CORONAVIRUS

Hallmarks of *Alpha-* and *Betacoronavirus* non-structural protein 7+8 complexes

Boris Krichel[1], Ganesh Bylapudi[2], Christina Schmidt[3], Clement Blanchet[4], Robin Schubert[3], Lea Brings[3], Martin Koehler[5], Renato Zenobi[5], Dmitri Svergun[4], Kristina Lorenzen[3], Ramakanth Madhugiri[2], John Ziebuhr[2], Charlotte Uetrecht[1,3]*

Coronaviruses infect many different species including humans. The last two decades have seen three zoonotic coronaviruses, with SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) causing a pandemic in 2020. Coronaviral non-structural proteins (nsps) form the replication-transcription complex (RTC). Nsp7 and nsp8 interact with and regulate the RNA-dependent RNA-polymerase and other enzymes in the RTC. However, the structural plasticity of nsp7+8 complexes has been under debate. Here, we present the framework of nsp7+8 complex stoichiometry and topology based on native mass spectrometry and complementary biophysical techniques of nsp7+8 complexes from seven coronaviruses in the genera *Alpha-* and *Betacoronavirus* including SARS-CoV-2. Their complexes cluster into three groups, which systematically form either heterotrimers or heterotetramers or both, exhibiting distinct topologies. Moreover, even at high protein concentrations, SARS-CoV-2 nsp7+8 consists primarily of heterotetramers. From these results, the different assembly paths can be pinpointed to specific residues and an assembly model proposed.

INTRODUCTION

Seven coronaviruses (CoVs) from six coronavirus species are known to cause infections in humans. While four of these viruses (HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1) predominantly cause seasonal outbreaks of (upper) respiratory tract infections with mild disease symptoms in most cases, three other CoVs (SARS-CoV, MERS-CoV, and SARS-CoV-2) of recent zoonotic origin are associated with lower respiratory tract disease including acute respiratory distress syndrome (ARDS) (*1–3*). SARS-CoV-2 is the etiologic agent of COVID-19 (coronavirus disease 2019), a respiratory disease with a wide spectrum of clinical presentations and outcomes (*4*). First detected in December 2019, it quickly became pandemic with numbers still growing (>100 million confirmed cases, >2,200,000 deaths, by early February 2021) (*5*). COVID-19 caused major perturbations of historical dimensions in politics, economics, and health care. Pets and domestic animals can also be infected by SARS-CoV-2 (*6*, *7*). Moreover, CoVs are important, widespread animal pathogens as illustrated by feline intestine peritonitis virus (FIPV) causing a severe and often fatal disease in cats (*8*) or porcine CoVs (*9*), such as transmissible-gastroenteritis virus (TGEV) or porcine epidemic diarrhea virus (PEDV), the latter causing massive outbreaks and economic losses in the swine industry.

The viral replication machinery is largely conserved across the different CoV species from the four currently recognized genera *Alpha-*, *Beta-*, *Gamma-*, and *Deltacoronavirus* (subfamily *Orthocoronavirinae*, family *Coronaviridae*) (*10*). The key components are generally referred to as non-structural proteins (nsps) and encoded by the viral replicase genes (ORFs 1a and 1b) and translated as parts of the replicase polyproteins pp1a (nsp1-11) or pp1ab (nsp1-16). Translation

of the ORF1b-encoded C-terminal part of pp1ab requires a ribosomal (−1)-frameshift immediately upstream of the ORF1a stop codon. Two proteases called PL^{pro} (one or two protease domains in nsp3) and M^{pro} (also called 3CL^{pro} or nsp5) facilitate polyprotein processing into 16 (sometimes 15) mature nsps. Most of these nsps form a membrane-anchored, highly dynamic protein-RNA machinery, the replication-transcription complex (RTC), which mediates replication of the ~30-kb single-strand (+)-sense RNA genome and production of subgenomic mRNAs (Fig. 1A) (*10*, *11*).

The main CoV-RTC building block is the fastest known RNA-dependent RNA-polymerase (RdRp) residing in the nsp12 C-terminal domain (*12*). For RdRp activity, nsp12 requires binding to its cofactors nsp7 and nsp8 (*13*). Recently, high-resolution structures illuminated two binding sites at nsp12: one for an nsp7+8 (1:1) heterodimer and a second for a single nsp8 (*14–18*). For in vitro RdRp activity assays, different methods were used to assemble the polymerase complex (*12*, *19*, *20*). So far, the highest processivity in vitro was obtained by mixing nsp12 with an nsp7L8 fusion protein containing a flexible linker between the nsp7 and nsp8 domains.

Recently, we reported that SARS-CoV nsp7 and nsp8 form a heterotetramer (2:2) in solution, in which nsp7 subunits have no self-interaction and rather sandwich an nsp8 scaffold with putative head-to-tail interactions (*21*). Current knowledge of full-length nsp7+8 complexes is mainly based on two X-ray crystal structures, each of which displays a different quaternary conformation. First, a SARS-CoV nsp7+8 (8:8) hexadecamer is assembled from four (2:2) heterotetramers with similar topologies but two distinct conformations, T1 and T2, which are both consistent with our in-solution structure (Fig. 1B) (*21*, *22*). Second, in a feline CoV (FIPV) nsp7+8 (2:1) heterotrimer, nsp8 is associated to two nsp7 molecules that self-interact (Fig. 1C) (*23*). Moreover, structures of SARS-CoV and SARS-CoV-2 with N terminally truncated forms of nsp8, thus lacking the self-interaction domain, revealed heterotetrameric nsp7+8 complexes around an nsp7 scaffold (*24*, *25*).

Current knowledge of CoV nsp7+8 complexes suggests a remarkable architectural plasticity but is unsupportive of deducing common
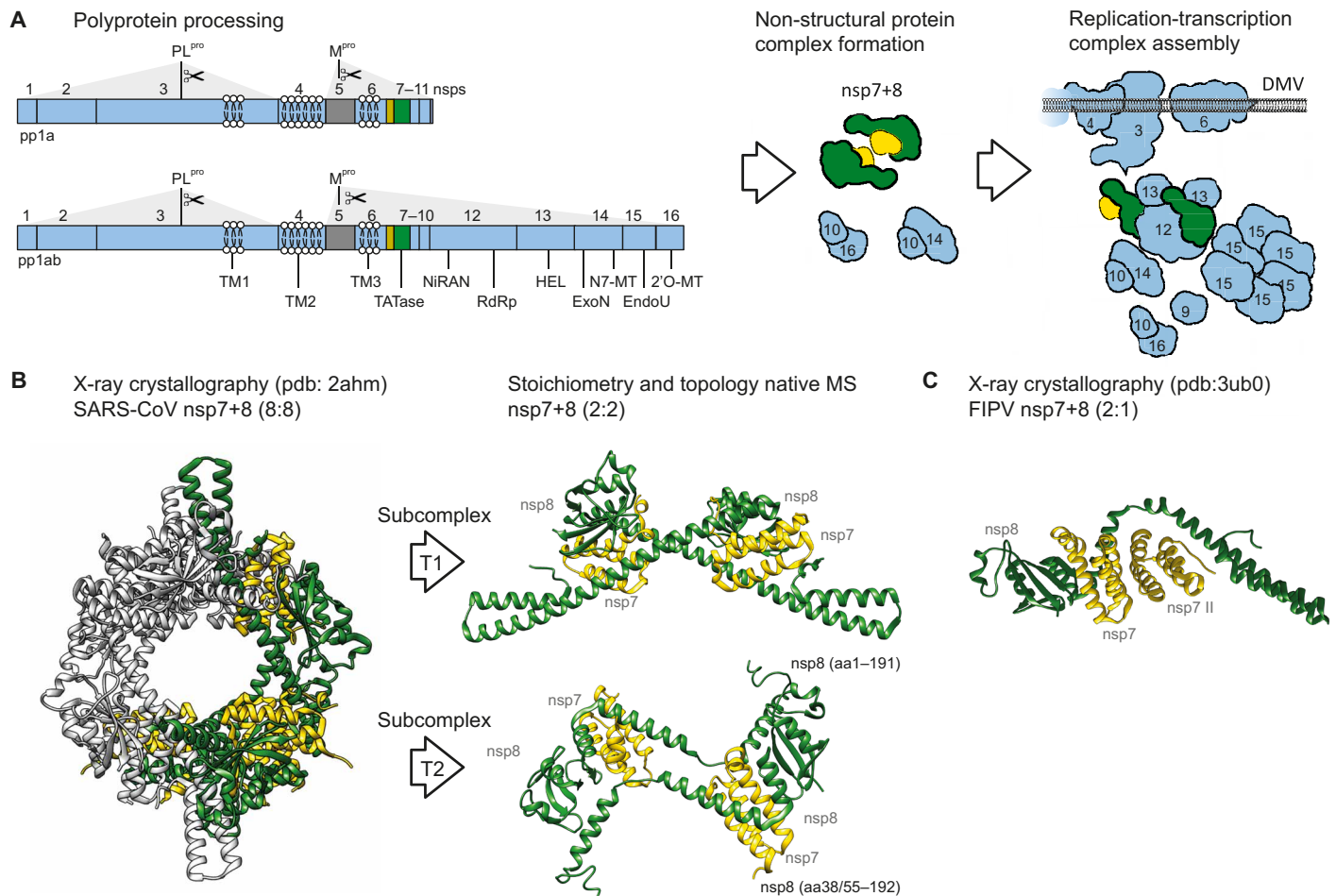
Fig. 1. Candidate structures in agreement with the observed stoichiometries and topologies. (**A**) Replicase polyproteins pp1a (nsp1-11) or pp1ab (nsp1-16) undergo processing by two internal proteases and subsequently release nsps that assemble into the CoV replication transcription complex residing in double-membrane vesicles (DMVs) within the cell. (**B**) For the full-length heterotetramer, an isolated structure does not exist. However, from the larger SARS-CoV nsp7+8 hexadecamer (*22*) [Protein Data Bank (pdb) 2ahm], two conformer subcomplexes of nsp7+8 (2:2), T1 and T2, can be extracted. Both conformers constitute a head-to-tail interaction of two heterodimers by an nsp8:nsp8 interface. Notably, nsp8 in T1 is more extended, containing an almost full-length amino acid sequence (2 to 193), while in T2, the nsp8 N-terminal 35 to 55 residues are unresolved. (**C**) For the trimeric complexes, the only deposited structure is FIPV nsp7+8 (2:1) trimer (*23*) (pdb 3ub0).

principles of complex formation. Moreover, it is unknown if the quaternary structure of nsp7+8 is conserved within a given CoV species or between genera. To fill these knowledge gaps, we analyzed nsp7+8 complexes derived from seven viruses of the *Alpha-* and *Betacoronavirus* genera, including a range of human CoV, namely, SARS-CoV, SARS-CoV-2, MERS-CoV, and HCoV-229E. We used native mass spectrometry (MS) to illustrate the landscape of nsp7+8 complexes in vacuo, collision-induced dissociation tandem MS (CID-MS/MS) to reconstruct complex topology, and complementary biophysical methods, such as gel electrophoresis, alternative MS, and scattering techniques, to verify the results (*26*, *27*). Our findings reveal distinct sets of nsp7+8 complexes for the different CoV species. The results hint at the properties that lead to complex heterogeneity and suggest common principles of complex formation based on two conserved binding sites.

RESULTS

Native MS illustrates the landscape of nsp7+8 complexes

To ensure authentic nsp7 and nsp8 N and C termini, which allow for optimal nsp7+nsp8 complex assembly, the proteins are expressed as

nsp7-8-His$_6$ polyprotein precursors (table S1). The precursors can be cleaved between nsp7/nsp8, and nsp8/linker to His$_6$ by their cognate protease M^{pro} so that no additional amino acid residues remain on nsp7 and nsp8 (Fig. 2A; table S2).

Native MS provides an overview of mass species in solution, while CID-MS/MS confirms the stoichiometry of protein complexes. Distinct oligomerization patterns of nsp7+8 (1:1) heterodimers, (2:1) heterotrimers, and (2:2) heterotetramers in the different CoVs allowed us to categorize their nsp7+8 complexes into three groups (Fig. 3). SARS-CoV and SARS-CoV-2 (species *Severe acute respiratory syndrome-related coronavirus*, genus *Betacoronavirus*) represent nsp7+8 group A complex formation pattern (Fig. 3, A and B). Consistent with our previous work, SARS-CoV nsp7+8 complexes exist primarily as a heterotetramer comprising two copies of each nsp7 and nsp8 (2:2) (*21*). As expected, SARS-CoV-2 nsp7+8 form identical (2:2) complexes given the high sequence identity of 97.5% in the nsp7-8 region (table S3). Next, relative peak intensities in native MS of nsp7+8 complexes are converted in a semiquantitative analysis into abundances of complex species (*28*). The heterodimer (2 to 4%) is much less abundant than the heterotetramer (96 to 98%), suggesting
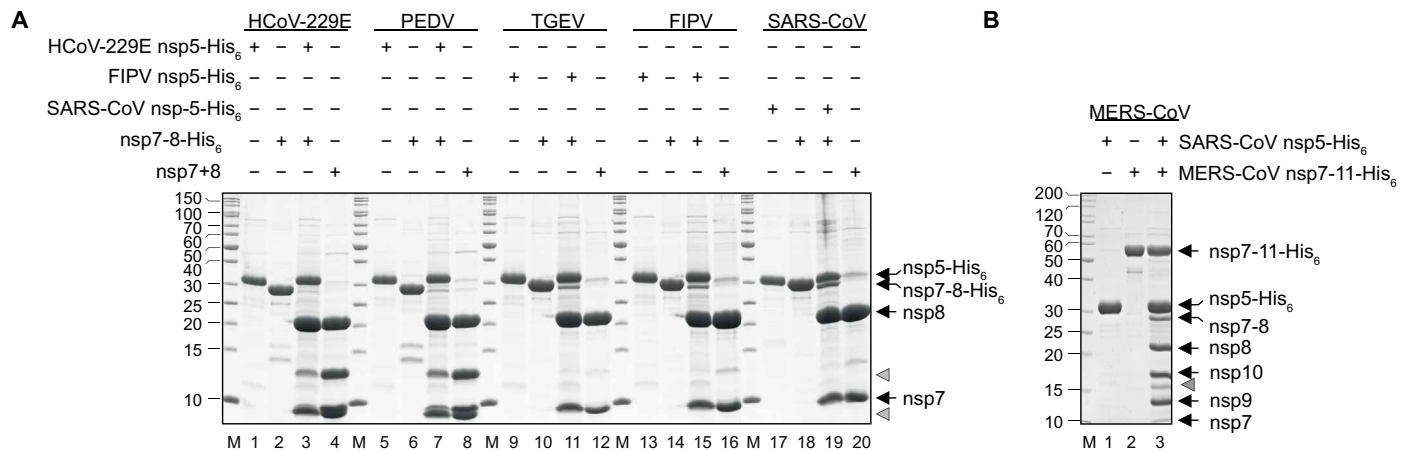
Fig. 2. M^{pro}-mediated processing of precursor protein constructs. SDS–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of M^{pro} (nsp5)-mediated processing and generation of CoV nsp7+8 complexes with authentic N and C termini from (A) polyprotein precursors nsp7-8 and (B) nsp7-11. (**A**) SDS-PAGE showing the purified M^{pro} (nsp5-His$_6$): lanes 1, 5, 9, 13, and 17; nsp7-8-His$_6$: lanes 2, 6, 10, 14, and 18; M^{pro}-mediated cleavage reaction: lanes 3, 7, 11, 15, and 19; enriched nsp7+8 complexes: lanes 4, 8, 12, 16, and 20. (**B**) SDS-PAGE showing the purified M^{pro} (nsp5-His$_6$): lane 1; nsp7-8-9-10-11-His$_6$: lane 2; M^{pro}-mediated cleavage reaction: lane 3. Lane M, marker proteins with molecular masses in kilodaltons indicated to the left. Black arrows on the right indicate the identities of proteins generated from precursor proteins by M^{pro}-mediated cleavage. Gray arrowheads indicate aberrant in vitro cleavage products of nsp8 as observed previously for SARS-CoV (*24*). +/− indicate the presence or absence of the respective proteins.

high affinity and, hence, efficient conversion of heterodimeric intermediates into heterotetramers. Hence, group A only forms two types of nsp7+8 complexes, heterodimers (1:1) and heterotetramers (2:2), with the latter clearly being predominant.

In FIPV and TGEV from the species *Alphacoronavirus 1*, genus *Alphacoronavirus*, nsp7 and nsp8 proteins share a sequence identity of 93.9% (table S3). Their nsp7+8 complexes are assigned to group B forming predominantly nsp7+8 (2:1) heterotrimers (83%) and, to a lesser extent, heterodimers (1:1) (~17%) (Fig. 3, C and D). An nsp7+8 (2:1) heterotrimeric structure has previously been reported for FIPV but not for TGEV or any other CoV. The association of a single nsp8 with two nsp7 indicates that group B nsp7+8 complexes lack the ability to form tetramers around an nsp8 scaffold.

The third oligomerization pattern is observed for nsp7+8 of HCoV-229E and PEDV, which represent different species in the genus *Alphacoronavirus*. They share only 70.9% sequence identity in the nsp7-8 region and even less (42 to 62%) with the other CoV species examined (table S3). PEDV and HCoV-229E nsp7+8 form three major types of oligomers with slightly different efficiencies: heterodimers (1:1), heterotrimers (2:1), and heterotetramers (2:2) (HCoV-229E: 20, 12, and 69%; PEDV: 52, 6, and 42%, respectively) (Fig. 3, E and F). By forming both heterotrimers and heterotetramers, these complexes combine properties described above for groups A and B and are hence categorized into a separate group named accordingly AB. This begs the question whether assembly pathways and structures of heterotetramers in groups A and AB are similar. Either two heterodimers form a heterotetramer around an nsp8 scaffold as in group A (*21*) or alternatively the heterotrimer recruits another nsp8 subunit to the complex, thus using an nsp7 core (*23*). The latter pathway has recently been reported for SARS-CoV-2 nsp7+8 heterotetramers containing N-terminally truncated nsp8 (*25*).

In addition, nsp7+8 complexation after M^{pro}-mediated cleavage of a MERS-CoV nsp7-11-His$_6$ precursor is compared. Initial attempts to cleave nsp7-8–only constructs failed. However, with the larger precursor nsp7-11-His$_6$ (comprising the domains nsp7, nsp8, nsp9, nsp10, and nsp11), cleavage was successful (Fig. 2B). Proteolytic

processing of this polyprotein precursor leads to cleavage intermediates (Fig. 3G). Such processing intermediates have been proposed to occur intracellularly and to function distinctly from the individual nsps in, e.g., regulation of RTC assembly and viral RNA synthesis (*29*). Here, signal intensities of these intermediates provide insights into the processing sequence. Because of the small size of nsp11, the nsp10/11 cleavage site is expected to have a high accessibility. However, a relatively large fraction of nsp10/11 remains uncleaved as indicated by the dominant intermediate nsp10-11-His$_6$. Therefore, slow cleavage and prolonged presence of an nsp10-11 intermediate may have functional implications warranting further studies. Notably, in many CoV polyproteins, the nsp10/11 and/or nsp10/12 cleavage sites contain replacements (Pro in MERS-CoV) of the canonical P2 Leu residue conserved throughout most M^{pro} cleavage sites, suggesting that slow or incomplete cleavage is beneficial for these particular sites. Moreover, this cleavage site has different C-terminal contexts in the two CoV replicase polyproteins, nsp10-11 in pp1a and nsp10-12 in pp1ab. While the structure of the small nsp11 (~1.5 kDa) is unknown, nsp12 is a large folded protein (~105 kDa), which potentially improves the accessibility of the nsp10/12 site for M^{pro}. Similar effects have been observed for the nsp8/9 cleavage site, which is efficiently cleaved in the protein but not in peptide substrates (*21, 30*). The question remains whether unprocessed nsp10-11 and/or nsp10-12 intermediates exist in virus-infected cells for prolonged times to fulfill specific functions. Other detected intermediates are nsp7-8-9 and nsp9-10 lacking nsp11-His$_6$. In particular, the nsp9-10 intermediate has not been identified in our analysis of SARS-CoV nsp7-10 processing, suggesting differences in the in vitro processing order between SARS-CoV and MERS-CoV.

MERS-CoV nsp7+8 forms heterodimers (1:1), heterotrimers (2:1), and heterotetramers (2:2) (73, 8, and 19%, respectively), thus demonstrating a group AB complexation pattern. However, we cannot confirm the heterotrimer (2:1) formation by CID-MS/MS owing to spectral complexity. Moreover, because of incomplete cleavage as is evident from the cleavage intermediates, signals assigned to the nsp7+8 heterodimer likely overlap with signals of unprocessed
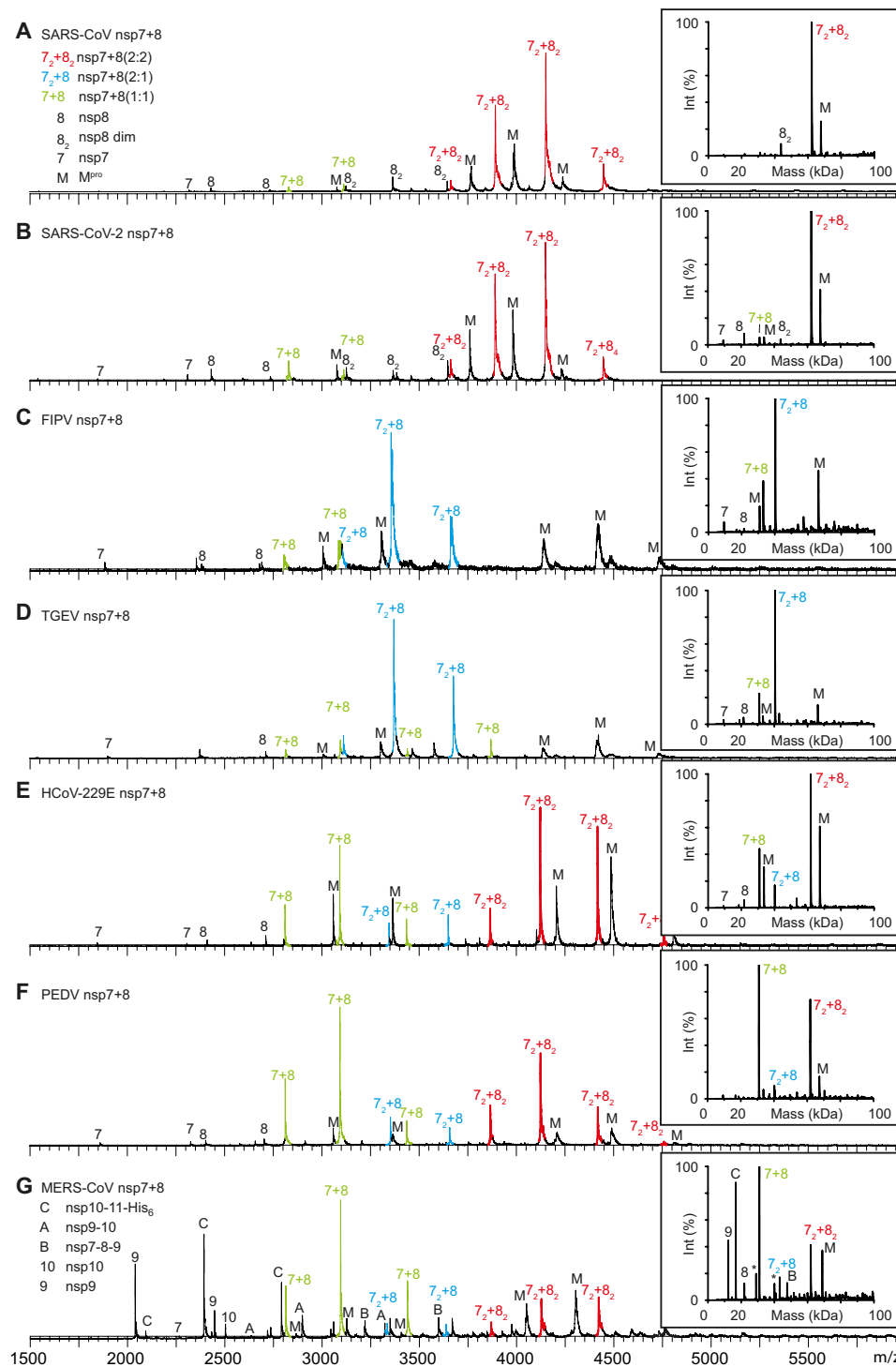
Fig. 3. Native MS of nsp7+8 complexes of seven CoVs representing five different CoV species. Representative mass spectra showing distinct nsp7+8 complexation patterns that were classified into the three groups A, B, and AB. Complex formation triggered by M^{pro} (M)–mediated cleavage of 15 μM nsp7-8-His$_6$ or MERS-CoV nsp7-11-His$_6$ precursors in 300 mM AmAc, 1 mM dithiothreitol (DTT) (pH 8.0). (**A**) SARS-CoV and (**B**) SARS-CoV-2 from group A forming nsp7+8 (2:2) heterotetramers (red), (**C**) FIPV and (**D**) TGEV from group B forming nsp7+8 (2:1) heterotrimers (blue), and (**E**) HCoV-229E and (**F**) PEDV from group AB forming both complex stoichiometries. (**G**) MERS-CoV, also from group AB, produced from an nsp7-11-His$_6$ precursor, additionally results in several processing intermediates that allow for an estimation of relative cleavage efficiencies at different cleavage sites. All groups form nsp7+8 (1:1) heterodimers as intermediate state (green).

nsp7-8. Thus, complete cleavage of nsp7-8 could shift the peak fractions from heterodimer to heterotrimer or heterotetramer.

Homodimerization of subunits and precursors

In the mass spectra of nsp7+8 complexes, monomers and homodimers of nsp7 and nsp8 are also observed. While nsp7 homodimers are identified for all seven CoV species tested, nsp8 homodimers are only detected for SARS-CoV and SARS-CoV-2, which belong to group A forming exclusively nsp7+8 heterotetramers putatively around the dimeric nsp8 scaffold observed for SARS-CoV (fig. S1) (21). Moreover, the oligomeric states of the different uncleaved nsp7-8 precursors are probed. Notably, precursors from group B CoVs are mostly monomeric, whereas precursors from group AB and A CoVs are in varying equilibria between monomers and dimers (Fig. 4). The different oligomerization propensities of precursors suggest that molecular interactions driving dimerization of nsp7-8 precursors
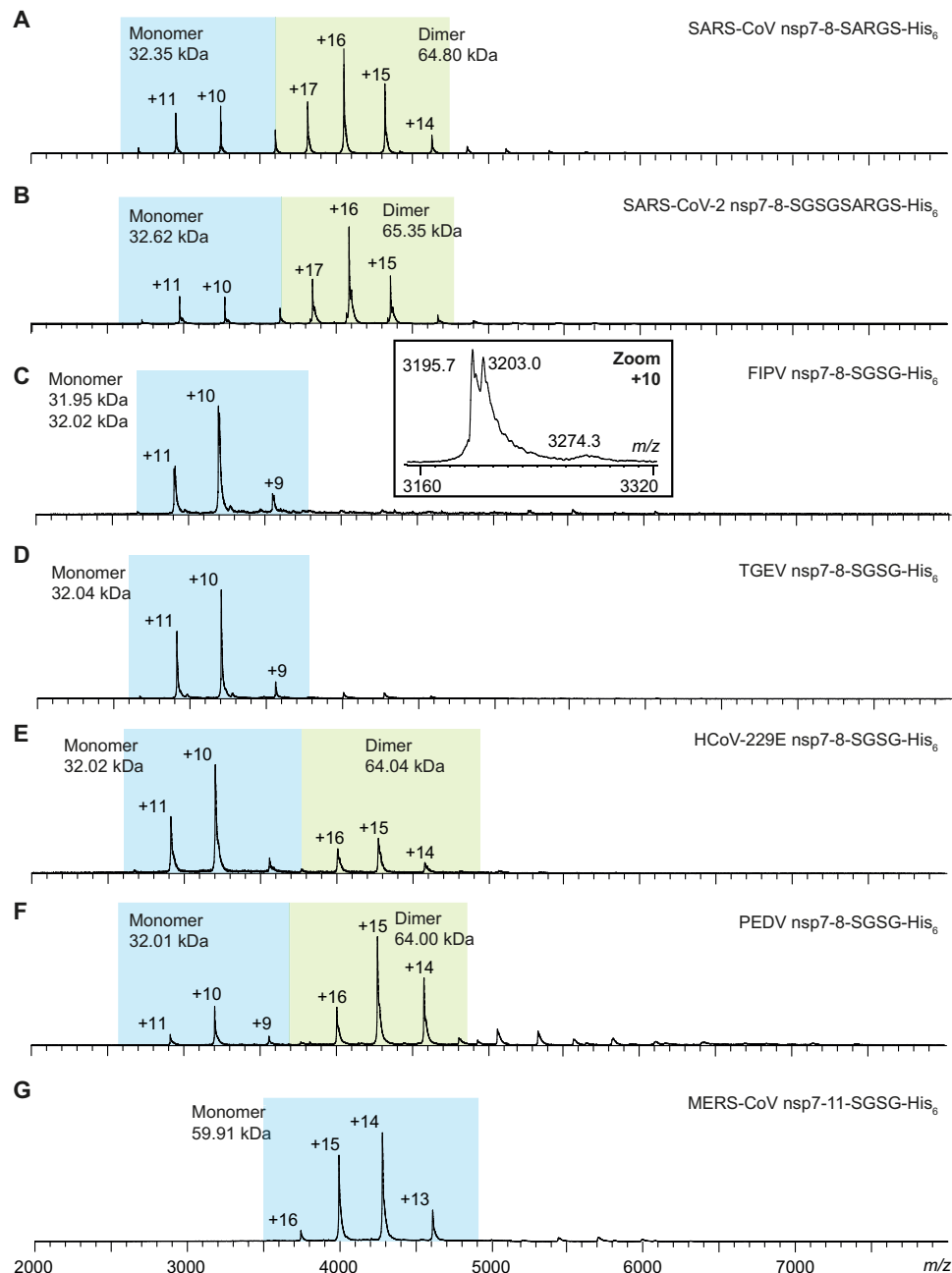

Fig. 4. Mass and oligomeric state of nsp7-8 precursors. Native MS of nsp7-8 precursors (**A** to **G**) sprayed at 18 μM from 300 mM AmAc (pH 8.0) and 1 mM DTT. Dominant charge envelope is highlighted (blue box). Labeled are charge states and molecular mass. (**C**) Inset shows mass heterogeneity in FIPV nsp7-8. The experimental molecular weight M_{exp} of the precursors agrees with the sequence-derived theoretical M_{theo} (table S2). Only FIPV nsp7-8 contained two mass species separated by ~110 Da. This heterogeneity was attributed to the precursor's central nsp8 domain following M^{pro} processing. Assignment to an amino acid variation failed but potentially was the result of codon heterogeneity in the plasmid. Nevertheless, both forms behaved identically and we refrained from further optimization.

could critically affect subsequent nsp7+8 oligomerization. The only exclusion being MERS-CoV nsp7-11-His$_6$, which is in line with our previous findings (*21*), in which C-terminally extended SARS-CoV nsp7-9-His$_6$ and His$_6$-nsp7-10 polyprotein constructs were mainly monomeric, suggesting that the presence of the extra C-terminal sequence further destabilizes an already weak dimerization.

Collision-induced dissociation reveals complex topology
To deduce the complex topology in the different groups of nsp7+8 interaction patterns, we applied CID-MS/MS using successive subunit dissociations to dissect conserved interactions (Fig. 5; MS/MS spectra for all complexes in fig. S2). CID-MS/MS of the HCoV-229E nsp7+8 heterotetramer (2:2) reveals two dissociation pathways, in which, first, one nsp7 subunit is ejected from the complex followed by another nsp7 or an nsp8 subunit. After two consecutive losses, the product ions are nsp7+8 (1:1) and nsp8$_2$ dimers, providing evidence for specific subunit interfaces in the complex (Fig. 5, A and B). From these results, the complex topology is deduced as a heterotetramer

based on an nsp8$_2$ dimer scaffold, in which each nsp8 binds only one nsp7 subunit. Notably, this is identical to our previously reported SARS-CoV nsp7+8 heterotetramer (2:2) architecture (*21*). All nsp7+8 (2:2) heterotetramers of groups A and AB (SARS-CoV-2, SARS-CoV, PEDV, HCoV-229E, and MERS-CoV) resulted in similar dissociation pathways, subunit interfaces, and topology maps, suggesting that these structures are similar across these diverse CoVs.

Next, the dissociation pathway of the HCoV-229E nsp7+8 (2:1) heterotrimers is monitored in CID-MS/MS (Fig. 5, C and D). After ejection of one nsp7 or nsp8 subunit, product dimers of nsp7+8 (1:1) and nsp7$_2$ are detected, indicating specific subunit interfaces between nsp7:nsp8 and nsp7:nsp7. Again, similar dissociation pathways and subunit interfaces are found for group B and AB heterotrimers (FIPV, TGEV, HCoV-229E, and PEDV). Topological reconstructions reveal a heterotrimer forming a tripartite interaction between one nsp8 and two nsp7 subunits. These results agree with the reported x-ray structure of FIPV nsp7+8 (*23*) and indicate that heterotrimers of these
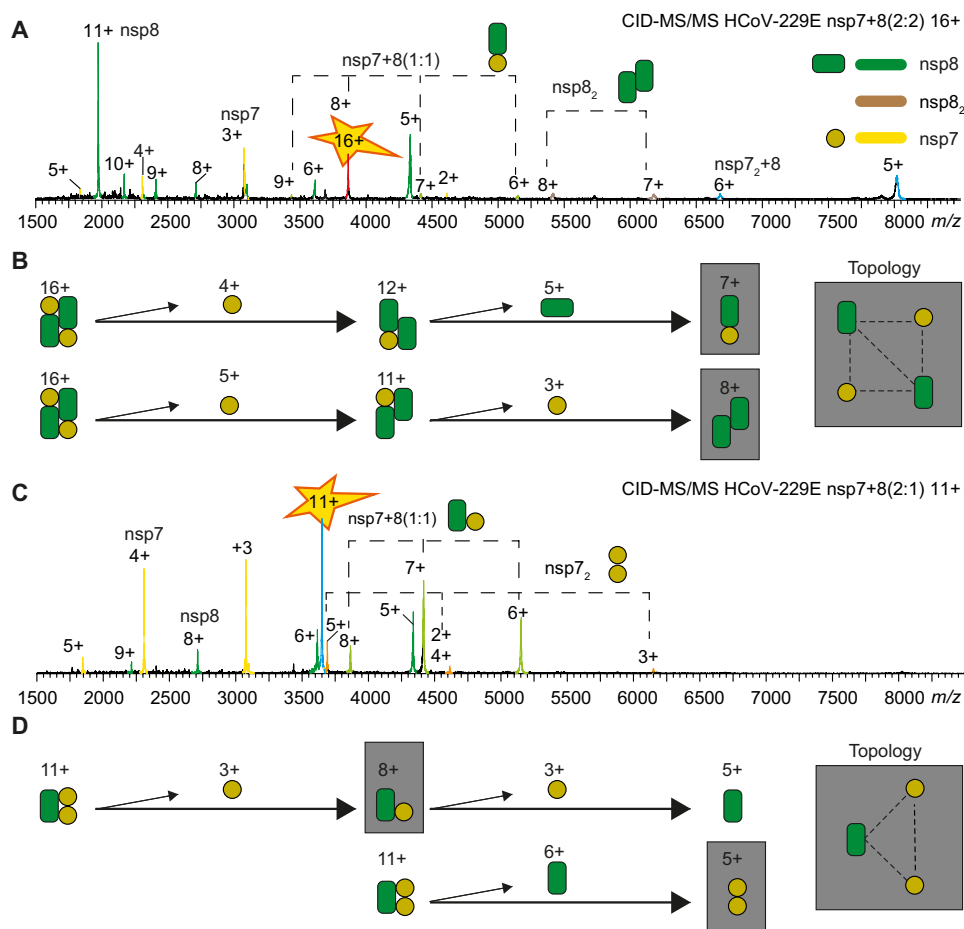

Fig. 5. Gas-phase dissociation reveals complex topology. CID-MS/MS product ion spectra (A and C) and dissociation pathways and topology maps (B and D) for HCoV-229E nsp7+8 (2:1) heterotrimers and (2:2) heterotetramers are shown. With increasing collisional voltage, protein complexes are successively stripped from their subunits revealing alternative dissociation pathways. The remaining dimeric species expose direct subunit interactions in the nsp7+8 complexes (gray boxes). Charge states are labeled. (**A** and **B**) The heterotetramers (2:2) undergo two consecutive losses, resulting in dimeric product ions of nsp7+8 (1:1) and nsp8$_2$. These products indicate that nsp7:nsp8 and nsp8:nsp8 have direct interfaces in heterotetramers. (**C** and **D**) HCoV-229E heterotrimers dissociate into the dimeric products nsp7+8 (1:1) and nsp7$_2$, indicating direct interfaces between nsp7:nsp8 and nsp7:nsp7 in heterotrimers. All CoV heterotrimers follow similar dissociation pathways; also, all CoV heterotetramers follow a common dissociation route, allowing a topological reconstruction of two distinct complex architectures (fig. S2).

CoV species have similar arrangements. In turn, this implies that heterotrimers and heterotetramers follow distinct assembly paths.

Chemical cross-linking confirms the formation of specific complexes

To further support the native MS results, which relies on spraying from volatile salt solutions [e.g., ammonium acetate (AmAc)], complementary methods compatible with conventional buffers supplemented with sodium chloride are applied. To provide additional evidence for specific nsp7+8 complex formation, the FIPV and HCoV-229E nsp7+8 complexes are stabilized via cross-linking with glutaraldehyde and subjected to XL-MALDI MS (cross-linking matrix-assisted laser desorption/ionization MS) (Fig. 6, A and B). Relative peak areas in the MALDI mass spectra are assigned to FIPV nsp7+8 heterodimer, heterotrimer, and heterotetramer (38, 45, and 17%, respectively, the latter being of similar intensity to unspecific neighboring peaks), and HCoV-229E nsp7+8 heterodimer, heterotrimer, and heterotetramer (35, 27, and 38%, respectively). The results suggest a higher abundance of nsp7+8 heterodimer and heterotrimer complexes in FIPV than in HCoV-229E, while HCoV-229E contains more heterotetramers. This largely agrees with the results from native MS. However, the MALDI mass spectra show high background of virtually all possible

nsp7+8 stoichiometries [<200,000 m/z (mass/charge ratio)], probably due to over–cross-linking with the rather unspecific glutaraldehyde.

To refine these results, nsp7+8 complexes are stabilized with the amine-specific cross-linker BS[3] and analyzed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 6C). Multiple stoichiometries are identified with a few prominent bands highlighting the main complexes generated. These bands are assigned to SARS-CoV nsp7+8 heterodimers and heterotetramers, FIPV nsp7+8 heterodimers and heterotrimers, and HCoV-22E nsp7+8 heterodimers, heterotrimers, and heterotetramers, providing additional support for the classification of nsp7+8 complexes into groups A (SARS-CoV), B (FIPV), and AB (HCoV-229E).

Light scattering provides insights into complexation at high protein concentrations

To test the stoichiometry at higher protein concentrations in solution, dynamic light scattering (DLS) of SARS-CoV-2 nsp7+8 from 1 to 15 mg/ml is performed (Fig. 7, A to C). No increase of the hydrodynamic radius (R_0) beyond the error occurs with increasing concentration. At the same time, the measured radii become more stable and fluctuate less, which suggests a shift toward higher complex homogeneity and a reduced fraction of free nsp7 and nsp8.
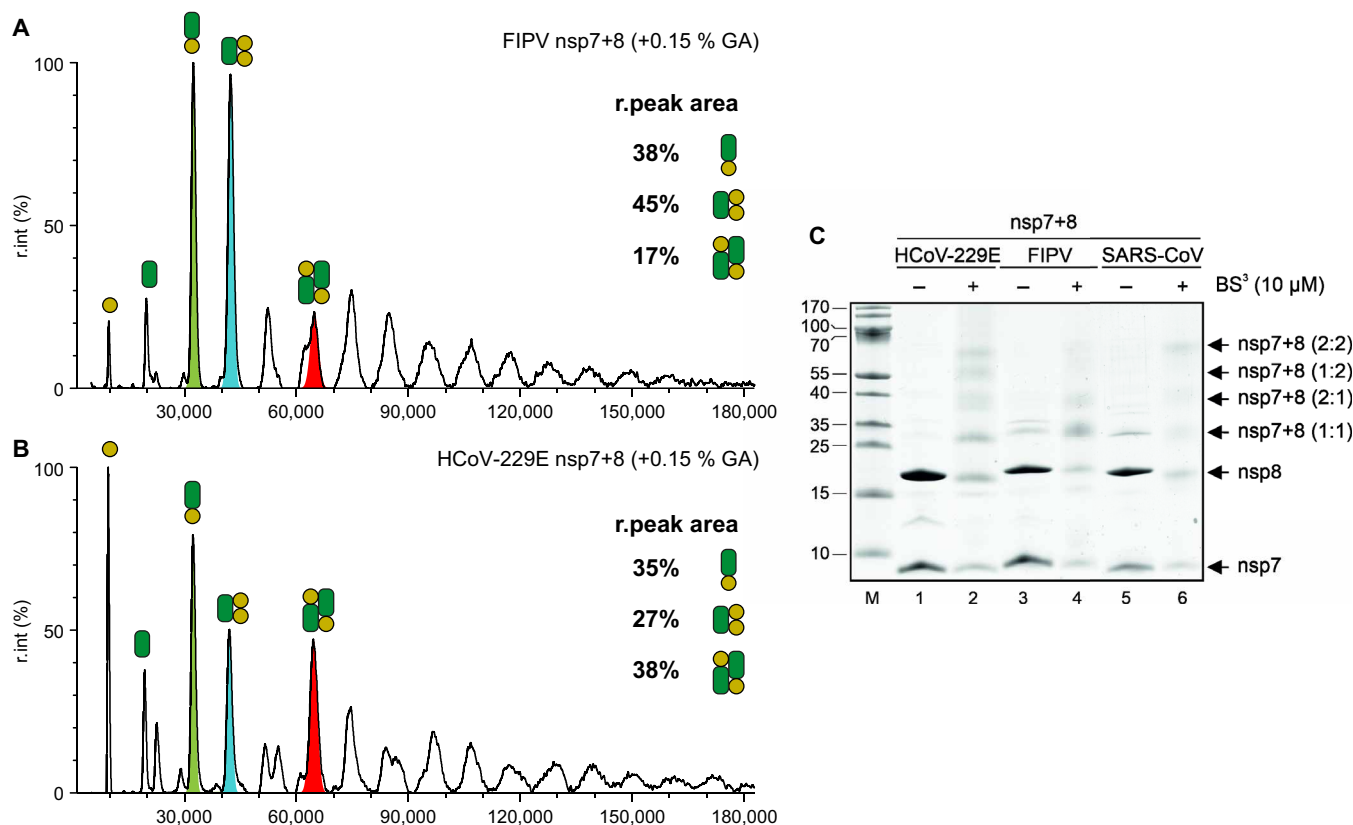

Fig. 6. Chemical cross-linking of prepurified nsp7+8 complexes shows species-specific complex formations. MALDI-MS of nsp7+8 complexes from FIPV (**A**) and HCoV-229E (**B**) stabilized with 0.15% glutaraldehyde (GA) for 25 min at 4°C. Mass spectra are background-subtracted and mass species of interest are labeled according to their stoichiometry with symbols for nsp7 (yellow) and nsp8 (green). Peak areas calculated from Gaussian fits of the respective peak for 1:1 heterodimers (green), 2:1 heterotrimers (blue), and 2:2 heterotetramers (red). Masses are higher in cross-linked samples owing to additional glutaraldehyde molecules. Mass spectra were not calibrated. Each spectrum shown generated from three MALDI spots. Signals above 50,000 m/z, except for the HCoV-229E heterotetramer, are low abundant and likely due to over–cross-linking. (**C**) SDS-PAGE analysis of chemically cross-linked HCoV-229E, FIPV, and SARS-CoV nsp7+8 complexes; 5 µg protein of nsp7+8 complexes cross-linked with 10 µM BS[3] at 37°C for 30 min. Lanes 1, 3, 5: nsp7+8 complexes not treated with BS[3] (−); lanes 2, 4, 6: nsp7+8 complexes treated with BS[3]. Lane M, marker proteins; molecular masses in kilodaltons are indicated to the left. Black arrows indicate the different oligomeric states of the nsp7+8 complexes obtained by cross-linking.
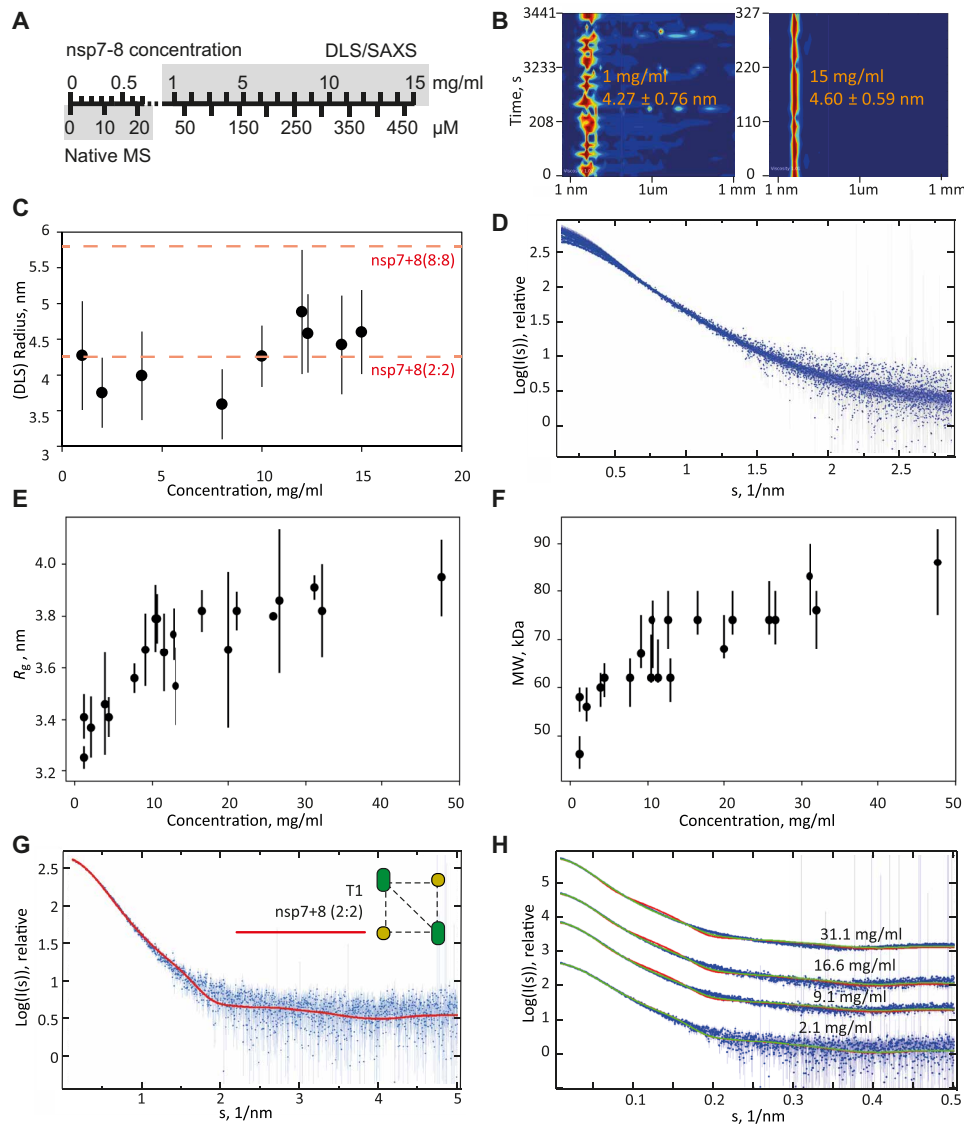
Fig. 7. DLS and SAXS reveal oligomeric state of SARS-CoV-2 nsp7+8 at higher protein concentrations. (**A**) Comparison of molar and mass concentration based on the molecular weight of the nsp7-8 polyprotein. Mass ranges for native MS and DLS/SAXS are indicated. (**B**) Two exemplary DLS plots (1 and 15 mg/ml) and (**C**) how the hydrodynamic radius (R_0) develops with increasing protein concentration. Theoretical radii ($R_{0,theo}$) of heterotetramer and hexadecamer candidate structures indicated (red dashed lines). Data points depicting R_0 with increasing complex concentrations. Error bars show SD. (**D**) SAXS curves collected at 1.2 to 47.7 mg/ml; (**E**) radius of gyration (R_g), error bars correspond to SD; and (**F**) molecular weight (MW), error bars correspond to credibility interval, estimated from the SAXS data. Both plots stabilize with increasing concentration on values that are in agreement with the $R_{0,theo}$ of the T1 nsp7+8 (2:2) heterotetramer. (**G**) Fit of the curve computed from T1 tetramer (red line) to the SAXS data collected at 1.2 mg/ml (blue dots with error bars). (**H**) Fits to all SAXS data. Experimental data (blue with experimental errors) for SARS-CoV-2 nsp7+8 complexes fitted with a mixture of T1 and hexadecamer (red), or T1 and dimer of T1 (green).

For SARS-CoV-2, no complex structure is available for full-length nsp7+8 proteins but, previously, a SARS-CoV nsp7+8 (8:8) has been reported using x-ray crystallography (*22*), where high protein concentrations are deployed. To relate the average experimental hydrodynamic radius ($R_{0,exp} = 4.25 \pm 0.61$ nm) to candidate structures, the theoretical hydrodynamic radius is calculated for the SARS-CoV nsp7+8 (8:8) hexadecamer ($R_{0,theo} = 5.80 \pm 0.29$ nm) and a subcomplex thereof, a putative nsp7+8 heterotetramer (2:2) in T1 conformation ($R_{0,theo} = 4.52 \pm 0.27$ nm) (Fig. 1B). This is the only model with full-length nsp8 that agrees with the stoichiometry and topology determined by native MS. At physiologically relevant concentrations from 1 to

10 mg/ml, the average experimental hydrodynamic radius remains relatively stable over the range of tested concentrations and agrees well with the theoretical hydrodynamic radius of the heterotetramer T1. Hence, a heterotetramer is likely the prevailing species in solution.

To underpin the DLS results, SAXS (small-angle x-ray scattering) data are collected on solutions of nsp7+8 at concentrations ranging from 1.2 to 47.7 mg/ml (Fig. 7D and table S4). The normalized SAXS intensities increase at low angles with increasing concentration, suggesting a change in the oligomeric equilibrium and a formation of larger oligomers. This trend is well illustrated by the evolution of the apparent radius of gyration and molecular weight of the solute
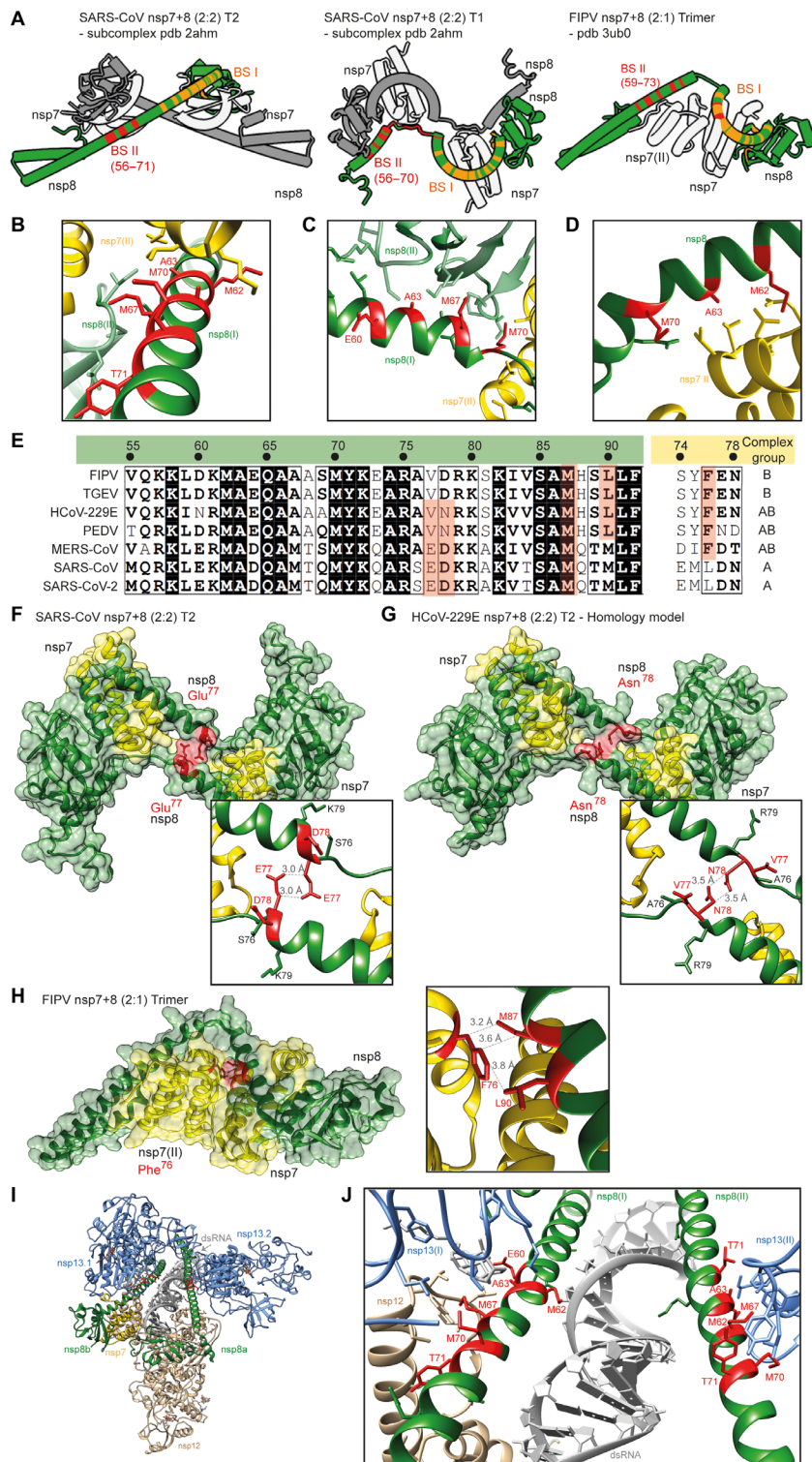
Fig. 8. Candidate structures and sequence conservation. Candidate structures for nsp7+8 heterotetramer and heterotrimer are chosen based on experimental stoichiometry and topology in solution and exhibit similar binding sites in nsp8, BS I and BS II. (**A**) Two conformers of SARS-CoV nsp7+8 (2:2) heterotetrameric subcomplexes (pdb 2ahm), T1 (left) and T2 (middle), and FIPV nsp7+8 (2:1) heterotrimer (right, pdb 3ub0). For BS II, residue (res) numbers are given (see also fig. S3). Candidate complexes involving similar conserved residues (red) in the nsp8 BS II are shown here for (**B**) SARS-CoV T1 and (**C**) T2 as well as (**D**) the FIPV heterotrimer. (**E**) Sequence alignment of BS II contact sites displayed for seven CoVs. Specific contact sites (red) exhibit sequence conservation well in line with the complexation groups determined by native MS. (**F**) Unique heterotetramer contact shown in SARS-CoV T2 (**G**) replaced by a neighboring amino acid in a homology model of HCoV-229E. (**H**) Unique heterotrimer contact shown in FIPV heterotrimer structure. Insets show magnifications with contact distances. (**I**) Cryo–electron microscopy (cryo-EM) of nsp7+8+12+13 (1:2:1:1) polymerase complex (pdb 6xez) (*18*). (**J**) Zoomed-in view of nsp8a and nsp8b BS II and its amino acids in contact with nsp12 thumb domain (brown), nsp13.1, and nsp13.2 (blue).

determined from the SAXS data (Fig. 7, E and F). The increase in the effective molecular weight, from about 50 to 80 kDa, suggests that the change in oligomeric state is limited and that the tetrameric state (MW_{theo}: 62 kDa) remains predominant in solution.

The SAXS data at low concentrations (<4 mg/ml) fit well the computed scattering from heterotetramer T1 but misfits appear at higher concentration (Fig. 7G, structure of T1 shown in Fig. 1B and the discrepancy χ^2 reported in table S5). Mixtures of heterotetramers and hexadecamers cannot be successfully fitted to the higher concentration data either. To further explore the oligomeric states of nsp7+8, a dimer of T1 is used to simultaneously fit the curves collected at different concentrations by a mixture of heterotetramers and heterooctamers. Reasonable fits to all SAXS data are obtained with volume fractions of heterooctamers growing from 0 to 0.52 with increasing concentration (Fig. 7H). On the basis of the flexibility of the molecule and the multiple possible binding sites between nsp7 and nsp8, it is expected that larger assemblies are observed at very high solute concentrations. The SAXS and DLS results provide evidence that the nsp7+8 (2:2) heterotetramer is the prevailing stoichiometry in solution at physiological concentrations (with volume fractions between 1.0 at 1 mg/ml and ~0.7 at 10 mg/ml).

Potential implications of sequence conservation on heterotrimer and heterotetramer formation

To extend this analysis, we select candidate structures in agreement with the stoichiometry and topology observed (Fig. 8A). For the heterotetramer, two conformers of nsp7+8 (2:2) subcomplexes, T1 and T2, of correct architecture can be extracted from the larger SARS-CoV nsp7+8 hexadecamer (22) [Protein Data Bank (pdb) 2ahm] (Fig. 1B). Both conformers constitute a head-to-tail interaction of two nsp7+8 heterodimers mediated by an nsp8-nsp8 interface. Notably, nsp8 in T1 is more extended, revealing an almost full-length amino acid sequence (2 to 193), while in T2, the nsp8 N-terminal 35 to 55 residues are unresolved. For the heterotrimeric complexes, the only deposited structure is the FIPV nsp7+8 (2:1) heterotrimer (23) (pdb 3ub0), which agrees well with our experimental topology (Fig. 1C).

To identify molecular determinants for heterotrimer or heterotetramer formation, the candidate structures are examined for molecular contacts (van der Waals radius −0.4 Å). The conservation of contact residues is evaluated in a sequence alignment to identify possible determinants of different stoichiometries (fig. S3). Notably, most amino acids lining subunit interfaces in heterodimers, heterotrimers, and heterotetramers are conserved. The interfaces in the candidate structures occupy two common structural portions of the nsp8 subunit. The first binding site (BS I) is located between the nsp8 head and shaft domain, responsible for binding of nsp7 (I) in heterodimer formation, as seen in all available high-resolution structures of nsp7+8 (22–25) and the polymerase complex (14–17). The second binding site (BS II) appears highly variable in terms of its binding partner and lies at the nsp8 elongated N terminus. One largely conserved motif (residues 60 to 70) is responsible for the main contacts in the entire candidate complexes selected on the basis of our data: nsp7+8 (2:2) T1 and T2 for the heterotetramer and nsp7+8 (2:1) for the heterotrimer. The respective side chains take positions on one side of the nsp8 α helix and have the ability to form interactions with either mainly nsp7 (partly nsp8) in the SARS-CoV nsp7+8 (2:2) heterotetramer T1, mainly nsp8 (partly nsp7) in the SARS-CoV nsp7+8 heterotetramer T2, or only nsp7 in the FIPV nsp7+8 (2:1) heterotrimer (Fig. 8, B to D). Because of its sequence conservation,

it is unlikely that this motif alone at BS II has a decisive impact for heterotrimer or heterotetramer formation.

Therefore, unique interactions could exist, which explain the shift in complex stoichiometry from heterotrimer to heterotetramer observed in the different CoVs categorized into groups A, AB, and B. Comparing unique and common amino acids in the candidate structures, relevant binding sites allowed us to suggest critical amino acids for the specific complex formation (Fig. 8E). Here, we identify a possibly heterotetramer stabilizing contact site in T2, where nsp8 Glu^{77} self-interacts with nsp8II Glu^{77}, which gives the complex density and compactness (Fig. 8F). This residue is only present in nsp8 of SARS-CoV and SARS-CoV-2 from group A and MERS-CoV of group AB. However, homology models suggest that in the other tetramer-forming complexes of group AB, HCoV-229E and PEDV, nsp8 Asn^{78} could partially replace this interaction (Fig. 8G). This is different in group B viruses, forming only heterotrimers, where residues at these positions are nsp8 Val^{77} and Asp^{78}, with the Asp^{78} possibly being solvent-exposed and hence unable to replace this interaction. Furthermore, we also identify a contact site possibly stabilizing the heterotrimer in the crystal structure of the FIPV nsp7+8 (2:1), which reveals that a second subunit of nsp7 (nsp7II) is locked via Phe^{76} to nsp8 (Fig. 8H). This residue is uniquely conserved among trimer-forming complexes of groups B and AB but replaced by nsp7 Leu^{76} in the strictly heterotetramer-forming group A.

These findings are compared to the recently released structure of the polymerase complex (pdb 6xez, Fig. 8I), comprising nsp7+8+12+13 (1:2:1:2) (18). The residues potentially responsible for a shift in quaternary structure, nsp8 Glu^{77} or Asp^{78} and nsp7 Phe^{76}, are distant from any protein-protein or protein-RNA interaction and thus are not expected to play a role in polymerase complex formation. Unexpectedly, the identical set of residues in BS II supports all interactions (Glu^{60}, Met^{62}, Ala^{63}, Met^{67}, and Met^{70}) between nsp8b and nsp12/nsp13.1 and between nsp8a and nsp13b (Fig. 8J). Notably, within the polymerase complex, amino acids involved in RNA binding point in the opposite direction of the protein interfaces and have little or no role in nsp7+8 complex formation.

DISCUSSION

Our findings reveal the nsp7+8 quaternary composition of seven CoVs representing five CoV species of the genera *Alpha-* and *Betacoronavirus*. Viruses of the same species (SARS-CoV/SARS-CoV-2 and TGEV/FIPV, respectively) produce the same type of nsp7+8 complexes. Next to a conserved nsp7+8 heterodimer (1:1), the inherent specificity of nsp7+8 complex formation categorizes them into three groups: group A forming only heterotetramers (2:2), group B forming only heterotrimers (2:1), and group AB forming both heterotetramers (2:2) and heterotrimers (2:1). Complexes of the same stoichiometry exhibit a conserved topology, consisting of an nsp8 homodimeric scaffold for the heterotetramers and an nsp7 homodimeric core for the trimers. Candidate structures based on our results highlight *Alpha-* and *Betacoronavirus*-wide conserved binding sites on nsp8, named BS I and BS II, which provide the modular framework for a variety of complexes. Furthermore, unique molecular contacts for the complex groups have the potential to determine the ability and preference for heterotrimer and/or heterotetramer formation (results overview in table S6).

We provide evidence that, even at high concentrations, the SARS-CoV-2 nsp7+8 heterotetramer (2:2) represents the predominant

species. To relate our results to in vivo conditions, we consider the following aspects: According to maximum molecular crowding (*31*), polyproteins pp1a and pp1ab can reach a maximum of 125 to 450 μM, which translates to 3.9 to 11.7 mg/ml nsp7+8. This range is covered by our DLS and SAXS analysis. In absence of other interaction partners, we expect that, in vivo, the nsp7+8 (2:2) heterotetramer represents the predominant nsp7+8 complex of SARS-CoV-2 and other heterotetramer-forming CoVs of complexation groups A and AB.

The heterotetramer candidate structures and models presented here are based on the conformers T1 or T2 of the SARS-CoV heterohexadecamer structure, which contains full-length nsp8 (*22*). Although our results cannot clarify if one of these conformers is the biologically relevant structure existing in solution, the combined evidence provided here strongly suggests structural similarity to T1/T2. Considering the crystallographic origin of T1/T2 and the overlap of binding sites, the heterotetramer could well be a flexible and dynamic structure in solution.

In contrast to our findings, a SARS-CoV nsp7+8 hexadecamer structure has been reported (*22*). However, this structure has been derived from x-ray crystallography, hence showing a static, frozen state, where the crystal lattice formation favors stabilized arrangements that could differ from the solution state of the protein complexes. In the case of nsp8, the flexible N terminus could inhibit crystal formation and has been removed in some studies (*24*, *25*). Alternatively, it may stabilize specific interactions, thereby promoting crystal formation by binding to one of the multiple interfaces presented between nsp7 and nsp8, resulting in a physiologically irrelevant larger oligomeric structure. The SAXS data presented here not only partially support this scenario at high protein concentrations but also confirm a predominantly heterotetrameric assembly in solution. Thus, a potential shift of quaternary structure from a heterotetramer toward a higher-order complex, such as a heterohexadecamer, appears unlikely unless triggered, e.g., by binding to nucleic acids as has been repeatedly described for nsp7+8 complexes (*22*).

All seven CoV nsp7 and nsp8 proteins shown here also form heterodimers (1:1). Such heterodimeric subcomplexes with nsp7 bound to nsp8 BS I have been observed in all deposited complex structures containing nsp7+8 (*22–25*) or nsp12 (*14–17*). Therefore, the heterodimer represents the most basic form of nsp7+8 complexes and likely serves as a universal substructure building block in the coordinated assembly of functional RTCs of CoVs from the genera *Alpha*- and *Betacoronavirus*. Moreover, heterotrimer and heterotetramer formation are based on a second canonical binding site at the nsp8 N-terminal domain, BS II. This site appears to have a high propensity
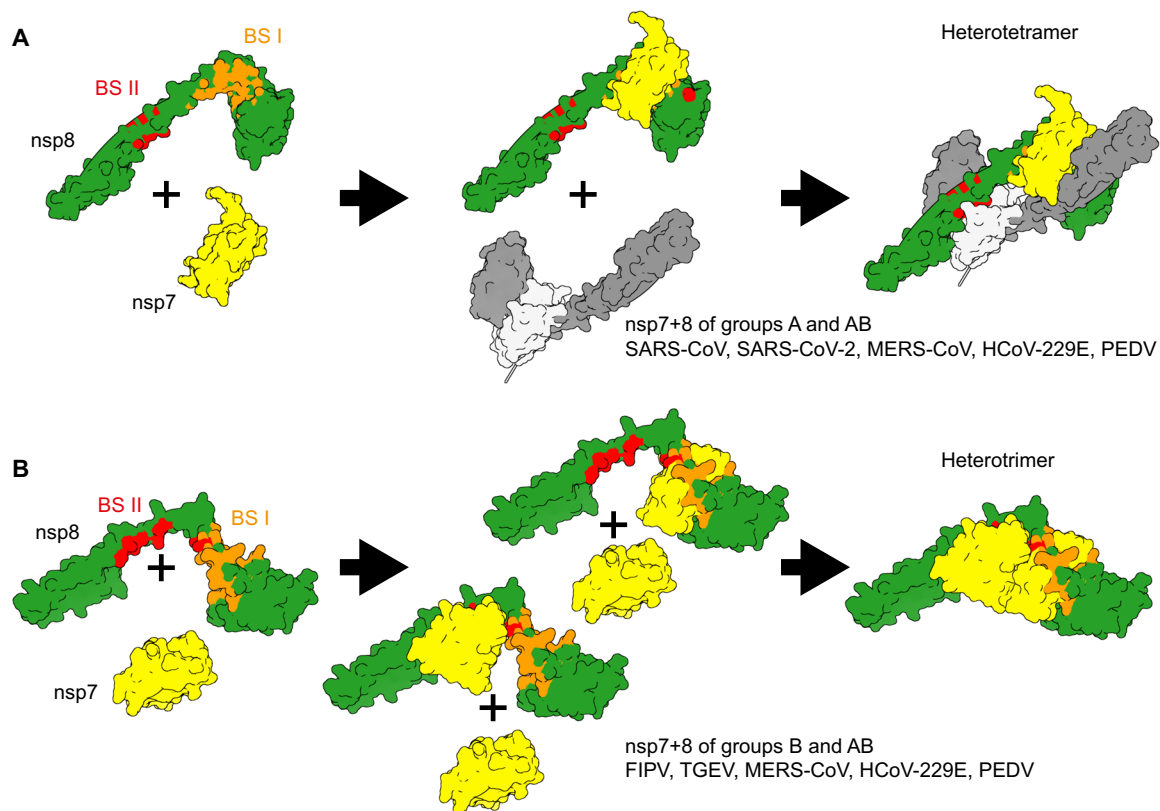

Fig. 9. Proposed model for nsp7+8 complex formation. (**A**) For complexes of group A, heterodimers form via nsp8 BS I, which quickly dimerize via BS II into a heterotetramer. A theoretic route via a preformed nsp8 scaffold is unlikely to play a role in heterotetramer formation since no nsp7+8 (1:2) intermediates are observed for complexes of SARS-CoV or SARS-CoV-2. Moreover, nsp7 and nsp8 occupy neighboring positions in the replicase polyproteins, thus favoring their interaction (in cis) at early stages in the infection cycle (when intracellular viral polyprotein concentrations are low) over intermolecular interactions between different replicase polyprotein molecules as is also evident from the low dimerization ability of the precursors. (**B**) For group B complexes, we propose the formation of a heterodimer intermediate via nsp8 BS I or BS II and subsequent recruitment of a second nsp7, resulting in an nsp7+8 (2:1) heterotrimer. This is also supported by the relatively high peak fractions of heterodimers detected. Group AB complexes can use both complexation pathways. In line with this, the proteins also produce a relatively high heterodimer signal but, ultimately, prefer to form heterotetramers rather than heterotrimers.

to form complexes with various binding partners (e.g., nsp7+8, nsp12, or nsp13). Accordingly, our analysis suggests that the nsp8 BS II strives for occupation. The nsp7+8 quaternary composition, topology, and analysis of binding sites presented here allow us to reconstruct and propose a model of the complex formation pathway (Fig. 9).

The preference for heterotrimer and heterotetramer can probably be pinpointed to just a few amino acids within nsp8 BS II or in nsp7 interacting with it. Here, we identify two contacts that could have unique discriminatory potential for promoting heterotrimeric (nsp7 Phe[76]) or heterotetrameric (nsp8 Glu[77] and Asn[78]) quaternary structures. Notably, in the presence of nsp7 Phe[76] and nsp8 Asn[78], as observed for group AB, the heterotetramer is always more abundant than the heterotrimer. However, compared to the entire BS II, these contacts only represent a small share of the binding interface and contribute little interaction energy through van der Waals forces. Nevertheless, the unique position of their contacts could critically determine the types of interactions with one or another binding partner.

Because the critical residues required for nsp7+8 complex formation have no overlap with nsp12 interaction sites, direct docking of preformed heterotrimers and heterotetramers to nsp12 can be expected. Furthermore, heterotrimeric and heterotetrameric structures are compatible with accommodation of specific RNA structures similar to what has been suggested for heterohexadecameric nsp7+8 by Rao and colleagues (15). Notably, if heterotrimeric or heterotetrameric nsp7+8 structures were associated with nsp12, the binding site for nsp13 would be blocked, which may have regulatory implications for CoV replication. Together, these conserved binding mechanisms and overlapping binding sites confirm the proposed role of nsp8 as a major interaction hub within the CoV RTC (32) and indicate critical regulatory functions by specific nsp7+8 complexes.

Last, we can only speculate about possible reasons for the existence of different nsp7+8 complexes: (i) similar kinetic stability due to occupation of both binding sites (both structures exist because they are equally efficient in occupying BS I and BS II), (ii) unknown functional relevance in CoV replication (e.g., specificity to RNA structures channeled to the nsp12 RdRp), or (iii) adaptation to host factors and possible regulatory functions.

In summary, our work shows, and provides a framework to understand, the characteristic distribution and structures of nsp7+8 (1:1) heterodimers, (2:1) heterotrimers, and (2:2) heterotetramers in representative *Alpha-* and *Betacoronaviruses*. The nsp7+8 structure in solution can be used to investigate its independent functional role in the formation of active polymerase complexes and, possibly, regulation and coordination of polymerase and other RTC activities, for example, in the context of antiviral drug development targeting different subunits of CoV polymerase complexes reconstituted in vitro.

MATERIALS AND METHODS
Cloning and gene constructs
The codon-optimized sequence for the SARS-CoV-2 nsp7-8 region (NC_045512.2) was synthesized by Eurofins scientific SE with overhangs suitable for insertion into pASK-IBA33plus plasmid DNA (IBA Life Sciences). A golden gate assembly approach using Eco31I (Bsa I) (Thermo Fisher Scientific) was used to shuttle the gene into the plasmid. Linker and tag of the expression construct SARS-CoV-2 nsp7-8-His$_6$ contained the C-terminal amino acids –SGSGSARGS-His$_6$ (SGSG residues as P1′-P4′ of M^{pro} cleavage site and SARGS-His$_6$

residues as the default linker of pASK vectors). The SARS-CoV nsp7-8 pASK33$^+$ plasmid generated previously (23, 33) was used for the expression of SARS-CoV nsp7-8-His$_6$ containing the C-terminal amino acids –SARGS-His$_6$. The expression plasmid for SARS-CoV M^{pro} was generated as described by Xue et al. (34). To produce nsp7-8-SGSGSARGS-His$_6$ precursor proteins in *Escherichia coli*, the nsp7-8 coding sequences of HCoV-229E (HCoV-229E; GenBank accession number AF304460), FIPV (FIPV, strain 79/1146; DQ010921), SARS-CoV (strain Frankfurt-1; AY291315), PEDV (PEDV, strain CV777, NC_003436), and TGEV (TGEV, strain Purdue; NC_038861) were amplified by reverse transcription polymerase chain reaction (RT-PCR) from viral RNA isolated from cells infected with the respective viruses and inserted into pASK3-Ub-CHis$_6$ using restriction- and ligation-free cloning methods as described before (35). Similarly, the nsp7-9 or nsp7-11 coding region of MERS-CoV (strain HCoV-EMC; NC_019843) was amplified by RT-PCR from infected cells and inserted into pASK3-Ub-CHis$_6$. The HCoV-229E and FIPV nsp5 coding sequences were cloned into pMAL-c2 plasmid DNA (New England Biolabs) for expression as maltose-binding protein (MBP) fusion proteins containing a C-terminal His$_6$-tag. Primers used for cloning and mutagenesis are available upon request.

Expression and purification
SARS-CoV M^{pro} was produced with authentic ends as described in earlier work (34). For amino acid sequences and protein IDs of nsp7-8-His$_6$ precursor proteins, see table S1. To produce the precursors, SARS-CoV and SARS-CoV-2 nsp7-8-His$_6$, BL21 Rosetta2 (Merck Millipore) were transformed, grown in culture flasks to OD$_{600}$ (optical density at 600 nm) = 0.4 to 0.6, and then induced with 200 μg anhydrotetracycline per liter culture; the cultures then continued to grow at 20°C for 16 hours. For pelleting, cultures were centrifuged (6000g for 20 min) and cells were frozen at −20°C. Cell pellets were lysed in 1:5 (v/v) buffer B1 [40 mM phosphate buffer and 300 mM NaCl (pH 8.0)] with one freeze-thaw cycle, sonicated (micro tip, 70% power, six times on 10 s, off 60 s; Branson digital sonifier SFX 150), and then centrifuged (20,000g for 45 min). Proteins were isolated with Ni^{2+}-NTA beads (Thermo Fisher Scientific) in gravity flow columns (BioRad). Proteins were bound to beads equilibrated with 20 column volumes (CV) of B1 + 20 mM imidazole and then washed with 20 CV of B1 + 20 mM imidazole followed by 10 CV of B1 + 50 mM imidazole. The proteins were eluted in eight fractions of 0.5 CV of B1 + 300 mM imidazole. Immediately after elution, fractions were supplemented with 4 mM dithiothreitol (DTT). Before analysis with native MS, Ni^{2+}-NTA eluted fractions containing the polyprotein were concentrated to 10 mg/ml and further purified over a 10/300 Superdex 200 column (GE Healthcare) in 20 mM phosphate buffer, 150 mM NaCl, and 4 mM DTT (pH 8.0). The main elution peaks contained nsp7-8. For quality analysis, SDS-PAGE was performed to assess the sample purity.

To obtain a prepurified SARS-CoV-2 nsp7+8 complex for DLS and SAXS, eluate fractions from the Ni^{2+}-NTA column containing the nsp7-8-His$_6$ were concentrated and the buffer was exchanged with a PD-10 desalting column (GE Healthcare) equilibrated with 50 mM tris (pH 8.0), 100 mM NaCl, 4 mM DTT, and 4 mM MgCl$_2$ [size exclusion chromatography (SEC) buffer]. Then, nsp7-8-His$_6$ was eluted with 3.5 ml of SEC buffer and subsequently cleaved with M^{pro}-His$_6$ (1:5, M^{pro}-His$_6$: nsp7-8-His$_6$) for 16 hours at RT. M^{pro}-His$_6$ was removed with Ni-NTA agarose, and the cleaved nsp7+8 complex was subjected to a HiLoad 16/600 Superdex 75 pg size exclusion column equilibrated with SEC buffer.

The HCoV-229E, PEDV, FIPV, and TGEV nsp7-8-His$_6$ and MERS-CoV nsp7-11-His$_6$ precursor proteins were produced and purified as described before (35) with a slightly modified storage buffer. Anion-exchange chromatography fractions of the peak containing the desired protein were identified by SDS-PAGE, pooled, and dialyzed against storage buffer [50 mM tris-Cl (pH 8.0), 200 mM NaCl, and 2 mM DTT].

MBP-nsp5-His$_6$ fusion proteins were purified using Ni^{2+}-IMAC as described before (35). To produce HCoV-229E and FIPV MBP-nsp5-His$_6$, *E. coli* TB1 cells were transformed with the appropriate pMAL-c2-MBP-nsp5-His$_6$ construct and grown at 37°C in LB medium containing ampicillin (100 µg/ml). When an OD$_{600}$ of 0.6 was reached, protein production was induced with 0.3 mM isopropyl β-D-thiogalactopyranoside and cells were grown for another 16 hours at 18°C. Thereafter, the cultures were centrifuged (6000*g* for 20 min) and the cell pellet was suspended in lysis buffer [20 mM tris-Cl (pH 8.0), 300 mM NaCl, 5% glycerol, 0.05% Tween 20, 10 mM imidazole, and 10 mM β-mercaptoethanol] and further incubated with lysozyme at 4°C (0.1 mg/ml) for 30 min. Subsequently, cells were lysed by sonication and cell debris was removed by centrifugation for 30 min at 40,000*g* and 4°C. The cell-free extract was bound to preequilibrated Ni^{2+}-NTA (Qiagen) matrix for 3 hours. Ni^{2+}-IMAC elution fractions were dialyzed against buffer composed of 20 mM tris-Cl (pH 7.4), 200 mM NaCl, 5 mM CaCl$_2$, and 2 mM DTT and cleaved with factor Xa to release nsp5-His$_6$. Then, nsp5-His$_6$ was passed through an amylose column and subsequently bound to Ni^{2+}-NTA matrix to remove any remaining MBP. Following elution from the Ni^{2+}-NTA column, nsp5-His$_6$ was dialyzed against storage buffer [20 mM tris-Cl (pH 7.4), 200 mM NaCl, and 2 mM DTT] and stored at −80°C until further use.

The enriched protein complexes for SDS-PAGE analysis were generated by cleaving 15 µg of nsp7-8-His$_6$ precursor protein with M^{pro} (nsp5-His$_6$, 5 µg) for 48 hours at 4°C. Subsequently, His$_6$-tag–containing cleavage products were removed by passing the material through a Ni^{2+}-IMAC column and nsp7+8 complexes were enriched by ion-exchange chromatography.

Native MS

To prepare samples for native MS measurements, M^{pro} was buffer-exchanged into 300 mM AmAc and 1 mM DTT (pH 8.0) by two cycles of centrifugal gel filtration (Biospin mini columns, 6000 MWCO (molecular weight cutoff), BioRad), and the precursors were transferred into 300 mM AmAc and 1 mM DTT (pH 8.0) by five rounds of dilution and concentration in centrifugal filter units (Amicon, 10,000 MWCO, Merck Millipore). Cleavage and complex formation was started by mixing nsp7-8-His$_6$ and protease M^{pro} with final concentrations of 15 and 3 µM, respectively. Three independent reactions were started in parallel and incubated at 4°C overnight.

Tips for nano-electrospray ionization (nanoESI) were pulled in-house from borosilicate capillaries (1.2-mm outer diameter, 0.68-mm inner diameter, with filament, World Precision Instruments) with a micropipette puller (P-1000, Sutter Instruments) using a squared box filament (2.5 mm by 2.5 mm, Sutter Instruments) in a two-step program. Subsequently, tips were gold-coated using a sputter coater (Q150R, Quorum Technologies) with 40 mA, 200 s, tooling factor 2.3, and end bleed vacuum of 8×10^{-2} mbar argon.

Native MS was performed at a nanoESI quadrupole time-of-flight (Q-TOF) instrument (Q-TOF2, Micromass/Waters, MS Vision) modified for higher masses (36). Samples were ionized in positive ion mode with voltages applied at the capillary of 1300 to 1500 V and at the cone of 130 to 135 V. The pressure in the source region was kept at 10 mbar throughout all native MS experiments. For desolvation and dissociation, the pressure in the collision cell was 1.5×10^{-2} mbar argon. For native MS, accelerating voltages were 10 to 30 V and quadrupole profile was 1000 to 10,000 *m/z*. For CID-MS/MS, acceleration voltages were 30 to 200 V. Raw data were calibrated with CsI (25 mg/ml) and analyzed using MassLynx 4.1 (Waters). Peak deconvolution and determination of relative intensity were performed using UniDec (37). All determined masses are provided (table S2).

Chemical cross-linking of prepurified complexes

To chemically cross-link prepurified nsp7+8 complexes, 5 µg of protein was incubated with 10 µM BS^3 (Thermo Fisher Scientific) in reaction buffer [20 mM Hepes-KOH (pH 8.0), 30 mM KCl, and 2 mM β-mercaptoethanol]. Cross-linking was carried out at 37°C for 30 min and quenched with 50 mM AmAc for another 30 min at 37°C. After terminating the cross-linking reaction, the samples were mixed with an excess of Laemmli sample buffer [50 mM tris-HCl (pH 6.8), 2.5% (w/v) SDS, 10% (v/v) glycerol, and 0.01% (w/v) bromophenol blue] and analyzed on 12% SDS-PAGE.

Cross-linking matrix-assisted laser desorption/ionization

Prepurified FIPV nsp7+8 and HCoV-229E nsp7+8 at 20 µM were cross-linked with 0.15% glutaraldehyde (Sigma-Aldrich) at 4°C for 25 min before diluting them to 1 µM in MALDI matrix solution (sinapinic acid 10 mg/ml in acetonitrile/water/TFA, 49.95:49.95:0.1, v/v/v) and spotting (1 µl) onto a stainless steel MALDI target plate. The MALDI-TOF/TOF mass spectrometer (ABI 4800, AB Sciex) equipped with a high-mass detector (HM2, CovalX) was used in linear mode. For acquiring mass spectra (1000 to 1,000,000 *m/z*), spots were ionized with a Nd:YAG laser (355 nm) and 500 shots per spectrum were accumulated. Obtained raw data were smoothed and analyzed using mMass [v5.5.0, by Martin Strohalm (38)] and origin2016.

Dynamic light scattering

To check the monodispersity of the samples and to study the stoichiometry of the nsp7+8 complexes, DLS measurements were performed with the Spectro Light 600 (Xtal Concepts). The complex was concentrated to various concentrations and samples were spun down for 10 min at 12,000 rpm and 4°C. A Douglas Vapour batch plate (Douglas Instruments) was filled with paraffin oil, and 2 µl of each sample was pipetted under oil. DLS measurements for each sample were performed at 20°C with 20 measurements for 20 s each, respectively. Data points depicting the R_0 with increasing complex concentrations were derived from 20 consecutive DLS measurements over 20 s each; error bars show SD.

Small-angle x-ray scattering

SAXS data were collected on the P12 beamline of EMBL at the PETRA III storage ring (DESY, Hamburg). An x-ray wavelength of 1.24 Å (10 keV) was used for the measurements, and scattered photons were collected on a Pilatus 6M detector (Dectris), with a sample-to-detector distance of 3 m. Data were collected on 22 concentrations ranging from 1.2 to 48 mg/ml nsp7+8 in 50 mM tris (pH 8.0), 100 mM NaCl, 4 mM DTT, and 4 mM MgCl$_2$; pure buffer was measured between samples. For each data collection, 20 frames of 100 ms were collected. 2D scattering images were radially averaged and normalized

to the beam intensity. The frames were compared using the program Cormap (*39*, *40*), and only similar frames were averaged and used for further analysis to avoid possible beam-induced effects. Scattering collected on the pure buffer was subtracted from that of the sample, and the resulting curves were normalized to the protein concentration to obtain the scattering of nsp7+8 complexes.

The data processing pipeline SASflow was used for data reduction and calculation of the overall SAXS parameters (*40*). For R_g values, error bars correspond to the SD of the experimental data from the fit of the linear Guinier region plus the SD of R_g values from all possible intervals from the R_g values from the selected interval (*41*). Molecular weights were inferred from different molecular calculation methods using a Bayesian assessment; the error bars correspond to the credibility interval computed using Bayesian assessment of the protein molecular weight (*42*). The program CRYSOL was used to compute the theoretical curves from the atomic structures (*43*). Volume fractions of the components of the oligomeric mixtures were computed and fitted to the data using the program OLIGOMER (*44*). The dimer of T1 was built by the program SASREFMX (*41*), which builds a dimeric model that fits best, in mixture with the monomeric T1, multiple scattering curves collected at different concentrations.

Sequence alignment
Amino acid sequences of nsp7-8-His6 precursor proteins (table S1) were aligned with Clustal Omega (*45*) and converted by ESPript (*46*) using the amino acid sequences without C-terminal linkers and His6 as input. For the multiple sequence alignment with identity matrix output, the SIAS Sequence identity and similarity tool has been used, provided by Secretaria General de Sciencia, Technologica e Innovacion of Spain (http://imed.med.ucm.es/Tools/sias.html). As input parameter, length of the smallest sequence was selected.

Visualization
Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (*47*).

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/10/eabf1004/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES
1. J. Cui, F. Li, Z. L. Shi, Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
2. A. M. Zaki, S. Van Boheemen, T. M. Bestebroer, A. D. Osterhaus, R. A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
3. C. Drosten, S. Gunther, W. Preiser, S. van der Werf, H. R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R. A. M. Fouchier, A. Berger, A. M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J. C. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H. D. Klenk, A. D. M. E. Osterhaus, H. Schmitz, H. W. Doerr, Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species *Severe acute respiratory syndrome-related coronavirus*: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).
5. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
6. E. M. Leroy, M. A. Gouilh, J. Brugère-Picoux, The risk of SARS-CoV-2 transmission to pets and other wild and domestic animals strongly mandates a one-health strategy to control the COVID-19 pandemic. *One Health.* **10**, 100133 (2020).
7. B. B. Oude Munnink, R. S. Sikkema, D. F. Nieuwenhuijse, R. J. Molenaar, E. Munger, R. Molenkamp, A. van der Spek, P. Tolsma, A. Rietveld, M. Brouwer, N. Bouwmeester-Vincken, F. Harders, R. Hakze-van der Honing, M. C. A. Wegdam-Blans, R. J. Bouwstra, C. GeurtsvanKessel, A. A. van der Eijk, F. C. Velkers, L. A. M. Smit, A. Stegeman, W. H. M. van der Poel, M. P. G. Koopmans, Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2021).
8. N. C. Pedersen, An update on feline infectious peritonitis: Virology and immunopathogenesis. *Veterinary J.* **201**, 123–132 (2014).
9. A. Vlasova, Q. Wang, K. Jung, S. Langel, Y. S. Malik, L. Saif, in *Porcine Coronaviruses in Emerging and Transboundary Animal Viruses* (Springer, 2020), pp. 79–110.
10. E. J. Snijder, E. Decroly, J. Ziebuhr, The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res.* **96**, 59–126 (2016).
11. M. J. van Hemert, S. H. van den Worm, K. Knoops, A. M. Mommaas, A. E. Gorbalenya, E. J. Snijder, SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLOS Pathog.* **4**, e1000054 (2008).
12. A. Shannon, B. Selisko, N.-T.-T. Le, J. Huchting, F. Touret, G. Piorkowski, V. Fattorini, F. Ferron, E. Decroly, C. Meier, B. Coutard, O. Peersen, B. Canard, Rapid incorporation of Favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat. Commun.* **11**, 4682 (2020).
13. L. Subissi, C. C. Posthuma, A. Collet, J. C. Zevenhoven-Dobbe, A. E. Gorbalenya, E. Decroly, E. J. Snijder, B. Canard, I. Imbert, One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3900–E3909 (2014).
14. R. N. Kirchdoerfer, A. B. Ward, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 (2019).
15. Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L. W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779–782 (2020).
16. Q. Peng, R. Peng, B. Yuan, J. Zhao, M. Wang, X. Wang, Q. Wang, Y. Sun, Z. Fan, J. Qi, G. F. Gao, Y. Shi, Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. *Cell Rep.* **31**, 107774 (2020).
17. H. S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, P. Cramer, Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154–156 (2020).
18. J. Chen, B. Malone, E. Llewellyn, M. Grasso, P. M. Shelton, P. D. B. Olinares, K. Maruthi, E. T. Eng, H. Vatandaslar, B. T. Chait, Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **182**, 1560–1573.e13 (2020).
19. F. Ferron, L. Subissi, A. T. S. De Morais, N. T. T. Le, M. Sevajol, L. Gluais, E. Decroly, C. Vonrhein, G. Bricogne, B. Canard, Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E162–E171 (2018).
20. C. J. Gordon, E. P. Tchesnokov, J. Y. Feng, D. P. Porter, M. Götte, The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J. Biol. Chem.* **295**, 4773–4779 (2020).
21. B. Krichel, S. Falke, R. Hilgenfeld, L. Redecke, C. Uetrecht, Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem. J.* **477**, 1009–1019 (2020).
22. Y. Zhai, F. Sun, X. Li, H. Pang, X. Xu, M. Bartlam, Z. Rao, Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12**, 980–986 (2005).
23. Y. Xiao, Q. Ma, T. Restle, W. Shang, D. I. Svergun, R. Ponnusamy, G. Sczakiel, R. Hilgenfeld, Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J. Virol.* **86**, 4444–4454 (2012).
24. S. Li, Q. Zhao, Y. Zhang, Y. Zhang, M. Bartlam, X. Li, Z. Rao, New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. *Protein Cell* **1**, 198–204 (2010).
25. E. Konkolova, M. Klima, R. Nencka, E. Boura, Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* **211**, 107548 (2020).
26. J. Dülfer, A. Kadek, J.-D. Kopicki, B. Krichel, C. Uetrecht, Structural mass spectrometry goes viral. *Adv. Virus Res.* **105**, 189–238 (2019).
27. E. Erba, L. Signor, C. Petosa, Exploring the structure and dynamics of macromolecular complexes by native mass spectrometry. *J. Proteomics* **222**, 103799 (2020).
28. G. Wang, A. J. Johnson, I. A. Kaltashov, Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates. *Anal. Chem.* **84**, 1718–1724 (2012).
29. J. Ziebuhr, The coronavirus replicase. *Curr. Top. Microbiol. Immunol.* **287**, 57–94 (2005).
30. A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**, 595–599 (2002).

31. B. Akabayov, S. R. Akabayov, S.-J. Lee, G. Wagner, C. C. Richardson, Impact of macromolecular crowding on DNA replication. *Nat. Commun.* **4**, 1615 (2013).

32. A. Von Brunn, C. Teepe, J. C. Simpson, R. Pepperkok, C. C. Friedel, R. Zimmer, R. Roberts, R. Baric, J. Haas, Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PLOS ONE* **2**, e459 (2007).

33. S. Falke, "Coronaviral Polyprotein Nsp7–10: Proteolytic Processing and Dynamic Interactions within the Transcriptase/Replicase Complex," thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky (2014); https://ediss.sub.uni-hamburg.de/handle/ediss/5271?mode=full.

34. X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, Z. Rao, Production of authentic SARS-CoV Mpro with enhanced activity: Application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **366**, 965–975 (2007).

35. J. Tvarogová, R. Madhugiri, G. Bylapudi, L. J. Ferguson, N. Karl, J. Ziebuhr, Identification and characterization of a human coronavirus 229E nonstructural protein 8-Associated RNA 3′-terminal adenylyltransferase activity. *J. Virol.* **93**, e00291-19 (2019).

36. R. H. van den Heuvel, E. van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. Brouns, D. Langridge, J. van der Oost, J. Hoyes, A. J. Heck, Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 (2006).

37. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. Hochberg, J. L. Benesch, C. V. Robinson, Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

38. M. Strohalm, M. Hassman, B. Košata, M. Kodíček, mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **22**, 905–908 (2008).

39. D. Franke, C. M. Jeffries, D. I. Svergun, Correlation map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12**, 419–422 (2015).

40. D. Franke, A. G. Kikhney, D. I. Svergun, Automated acquisition and analysis of small angle x-ray scattering data. *Nucl. Instrum. Methods Phys. Res.* **689**, 52–59 (2012).

41. M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. Mertens, P. V. Konarev, D. I. Svergun, New developments in the *ATSAS* program package for small-angle scattering data analysis. *J. Appl. Cryst.* **45**, 342–350 (2012).

42. N. R. Hajizadeh, D. Franke, C. M. Jeffries, D. I. Svergun, Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci. Rep.* **8**, 7204 (2018).

43. D. Svergun, C. Barberato, M. H. Koch, CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).

44. P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. Koch, D. I. Svergun, *PRIMUS*: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).

45. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

46. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

47. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).