Distance (Å)	Distance (Å)
27	79	29.3	30.9	NF	NF
27	196	NF	NF	NF	NF
27	127	16.9	17.6	17.2	NF
70	79	18.5	22.1	NF	NF
7	79	26.8	23.4	NF	NF
43	79	37.3	33.9	NF	NF
57	97	17.9	17.7	18.3	19.3
7	97	13.0	13.3	13.2	13.1
2	97	9.4	10.3	10.4	9.5

[a]Cross-links to residues not resolved in the structure are labeled as not found (NF).

structure for the complex in solution. This structure, represented by the published structure PDB 6YHU, proposes a 2:2 heterotetrameric structure formed via interaction of $H1^{nsp7}$ and $H3^{nsp7}$ with $H1^{nsp8}$ and $H2^{nsp8}$ at the dimer interface and $H1^{nsp7}$ and $H2^{nsp7}$ with $H1^{nsp8}$ at the heterotetramer interface. HDX-MS revealed protection from exchange in these regions, and XL-MS identified inter nsp7-nsp8 cross-links that were mapped to these protected regions. Furthermore, all interprotein cross-links were below the 26 Å distance limit when mapped to either the dimer or heterotetramer interface of this nsp7:nsp8 2:2 heterotetrameric structure. Additional protection to exchange at the C-terminus of nsp7, as well as a cross-link from $H1^{nsp7}$ to the C-terminus of nsp8, suggests that additional contact surfaces are likely to be present, involving the N- and C-termini of nsp7 and nsp8, that are not present in the PDB structures due to the use of truncated proteins for crystallization or lack of electron density.

As the nsp7:nsp8 complex is dynamic and interacts with other nsps including nsp12 to form the RTC responsible for RNA replication, we used our XL-MS results to interrogate changes in nsp7-nsp8 interaction dynamics in these larger complexes. Our results concluded that the nsp7:nsp8-I structural arrangement in the apo and pre- and post-translocated polymerase complexes is largely similar to the arrangement of the nsp7:nsp8 dimer in the heterotetramer. The nsp8 N-terminal is the only region slightly influenced by binding of nsp12 and/or RNA, as it appears to become stabilized by binding of RNA in the pretranslocation state and then destabilized in the post-translocation state as the RNA duplex is extended.

While this work was under review and publicly available in the preprint server BioRxiv, additional studies have been published or made available in BioRxiv to support our conclusion that the CoV-2 nsp7:np8 heterodimer dissociates into stable dimers rather than associating into a hexadecamer like CoV nsp7:nsp8. Jochheim and co-workers have detected (in reanalyzing their previous RdRp cryo-EM data sets) a "dimeric RdRp", constituted by two copies of nsp12:nsp7:nsp8 "monomers" interacting via an nsp7:nsp7 interface.[45] They conclude that formation of this complex is only possible through the dissociation of the heterotetrameric nsp7:nsp8 complex. Overall, the previous work, our results, and the recent publication from Krichel and co-workers[40] suggest that the nsp7:nsp8 heterotetrameric arrangement is the predominant form and most biologically relevant as it may facilitate its dissociation into a dimer that can bind to nsp12 in the RTC without significant alteration of nsp7-nsp8 interactions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.1c00086.

Additional experimental details for differential HDX-MS experiments, SDS-PAGE validation of all cross-linked samples, mapping of all observed cross-links in the nsp8 alone sample and dead-end cross-links in the nsp7 alone sample, and all cross-link distances mapped to available nsp7:nsp8 structures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Eddy Arnold − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0003-2612-9622; Email: arnold@cabm.rutgers.edu

Patrick R. Griffin − *Department of Molecular Medicine and Department of Integrative Structural and Computational Biology, The Scripps Research Institute, Jupiter, Florida 33458, United States;* orcid.org/0000-0002-3404-690X; Email: pgriffin@scripps.edu

Authors

Valentine V. Courouble − *Department of Molecular Medicine, The Scripps Research Institute, Jupiter, Florida 33458, United States*

Sanjay Kumar Dey − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; Dr. B.R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi 110007, India*

Ruchi Yadav − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States*

Jennifer Timm − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology,*

Rutgers University, Piscataway, New Jersey 08854, United States

Jerry Joe E. K. Harrison − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; Department of Chemistry, School of Physical and Mathematical Sciences, University of Ghana, Accra, Ghana*

Francesc X. Ruiz − *Center for Advanced Biotechnology & Medicine, and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.1c00086

Author Contributions

⊥V.V.C., S.K.D., and R.Y. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco. The E.A. and P.R.G. laboratories are grateful for support from NIH U54 AI150472, and the E.A. laboratory is also grateful for financial support from a Rutgers Center for COVID-19 Response and Pandemic Preparedness COVID research award. We are grateful to Vikas Nanda and Paul Falkowski for providing financial support for Jennifer Timm's efforts.

■ REFERENCES

(1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. China Novel Coronavirus, I. Research, T. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727−733.

(2) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497−506.

(3) Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; Sheng, J.; Quan, L.; Xia, Z.; Tan, W.; Cheng, G.; Jiang, T. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **2020**, *27*, 325−328.

(4) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **2020**, *26*, 450−452.

(5) V'Kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021**, *19*, 155.

(6) 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. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**, *368*, 1499−1504.

(7) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020**, *584*, 154−156.

(8) Wang, Q.; Wu, J.; Wang, H.; Gao, Y.; Liu, Q.; Mu, A.; Ji, W.; Yan, L.; Zhu, Y.; Zhu, C.; Fang, X.; Yang, X.; Huang, Y.; Gao, H.; Liu, F.; Ge, J.; Sun, Q.; Yang, X.; Xu, W.; Liu, Z.; Yang, H.; Lou, Z.; Jiang, B.; Guddat, L. W.; Gong, P.; Rao, Z. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **2020**, *182*, 417−428.

(9) Peng, Q.; Peng, R.; Yuan, B.; Zhao, J.; Wang, M.; Wang, X.; Wang, Q.; Sun, Y.; Fan, Z.; Qi, J.; Gao, G. F.; Shi, Y. Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell Rep.* **2020**, *31*, 107774.

(10) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; Ge, J.; Zheng, L.; Zhang, Y.; Wang, H.; Zhu, Y.; Zhu, C.; Hu, T.; Hua, T.; Zhang, B.; Yang, X.; Li, J.; Yang, H.; Liu, Z.; Xu, W.; Guddat, L. W.; Wang, Q.; Lou, Z.; Rao, Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020**, *368*, 779−782.

(11) Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Hobartner, C.; Cramer, P. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **2021**, *12*, 279.

(12) Shi, W.; Chen, M.; Yang, Y.; Zhou, W.; Chen, S.; Yang, Y.; Hu, Y.; Liu, B. A dynamic regulatory interface on SARS-CoV-2 RNA polymerase. *bioRxiv*, 2020. DOI: 10.1101/2020.07.30.229187.

(13) Konkolova, E.; Klima, M.; Nencka, R.; Boura, E. Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* **2020**, *211*, 107548.

(14) Wilamowski, M.; Kim, Y.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Michalska, K.; Joachimiak, A. The 1.5 A Crystal Structure of the Co-factor Complex of NSP7 and the C-terminal Domain of NSP8 from SARS CoV-2; *RCSB PDB* **2020**. DOI: 10.2210/pdb6XIP/pdb

(15) Kim, Y.; Wilamowski, M.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Michalska, K.; Joachimiak, A. The 1.95 A Crystal Structure of the Co-factor Complex of NSP7 and the C-terminal Domain of NSP8 from SARS CoV-2; *RCSB PDB* **2020**. DOI: 10.2210/pdb6WQD/pdb

(16) Yan, L. M.; Ge, J.; Lou, Z. Y.; Rao, Z. H. Crystal structure of 2019-nCoV nsp7-nsp8c complex; *RCSB PDB* **2020**. DOI: 10.2210/pdb6M5I/pdb

(17) Biswal, M.; Hai, R.; Song, J. Crystal Structure of SARS-CoV-2 NSP7-NSP8 complex; *RCSB PDB* **2020**. DOI: 10.2210/pdb7JLT/pdb

(18) Wilamowski, M.; Kim, Y.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Michalska, K.; Joachimiak, A. Crystal structure of the co-factor complex of NSP7 and the C-terminal domain of NSP8 from SARS CoV-2; *RCSB PDB* **2020**. DOI: 10.2210/pdb6WIQ/pdb

(19) Wilamowski, M.; Kim, Y.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Michalska, K.; Joachimiak, A. Crystal Structure of the Second Form of the Co-factor Complex of NSP7 and the C-terminal Domain of NSP8 from SARS CoV-2; *RCSB PDB* **2020**. DOI: 10.2210/pdb6WTC/pdb

(20) Zhai, Y.; Sun, F.; Li, X.; Pang, H.; Xu, X.; Bartlam, M.; Rao, Z. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **2005**, *12*, 980−986.

(21) Xiao, Y.; Ma, Q.; Restle, T.; Shang, W.; Svergun, D. I.; Ponnusamy, R.; Sczakiel, G.; Hilgenfeld, R. Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J. Virol.* **2012**, *86*, 4444−4454.

(22) Krichel, B.; Falke, S.; Hilgenfeld, R.; Redecke, L.; Uetrecht, C. Processing of the SARS-CoV pp1a/ab nsp7−10 region. *Biochem. J.* **2020**, *477*, 1009−1019.

(23) Liu, F.; Lossl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **2017**, *8*, 15473.

(24) Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **2015**, *12*, 1179−1184.
(25) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966−968.
(26) Chen, Z.; Chen, Z.; Rappsilber, J. A generic solution for quantifying cross-linked peptides using software Skyline. *Protocol Exchange*, 2018. DOI: 10.1038/protex.2018.001
(27) Chen, Z. A.; Rappsilber, J. Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat. Protoc.* **2019**, *14*, 171−201.
(28) Choi, M.; Chang, C. Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **2014**, *30*, 2524−2526.
(29) Combe, C. W.; Fischer, L.; Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell Proteomics.* **2015**, *14*, 1137−1147.
(30) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Perez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442−D450.
(31) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; He, Y.; Hendrickson, C. L.; Marshall, A. G.; Griffin, P. R. Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2006**, *78*, 1005−1014.
(32) Zhang, Z.; Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein science: a publication of the Protein Society.* **1993**, *2*, 522−531.
(33) Pascal, B. D.; Willis, S.; Lauer, J. L.; Landgraf, R. R.; West, G. M.; Marciano, D.; Novick, S.; Goswami, D.; Chalmers, M. J.; Griffin, P. R. HDX workbench: software for the analysis of H/D exchange MS data. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1512−1521.
(34) Keppel, T. R.; Weis, D. D. Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 547−554.
(35) Li, J.; Guo, M.; Tian, X.; Wang, X.; Yang, X.; Wu, P.; Liu, C.; Xiao, Z.; Qu, Y.; Yin, Y.; Wang, C.; Zhang, Y.; Zhu, Z.; Liu, Z.; Peng, C.; Zhu, T.; Liang, Q. Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis. *Med. (N Y).* **2021**, *2*, 99−112 e117.
(36) Xu, W., Pei, G., Liu, H., Wang, J., Li, P.: Extensive High-Order Complexes within SARS-CoV-2 Proteome Revealed by Compartmentalization-Aided Interaction Screening. *bioRxiv*, 2020. DOI: 10.1101/2020.12.26.424422
(37) Kao, A.; Chiu, C. L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell Proteomics.* **2011**, *10*, 2170.
(38) Kosinski, J.; von Appen, A.; Ori, A.; Karius, K.; Muller, C. W.; Beck, M. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **2015**, *189*, 177−183.
(39) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(40) Krichel, B.; Bylapudi, G.; Schmidt, C.; Blanchet, C.; Schubert, R.; Brings, L.; Koehler, M.; Zenobi, R.; Svergun, D.; Lorenzen, K.; Madhugiri, R.; Ziebuhr, J.; Uetrecht, C. Hallmarks of Alpha- and Betacoronavirus non-structural protein 7 + 8 complexes. *Sci. Adv.* **2021**, *7*, eabf1004.
(41) Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **2018**, *27*, 245−258.
(42) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(43) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725−738.
(44) Orban-Nemeth, Z.; Beveridge, R.; Hollenstein, D. M.; Rampler, E.; Stranzl, T.; Hudecz, O.; Doblmann, J.; Schlogelhofer, P.; Mechtler, K. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **2018**, *13*, 478−494.
(45) Jochheim, F. A.; Tegunov, D.; Hillen, H. S.; Schmitzova, J.; Kokic, G.; Dienemann, C.; Cramer, P. Dimeric form of SARS-CoV-2 polymerase. *bioRxiv*, 2021. DOI: 10.1101/2021.03.23.436644