Chapter 5

The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both ***de novo*** initiation and primer extension

Aartjan J.W. te Velthuis[1], Sjoerd H.E. van den Worm[1#] and Eric J. Snijder[1]

Molecular Virology laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, PO Box 9600, 2300RC Leiden, The Netherlands.

Present address: Vaccine and Gene Therapy Institute, Oregon Health and Science University, Beaverton, Oregon, USA

Published in: Nucleic Acids Research, Volume 40, Pages 1037-1047, 2012. Epub October 2011.

ABSTRACT

Uniquely among RNA viruses, replication of the ~30-kilobase SARS-coronavirus genome is believed to involve two RNA-dependent RNA polymerase (RdRp) activities. The first is primer-dependent and associated with the 106-kDa nonstructural protein 12 (nsp12), whereas the second is catalysed by the 22-kDa nsp8. This latter enzyme is capable of *de novo* initiation and has been proposed to operate as a primase. Interestingly, this protein has only been crystallised together with the 10-kDa nsp7, forming a hexadecameric, dsRNA-encircling ring structure (*i.e.*, nsp(7+8), consisting of 8 copies of both nsps). To better understand the implications of these structural characteristics for nsp8-driven RNA synthesis, we studied the prerequisites for the formation of the nsp(7+8) complex and its polymerase activity. We found that in particular the exposure of nsp8's natural N-terminal residue was paramount for both the protein's ability to associate with nsp7 and for boosting its RdRp activity. Moreover, this "improved" recombinant nsp8 was capable of extending primed RNA templates, a property that had gone unnoticed thus far. The latter activity is, however, ~20-fold weaker than that of the primer-dependent nsp12-RdRp at equal monomer concentrations. Finally, site-directed mutagenesis of conserved D/ExD/E motifs was employed to identify residues crucial for nsp(7+8) RdRp activity.

INTRODUCTION

In the replicative cycle of RNA viruses, the crucially important process of RNA-templated RNA synthesis is generally performed by an RNA-synthesizing complex of viral enzymes [235,264]. Commonly, its core subunit is a single RNA-dependent RNA polymerase (RdRp) that drives the production of template strands for replication, new genome molecules, and - in many RNA virus groups - also subgenomic (sg) mRNAs. This canonical RdRp is structurally conserved among RNA viruses and widely accepted to drive catalysis of phosphodiester bond formation via a well-established reaction mechanism involving two metal ions that are coordinated by aspartate residues in its motifs A and C [89,279,281].

Uniquely among RNA viruses however, current evidence suggests that at least two RdRp activities are encoded by the genomes of members of the coronavirus (CoV) family, the +RNA virus group that infects a wide range of vertebrates and is renowned for its exceptionally large polycistronic genome of approximately 30 kilobases [65]. Both CoV RdRps belong to the set of 16 nonstructural proteins (nsps) that are produced through proteolytic processing of the pp1a and pp1ab replicase precursor polyproteins, which both derive from translation of the genomic RNA [58,323]. For the Severe Acute Respiratory Syndrome-associated coronavirus (SARS-CoV), which emerged in 2003 and caused worldwide concern due to the ~10% mortality rate associated with infection of humans [65,324], the two replicase subunits with RdRp activity have been studied in some detail. The first is the 106-kDa nsp12, which contains the canonical viral RdRp motifs in its C-terminal part and employs a primer-dependent initiation mechanism [154,325]. The second polymerase, the 22-kDa nsp8, is unique for CoVs and was reported to be only capable of *de novo* RNA synthesis on ssRNA templates, albeit with low fidelity [155]. Together, these observations inspired a hypothesis in which nsp8 would serve as an RNA primase, *i.e.*, would synthesise short oligonucleotide primers for subsequent extension by the nsp12 "main RdRp" [155].

In spite of this attractive model, however, many questions regarding CoV RNA synthesis remain unanswered thus far. For instance, it is unclear whether the homomeric form of nsp8, for which *in vitro* RdRp activity was previously documented [155], actually occurs *in vivo*, as nsp8 was also shown to co-crystallise and form a unique hexadecameric ring-structure with the 10-kDa nsp7 subunit, which resides immediately upstream in the replicase polyprotein precursors (Fig. 1) [153]. In a similar fashion, it is presently unknown whether the postulated double-stranded RNA (dsRNA) binding channel of this complex plays a role in the RdRp activity of nsp8 and whether this activity is influenced by nsp7, particularly given the observed low fidelity and low processivity of nsp8 [155].

To investigate the properties of the nsp7+nsp8 (nsp(7+8)) hexadecamer in more detail, and seek answers to the above questions, we here generated and purified re-

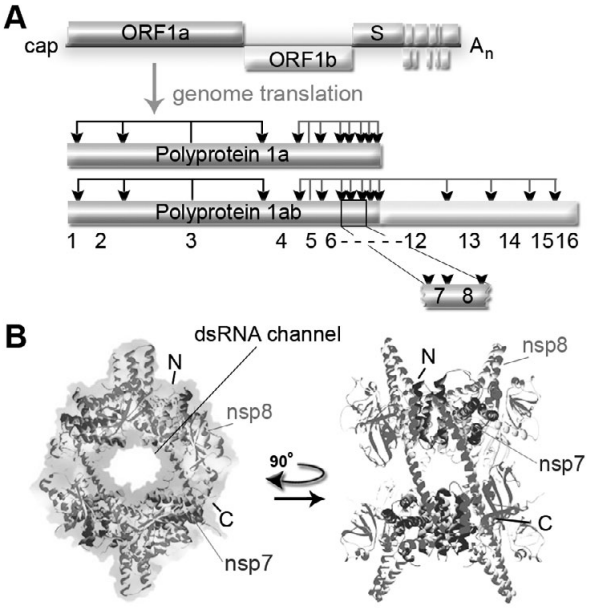

Figure 1: SARS-coronavirus genome organization and structure of the nsp7+nsp8 hexadecamer. (**A**) The coronavirus genome contains two large 5′-proximal ORFs (ORF1a and 1b) that encode the two replicase polyproteins, whose mature products bring about the formation of the viral replication and transcription complex. Both polyproteins are cleaved (cleavage sites indicated with arrow heads) by the proteinase activities of nsp3 (left hand side) and nsp5 (right hand side), which releases the maturensps. Also indicated are the 5′ cap structure and the 3′polyA tail (A_n). (**B**) The SARS-CoV nsp8 crystal structure (pdb 2AHM) resembles a "golf club-like" shape. This nsp8 conformation connects to a much larger, hexadecameric structure that is composed of seven additional nsp8 subunits (light grey) and eight nsp7 subunits (dark grey). The hollow hexadecameric ring structure has a positively charged channel (darker grey background shading) that was proposed to mediate RNA binding. The outside of the structure is predominantly negatively charged (light grey background shading).

combinant forms of SARS-CoV nsp8 and nsp(7+8) that have natural N-terminal residues. This technical refinement was found to greatly improve nsp8's ability to associate with nsp7. Moreover, and in contrast to previous observations [155], exposure of the natural N-terminus proved crucial for the enzymatic activity of the complex on partially double-stranded RNA templates, demonstrating that nsp(7+8) is capable of primer-dependent RdRp activity as well. Site-directed mutagenesis of nsp8 in the context of the nsp(7+8) complex identified a conserved D/ExD/E motif that is important for catalysis *in vitro*, possibly providing a first indication of the location of the presently unknown nsp8 active site. Overall, these results define the SARS-CoV nsp(7+8) complex as an intriguing multimeric RNA polymerase that is capable of primer extension.

RESULTS

N-terminal processing defines nsp8 multimerisation and nsp(7+8) complex formation

SARS-CoV nsp7 and nsp8 were previously reported to interact and form a hollow ring structure that is composed of an intricate nsp8 octamer supported by eight copies of nsp7 [153,289] (Fig. 1B). Based on the large diameter, positive charge of the hexadecamer's channel and *in silico* docking, it was proposed to be able to encircle dsRNA (Fig. 1B). However, the functional significance of the compound interactions between nsp7 and nsp8 is poorly understood, as are the polymerase activities associated with monomeric nsp8 or nsp8-containing multimers. So far, strategies for the purification of recombinant nsp8 have involved the use of affinity tags (*e.g.*, His_6 or glutathione-S-transferase (GST) [153,155]) that were fused to one terminus to facilitate protein recovery. Inadvertently though, such tags or other exogenous sequences may significantly impede the correct folding of enzymes and thus alter their stability or activity, as exemplified by studies of the poliovirus ($3D^{pol}$) and SARS-CoV (nsp12) RdRp subunits [154,277,278]. To circumvent this issue, we developed a protocol in which SARS-CoV nsp8 was expressed as a ubiquitin (ub) fusion protein carrying a C-terminal His_6-tag (ub-nsp8-His), which was subsequently processed at both termini in two steps. The first step was co-translational and involved the release of the N-terminal ub fusion partner by the co-expressed ubiquitin carboxyl-terminal hydrolase 1 (Upb1, Fig. 2A) [154,277]. The second proteolytic step, catalysed by a recombinant form of the SARS-CoV nsp5 main protease [326], removed the C-terminal His_6-tag and was performed either in solution (Fig. 2A and 2B) or when nsp8-His was immobilised to Talon beads. This procedure yielded SARS-CoV nsp8 with its exact natural N- and C-terminus (replicase residues Ala-3920 and Gln-4117, respectively; Fig. 2A), the product that is normally liberated by the nsp5-driven autoprocessing of the SARS-CoV replicase polyproteins [44].

In accordance with the octameric state observed in cross-linking experiments using glutaraldehyde (Fig. S1) or ethylene glycolbis [153], the hydrodynamic profile of the untagged nsp8 corresponded to a mass of ~160 kDa (Fig. 2D). To identify and explain differences with previously published observations, we also produced and characterised N- and C-terminally tagged forms of nsp8 (Fig. 2C). Importantly, under the same assay conditions, the N-terminally His_6-tagged nsp8 (His-nsp8) that was used in the original nsp8 RdRp activity study [155] showed a marked difference in multimerisation behaviour (Fig. 2D and Fig. S1). On the other hand, little difference was observed between untagged nsp8 and a C-terminally His_6-tagged version of the protein (nsp8-His; Fig. 2E).

To investigate whether nsp7 could influence the change in multimerisation behaviour, we next added separately purified and C-terminally processed nsp7 to the different nsp8 preparations. Interestingly, we found that nsp8 and nsp8-His could both associate with

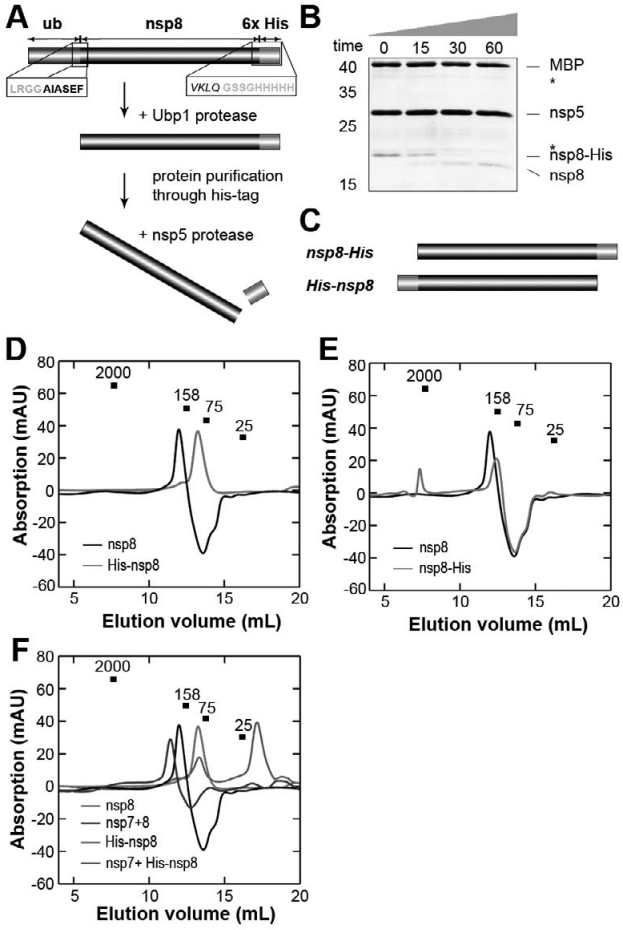

Figure 2: Purification and multimerisation of recombinant SARS-CoV nsp7-8 precursor and different nsp8 variants. (**A**) Expression of nsp8 in the presence of the ubiquitin protease Ubp1 to liberate the natural N-terminal sequence (AIASEF), followed by purification and cleavage by recombinant SARS-CoV nsp5 main protease to remove the C-terminal His_6-tag and its upstream GSSG linker. (**B**) Eighteen per cent SDS-PAGE analysis of nsp5-treated, purified nsp8-His demonstrates near-complete release of the C-terminal His_6-tag within 60 min. The maltose binding protein (MBP) was added to the reaction to serve as an independent loading control. Asterisks indicate nonspecific bands. (**C**) In addition to the tag-less nps8 and nsp8-His, we also produced the N-terminally His_6-tagged nsp8 (His-nsp8) used by Imbert *et al*. [155]. (**D**) Comparative gel filtration analysis of nsp8 (22 kDa as a monomer) versus His-nsp8 and (**E**) nsp8 versus nsp8-His. In all three cases, nsp8 formed multimers in solution, but the apparent molecular mass of complexes formed by both nsp8 and nsp8-His was ~2 fold higher than for complexes formed by His-nsp8. (**F**) Comparative analysis of nsp8, nsp(7+8), His-nsp8 and nsp7+nsp8-His. Only nsp(7+8) showed a molecular weight shift to the ~225-kDa size range with a standard deviation of 15-kDa (n = 3). This size is indicative of hexadecamer formation, whereas the analysis of nsp7+nsp8-His showed dominant peaks of nsp8-His and nsp7 (which is ~10 kDa as a monomer).

this protein, in accordance with published data [153], but that His-nsp8 was unable to do so within the frame of our experimental conditions (Fig. 2F). Consequently, although various lines of evidence support the observation that nsp7 and nsp8 can form a hexadecamer, it now appears that the correct N-terminal processing of nsp8 is a significant factor in determining the final oligomeric state of the protein.

SARS-CoV nsp7 enhances RNA binding by nsp8

A unique feature of the hexadecameric SARS-CoV nsp(7+8) structure is the fact that it does not derive from stacking of its protein subunits, but rather from stable interconnections of the "golf club-like" nsp8 molecules (Fig. 1B) [153]. The structural support of the nsp8 octamer by eight copies of nsp7 thus appears to be redundant, in line with the critical role for the nsp8 N-terminal domain described above. We surmised therefore that the additional complexity must have evolved to improve nsp8's function and set out to compare the RNA binding capabilities of the purified nsp8 octamer and nsp(7+8) hexadecamer.

By analysing the steady-state ribonucleotide-protein (RNP) complexes formed through binding of nsp8 to 5′ ^{32}P-labelled dsRNA (Fig. 3A), we estimated the nsp8 dissociation constant (K_d) for dsRNA to be ~3.3 μM (Fig. 3F), which is about ~25 fold higher than the apparent K_d of nsp12 under comparable conditions [154]. A comprehensive analysis of the influence of nsp7 on nsp8-dependent RNA binding required an nsp8 mutant that was incapable of RNA binding. To this end, we engineered an alanine substitution of the conserved residue K58, which resides in nsp8's proposed dsRNA-binding channel (residues 55-78 [153]). As is evident from the electromobility shift assay in Fig. 3B, this mutation was sufficient to significantly disrupt RNA binding. As a control, we also performed an aspartate-to-alanine substitution at position 52, which is partially conserved, yet not expected to participate in RNA backbone binding due to its negative charge and position just outside the proposed RNA binding channel. Indeed, the D52A mutation only induced a migratory shift of the dominant RNP signal towards the anode, likely as a result of the lost negative charge (Fig. 3C).

With the results obtained with these control proteins in mind, we next explored the contribution of nsp7 to RNA binding by the nsp(7+8) complex. We used a fixed concentration of nsp7 and added either wild-type or mutant nsp8 up to the point where the nsp7:nsp8 ratio reached equimolarity. No RNA binding was observed in the absence of nsp8, but upon nsp(7+8) complex formation the amount of bound dsRNA rapidly increased (Fig. 3D). Indicative of successful complex formation, we also observed a shift in the molecular weight of the major RNP complex formed (Fig. 3D). Western blot analysis confirmed that both nsp7 and nsp8 were present at this position in the gel (not shown), but due to the generally unpredictable migration behaviour of proteins and RNPs in native PAGE, it was not possible to assess whether this band indeed corresponded to the

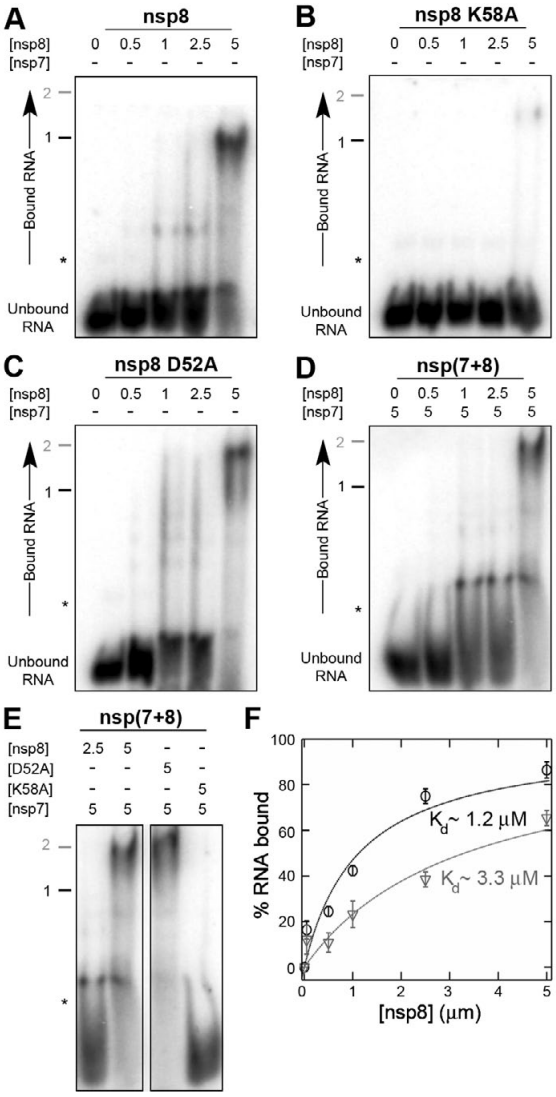

Figure 3: SARS-CoV nsp7 stimulates nsp8-dependent RNA binding. (**A**) Five prime ^{32}P-labelled dsRNA was incubated with increasing concentrations (0-5 μM) of wild-type nsp8, (**B**) nsp8 K58A, (**C**) or nsp8 D52A. Clearly, mutation of K58 to alanine significantly reduced the binding affinity of nsp8, whereas mutation of D52 to alanine did not. We also noted that the change in charge due to the mutation (up to eightfold in the octamer) resulted in an upward shift of the dominant RNP band, relative to the dominant RNP in panel 3A (labelled with black 1). (**D**) Five prime ^{32}P-labelled duplex RNA was incubated with a fixed concentration of nsp7 (5 μM) and increasing concentrations of wild-type nsp8 (0-5 μM). Note the migration shift of the dominant ribonucleotide-protein (RNP) complex in the presence of nsp7 (compare RNPs labelled with black 1 and grey 2). (**E**) Addition of an equimolar amount of nsp7 to the nsp8 mutants D52A and K58A stimulated binding of dsRNA. For reference the 2:1 and 1:1 ratios of wild-type nsp8 and nsp7 are shown in the left panel. Asterisks indicate nonspecific bands. (**F**) RNA-binding curves for nsp8 in the absence (grey triangles) or presence of a fixed (5 μM) concentration of nsp7 (black circles). Lines represent fits to the Hill equation, while error bars represent standard deviations (n = 3).

nsp(7+8) hexadecamer. The K_d of the nsp(7+8) complex was estimated at ~1.2 μM, about 3-fold lower than that of nsp8 alone (Fig. 3F).

When we next added an equimolar amount of nsp7 to the nsp8 RNA-binding mutant K58A, we observed a minor increase in the binding affinity for RNA (compare Fig. 3B with Fig. 3E). Mutant D52A, on the other hand, behaved similar to the wild-type protein (Fig. 3E). Together, these results complement the observation that various positively charged nsp7 residues line the inside of the nsp8-scaffolded RNA binding channel [153], and they provide the first direct evidence for a functional role of nsp7 in the SARS-CoV nsp(7+8) structure.

The nsp(7+8) complex has primer extension activity

Given nsp(7+8)'s ability to bind dsRNA, we wondered whether this protein complex would also be catalytically active on this type of template and able to incorporate nucleoside monophosphates (NMPs) into partially double-stranded RNA molecules, *i.e.*, primed templates. We therefore examined the ability of nsp8 to extend a 20-nt primer that was pre-annealed to a heteromeric template with relatively low secondary structure, to rule out potential adverse effects of hairpins (Fig. 4A). Interestingly and in contrast to previous observations [155], the nsp(7+8) complex readily extended the primer up to template length, resulting in the formation of a 40-base pair RNA duplex (Fig. 4B).

The negatively charged and helical polymer heparin is able to occupy the binding sites of RNA and DNA polymerases, and can thus directly compete with RNA and DNA templates. To verify that the full-length and longer RNA products were derived from single nsp(7+8) complexes bound to the template (*i.e.*, from a processive activity), and not from multiple binding and extension events (*i.e.*, a distributive activity), we performed the primer extension reaction in the presence of heparin to trap any unbound nsp(7+8). We first tested the concentration required to saturate all nsp(7+8) complexes in the reaction by titrating 0-100 μM into the reaction (Fig. S2A) and observed that the incorporation levels were stable above 1 μM (Fig. S2B), suggesting that these reactions represent single initiation-extension events. We next assessed whether the activity of nsp8 or nsp(7+8) was distributive or processive by quantifying the incorporated signal in full-length or longer products in the presence of 1 μM heparin (Fig. 4C). As shown in Fig. 4D, 66 ± 4% (mean ± standard deviation) of the nsp8 products were full length compared to 61 ± 2% of the nsp(7+8) products, suggesting that both enzymes complexes are mostly processive and that nsp7 does not confer additional processivity to nsp8. Interestingly, both nsp8 and nsp(7+8) are able to extend the RNA primers beyond template length in the presence of heparin (Fig. 4D and Fig. S2B), suggesting that these extensions result from terminal transferase activity and not from template switching, as was previously observed for poliovirus $3D^{pol}$ [284].

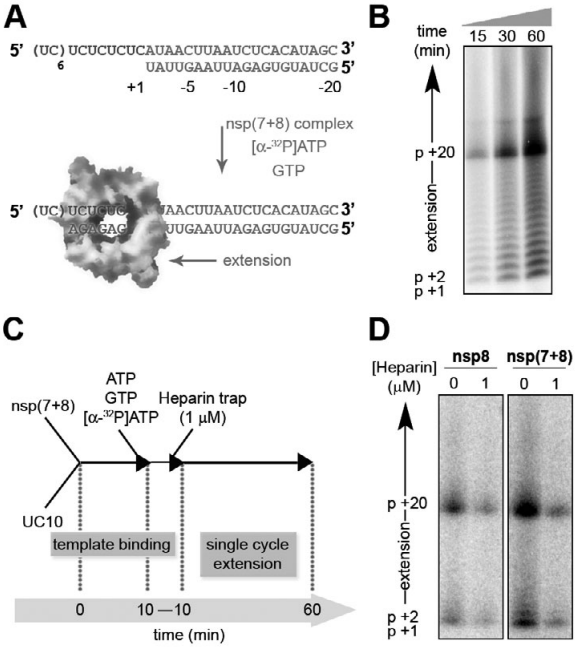

Figure 4: The nsp(7+8) complex has primer extension activity. (**A**) Schematic presentation of the nsp8 primer extension assay, in which [α-^{32}P]AMP and GMP are incorporated into a primed RNA template. (**B**) Incorporation of [α-^{32}P]AMP by the nsp(7+8) complex. Samples were taken at the indicated time points and resolved on a 20% PAGE/7M urea gel. (**C**) Schematic presentation of the single-cycle reaction. Template and nsp(7+8) complex were pre-incubated for 10 min before nucleotides were added. The mixture was then rapidly split into equal aliquots that were immediately mixed with heparin to trap unbound or released enzyme. (**D**) Samples were taken after 60 min and resolved on 20% PAGE/7M Urea.

The nsp(7+8) complex requires a D/ExD/E motif for catalysis

Intrigued by the primer extension activity of the SARS-CoV nsp(7+8) complex described above, we next designed a set of mutations to verify that the activity indeed was nsp(7+8) derived and to identify the most critical residues for activity in the complex. We first tested RNA-binding mutant K58A (Fig. 2) at varying concentrations and observed a ~95% loss of nucleotide incorporation activity compared to the wild-type protein (Fig. 5). Other likely candidates for a direct role in RdRp catalysis generally are Mg^{2+}-coordinating aspartate residues and lysine or histidine residues that can function as general acid [279]. In canonical RNA polymerases, the aspartates commonly reside in motifs A and C [84,279], while in DNA-dependent RNA primases they are usually found in a central D/ExD/E motif [328]. Given the absence of classical RdRp A and C motifs in the nsp8 sequence [155], we screened an alignment of CoV nsp8 sequences for conserved D/ExD/E motifs. Interestingly, we found such a motif in both the N-terminal and the C-terminal domain (Fig. 5A). Subsequent alanine substitution of the N-terminal D/

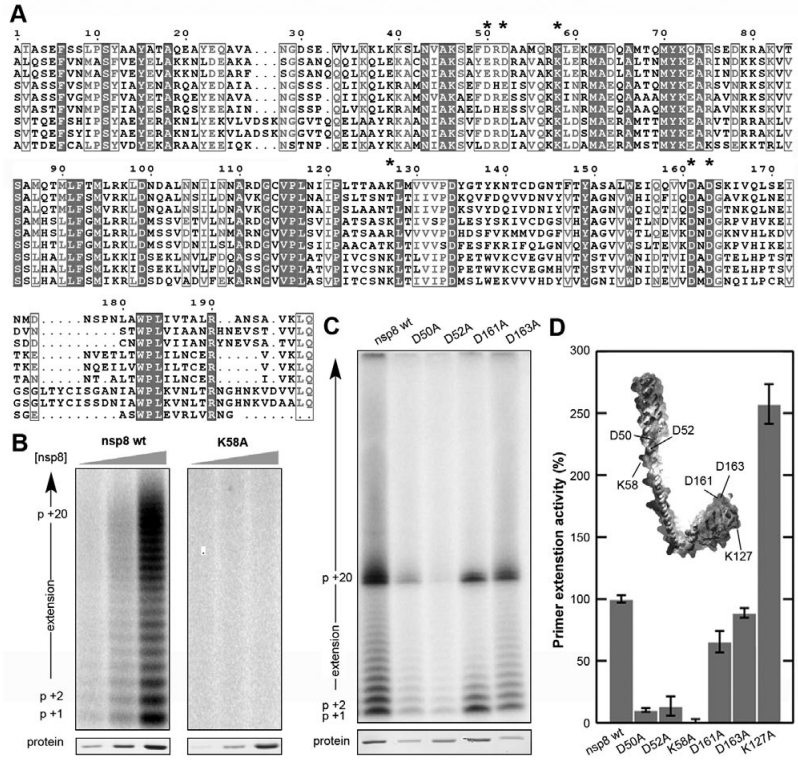

Figure 5: Mutagenesis of SARS-CoV nsp8. (**A**) Alignment of nsp8 sequences from representative alpha-, beta- and gammacoronaviruses. Fully conserved residues are shaded dark grey, while partially conserved residues are boxed. The residues targeted by mutagenesis are indicated with asterisks. Please see Material and Methods for the Genbank accession numbers associated with the presented sequences. Sequences are presented in the order SARS-CoV, MHV, HCoV OC43, HCoV NL63, HCoV 229E, Bat-CoV HKU8, IBV, Turkey CoV and BW-CoV. (**B**) To verify that the observed extension activity was nsp8-dependent, we tested the incorporation of AMP into the primed U_{20} template by 1, 5 or 10 µM of wild-type nsp8 or template-binding mutant K58A. Mutation of K58 resulted in a ~95% reduction of AMP incorporation. (**C**) To assess the importance of the two D/ExD/E motifs in nsp8, we engineered alanine substitution mutants of these residues and tested their primer extension activity on the primed UC_{10} template (see Fig. 4). Reactions were stopped after 60 min and compared to the activity of the wild-type nsp(7+8) complex on a 20% PAGE/7M urea gel. The bottom panel shows the nsp8 protein concentration present in each of the reactions. (**D**) Quantification of the primer extension activities on the CU_{10} template of the D/ExD/E alanine substitution mutants and control substitutions K58A and K127A. Values are normalised to the protein concentration. Error bars represent standard deviations (n=3).

ExD/E motif, composed of D50 and D52 in SARS-CoV, greatly affected primer extension activity on the CU_{10} template as shown in Fig. 5C. Mutation of the downstream domain (residues D161 and D163 in SARS-CoV), however, had a much smaller effect on polymerase activity, suggesting that this C-terminal D/ExD/E motif is not critical for catalysis. Controls included mutant K58A and a mutant carrying a lysine-to-alanine substitution

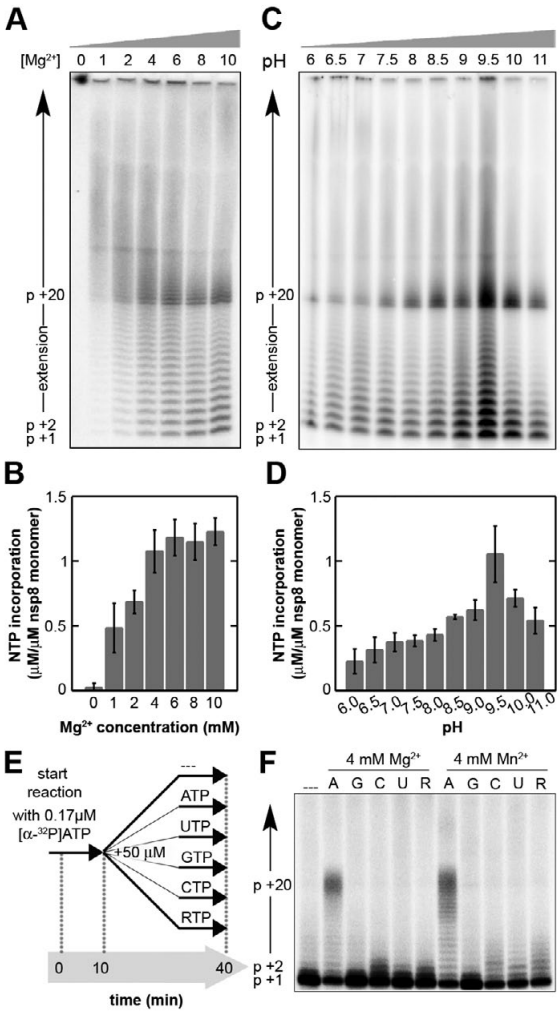

Figure 6: Influence of divalent ions and pH on nsp(7+8) activity. (**A**) To test the influence of magnesium ions on nsp(7+8) activity, we performed nsp(7+8) primer extension reactions at 0-10 mM Mg^{2+}. (**B**) Quantification of the results presented in Fig. 6A, presented as the amount of NTP incorporated per µM nsp8 monomer. Error bars represent standard deviations (n=3). (**C**) The influence of the pH on nsp(7+8) activity was tested for a pH range of 6-11. A clear optimum was observed around 9.5. (**D**) Quantification of the results in Fig. 6C, presented as the amount of NTP incorporated per µM nsp8 monomer. Error bars indicate standard deviations (n=3). (**E**) Schematic presentation of the pulse-chase experiment that was used to test the nsp(7+8) nucleotide incorporation specificity on a primed poly(U) template (see Table 1). The reactions were initiated with a limiting concentration of [α-^{32}P]ATP to allow the formation of a stable polymerase-template complex. Unlabelled nucleotides were used at a final concentration of 50 µM. (**F**) SARS-CoV nsp(7+8) allowed only limited transversional and transitional mutations. Use of manganese ions as cofactor for polymerase activity resulted in a minor, though noticeable loss of fidelity. Lane 1 represents the input signal to which no unlabelled nucleotides were added. Nucleoside triphosphates are abbreviated to single letters (i.e., A for ATP, G for GTP, U for UTP, C for CTP, and R for RTP).

of the non-conserved residue 127. In line with the observation of the U_{20} template and its conservation in CoVs, the loss of a lysine at position 58 resulted in a near complete loss of RdRp activity, whereas mutation of K127 positively influenced RNA synthesis (Fig. 5).

Influence of divalent ions and protons on nsp(7+8) activity

As outlined above, magnesium ions are well-known cofactors of nucleic acid polymerases and assist in the coordination and activation of incoming nucleoside triphosphates. Also the activity of SARS-CoV nsp(7+8) was found to be positively correlated with the Mg^{2+} concentration, albeit with a broad optimum running from 4-10 mM (Fig. 6A). At this optimum, nsp(7+8) incorporates ~1 μM NMP into the primed template per μM of monomeric nsp7 and nsp8 present in the reaction.

Similar to the presence of divalent cations, the pH greatly affects the activity of RdRps and has been shown to play a role in both catalysis and fidelity [242,279]. To investigate the influence of the pH on nsp(7+8), we tested the activity of the complex in a pH range of 6-11. As shown in Fig. 6C, we observed a sharp optimum at pH 9.5, which is considerably higher than the optimum that was previously observed for the SARS-CoV nsp12-RdRp and the His-nsp8 homomer (pH optimum 7.5 and 8.0, respectively) [154,155].

Interestingly, the primer extension activity of nsp(7+8) did not require manganese ions as was previously reported for the His-nsp8 homomer [155]. In fact, similar to the SARS-CoV nsp12-RdRp [154], the addition of Mn^{2+} was found to reduce the fidelity of nsp(7+8) and induce both transversional and transitional misincorporations in a pulse-chase experiment (Fig. 6E and 6F). Interestingly, the assay also revealed a discrimination against the widely used ATP and GTP analogue ribavirin triphosphate (RTP) [56,94]. Whether this may offer an explanation for SARS-CoV's relative resistance to this antiviral drug [329,330] remains an open question for future research.

N-terminal extensions other than nsp7 frustrate the primer extension activity of nsp8

The primer-extension and terminal transferase activity documented in Fig. 4 for the complex containing the untagged nsp8 was not observed by Imbert *et al*. when they first purified and analysed His-nsp8 [155]. To investigate whether this difference could be attributed to complex formation with nsp7 or the removal of the affinity tag, we performed the primer extension assay with three different recombinant nsp8 versions of which the gel filtration analysis is documented in Fig. 2. Interestingly, for all three variants primer-extension activity was observed (Fig. 7A), but the activity was most pronounced for nsp8-His and the untagged nsp8 (Fig. 7A). To estimate the effect of nsp7 on the nsp8-driven primer extension activity, we performed a direct comparison of the two enzyme complexes and found that the activity of nsp8 alone was >2 fold

lower than when nsp7 and nsp8 were present at equal molarity in the reaction (Fig. 4D and Fig. 7B). A similar comparison was performed for the *de novo* activity of nsp8, using the assay published by Imbert *et al*. [155] and taking the first dinucleotide (pppGpA) product as readout. Interestingly, both nsp8 and nsp(7+8) synthesised equal amounts of the pppGpA dinucleotide (Fig. 7C), suggesting that the effect of nsp8 is limited to the primer-extension activity of nsp8. In addition, we observed that the *de novo* initiation activity of nsp8 was ~2 fold higher than that of His-nsp8 (Fig. 7D).

Our comparative study revealed that the N-terminal His_6-tag of His-nsp8 greatly influences the primer-extension activity of nsp8 (Fig. 7A), its multimerisation profile and its association with nsp7 (Fig. 2). To test if this inhibitory effect was His_6-tag specific, we assessed the activity of a ub-nsp8-His fusion protein. At the same time, control reactions were performed in which we i) followed the activity of this protein as it was being processed by a recombinant form of the ubiquitin-cleaving nsp2 protease of equine arteritis virus [88] or ii) monitored the activity of nsp8-His. As shown in Fig. S3, the presence of the ub-tag decreased nsp8 activity to a level that was comparable to that of N-terminally His_6-tagged nsp8. Upon cleavage by EAV nsp2, however, a partial recovery of the primer extension activity was observed (Fig. S3). Unfortunately, we were not able to perform the same experiment with purified ub-nsp8, since our recombinant nsp5 removed the N-terminal ub-tag with similar efficiency as the C-terminal His_6-tag (Fig. S4).

Extrapolating to the situation in the viral pp1a and pp1ab precursor polyproteins, in which the nsp8 N-terminus is initially fused to nsp7 (Fig. 1A), our observations suggested that nsp8 may thus be inactive in the polyprotein context. This would constitute a form of regulation of viral enzyme activity that is not without precedent, since also the poliovirus 3Dpol is inactive as long as it is fused to the 3C protease in the 3CD precursor [331]. To verify this hypothesis, we expressed nsp7-8-His and tested this protein for RdRp activity. Interestingly, this fusion protein, a potential intermediate of CoV replicase polyprotein processing and a multimer in solution (Fig. 7E), showed primer extension activities that were comparable to or higher than the activity of nsp(7+8-His) (Fig. 7F). The *de novo* initiation activity of nsp7-8-His was, however, ~2 fold lower than the activity of nsp8 and nsp(7+8) (Fig. 7D). In conclusion, this result clearly underlines that the two N-terminal fusion partners other than nsp7 are specifically detrimental to SARS-CoV nsp8 primer-dependent RdRp activity *in vitro*. It also demonstrates that nsp8 alone may be sufficient to act as a primase.

Discussion

The complex replication and transcription process that coronaviruses initiate upon infection involves up to 16 viral nsps and at least one host factor [51,79,290]. Both individually as well as in complex with each other, these subunits engage in numerous protein-protein interactions [85,289] and embody various enzymatic activities, includ-

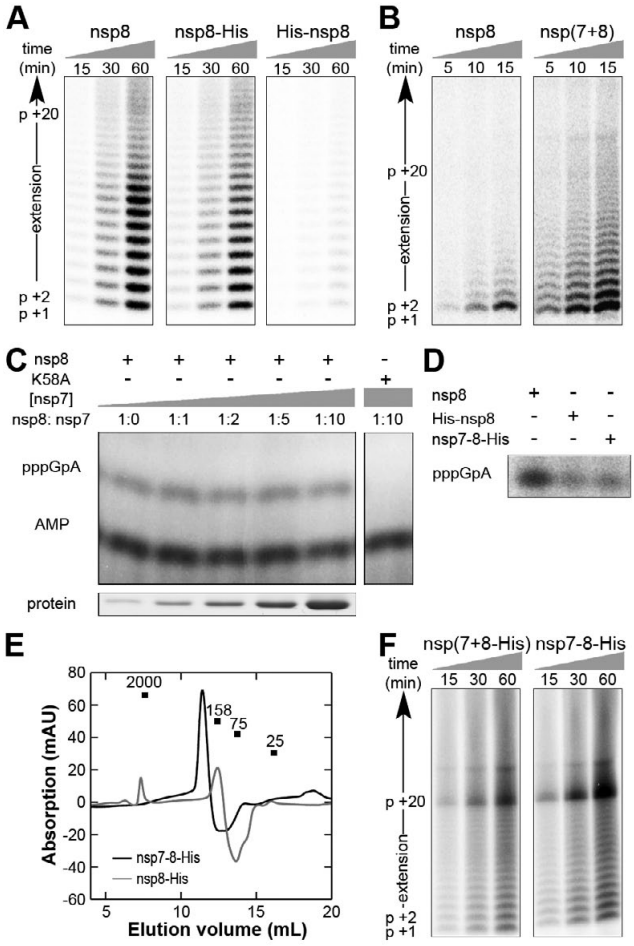

Figure 7: Influence of His_6-tags and nsp7 on the RdRp activity of SARS-CoV nsp8. (**A**) The UC_{10} template (see Fig. 4) was incubated with 1 µM of wild-type nsp8 or either of three nsp8 variants to investigate the influence of the His_6-tag. Samples were taken at the indicated time points and checked for [α-^{32}P]AMP incorporation by 20% PAGE/7M urea analysis. (**B**) Side-by-side comparison of the primer-extension activities of nsp8 and nsp(7+8). Shorter incubations are shown to better demonstrate the difference in activity. (**C**) *De novo* activity of nsp8 and nsp(7+8) on template AFMB131 (see Table 1), using the synthesis of the first dinucleotide pppGpA, as previously described by Imbert*et al*. [155], as readout. Nsp8 template binding mutant K58A was used as negative control. The AMP contaminant present in the used [α-^{32}P]ATP label is marked as loading control and size reference. (**D**) Side-by-side comparison of the *de novo* initiation activities of nps8, His-nsp8 and nsp7-8-His. (**E**) Elution profile of the nsp7-8-His fusion protein relative to nsp8-His. (**F**) Primer-extension activities of putative cleavage intermediate nsp7-8 on the U_{20} template (see Fig. 4 and Table 1).

ing proteolytic [44,332], ATPase [108], and 5′ cap modifying reactions [333]. Remarkably though, the mechanism and enzymes required to catalyse RNA synthesis in the CoV RTC

remain very poorly understood. Moreover, uniquely among RNA viruses which generally employ a single RNA polymerase to drive their RNA synthesis [235,264], the polymerase activity assays and nsp8 mutagenesis documented in this and other studies suggest that, in addition to the presumed nsp12 "main RdRp", other polymerase activities could play a critical role in the synthesis of SARS-CoV RNAs [154,155,285,334].

Following up on the description of an nsp8- and nsp7-containing hexadecameric ring structure [153] and the nsp8-associated polymerase activity [155], we here demonstrate that the nsp(7+8) hexadecamer is the most probable conformation of the second SARS-CoV polymerase, given the near-complete association of nsp7 and nsp8 when mixed 1:1 in solution (Fig. 2F). Significant for our understanding of CoV RNA synthesis, we find that this complex is capable of binding dsRNA molecules and extending partially double-stranded RNA templates. This activity is therefore essentially comparable to the activity reported for the nsp12-RdRp [154].

A direct comparison with the nsp12 activity is difficult, however. In the course of a one-hour reaction, 0.1 μM monomeric nsp12-RdRp incorporates ~2 μM NMP into a primed $(CU)_{10}$ template [154]. The nsp(7+8) complex, at a 1 μM concentration of nsp7 and nsp8 monomers, incorporates ~1 μM NMP. Per monomer, the activity difference is therefore 20-fold, but if we assume that most nsp7 and nsp8 monomers assemble into hexadecamers and that each hexadecamer contributes only one functional active site per incorporation event, the difference would be much smaller and only ~2.5 fold. Presently, however, we do not yet have an estimate for the efficiency and stability of the nsp(7+8) complex, nor do we know the number of active sites in the complex that determine its overall activity.

Mutagenesis of nsp8 was performed to identify residues that may contribute to the catalytic centre of the nsp(7+8) polymerase, while differently tagged nsp8 recombinant proteins were constructed to explain some striking differences with previous observations. These efforts resulted in two intriguing observations. First, mutation of the conserved N-terminal D/ExD/E motif, comprising D50 and D52 in SARS-CoV, abolished RdRp activity, whereas mutation of the C-terminal motif, including SARS-CoV residues D161 and D163, did not affect polymerase activity (Fig. 5). Given the general importance of acidic residues for metal-ion coordination in polymerase active sites [89,279,281,328] and the D/ExD/E consensus sequence in coronaviruses at positions 50-52, we now postulate that these residues are part of the Mg^{2+}-binding active site in spite of the more conserved nature of D161 and D163 (Fig. 5), and their position in the nsp(7+8) structure (see below for further discussion).

Second, the presence of N-terminal extensions other than nsp7, such as ubiquitin and His_6, severely affected the primer extension activity of nsp8 (Fig. 7), potentially by changing its oligomeric state (Fig. 2). However, the relatively strong activity of nsp7-8 (Fig. 7), a potential naturally occurring replicase processing intermediate, implies that

nsp8's activity is unlikely to be directly controlled by an N-terminal cleavage event, as was observed for, *e.g.*, the poliovirus polymerase [331]. In addition, these observations suggest that a more diverse array of nsp8-containing RdRps may be involved in CoV replication and transcription.

Interestingly, our study and the published nsp(7+8) structural data [153] deviate at four main points. First, we observe that in the published nsp(7+8) crystal structure four of the eight N-terminal D/ExD/E motifs in the complex reside at the border of partially unresolved N-terminal nsp8 domains, where the coordinates of up to 49 nsp8 residues and 5 exogenous amino acids derived from the removed GST fusion partner were not determined. In light of our own finding that unnatural N-terminal extensions severely impair nsp8's RdRp activity (Fig. 5), this suggests that the published crystal structure may not represent an active conformation of the nsp(7+8) polymerase. Second, we observe that residues D50 and D52, which are both crucial for nsp(7+8) activity, are residing in an α-helix in the nsp(7+8) structure (Fig. S5), whereas in canonical primases and polymerases, the catalytic centre is preferentially located on β-strands or turns [155,328]. Third, we note that Mg^{2+} was lacking from the published nsp(7+8) crystal structure [153], even though it is required for nsp(7+8) activity. Fourth and last, we observe that a 1:1 ratio of nsp7:nsp8 is sufficient to capture all nsp8 in a higher molecular weight complex (Fig. 2F) whereas previously a 2:1 ratio was required [153], potentially due to the additional N-terminal residues that altered the dynamics of complex formation.

The (functional) implications of these observations are not clear at present, but additional structural studies will likely be required to address these issues in detail, and gain insights that may aid in explaining the *in vitro* results presented here. Likely, such experiments will also offer further information regarding the residues that are involved in nucleotide positioning, Mg^{2+} coordination and RdRp chemistry.

In summary, our results provide important novel insights into the functionality of the SARS-CoV hexadecameric nsp(7+8) complex and demonstrate its activity as an RNA polymerase. In addition, our experiments and controls revealed and address a number of disparities between previous claims and hypotheses [155], and our own observations. The "primase hypothesis" previously formulated by Imbert and co-workers [155] remains an intriguing model to explain the initiation of SARS-CoV RNA synthesis and is a topic that will be addressed in detail elsewhere. Nevertheless, based on the primer extension activity of nps7+8 on non-structured RNA templates, we can no longer exclude the possibility that nsp(7+8) may synthesise substantially longer products than mere oligonucleotide primers *in vivo*, possibly stimulated by the presence of additional viral protein factors that could,*e.g.*, provide RNA-unwinding activity. Consequently, it is now a distinct possibility that CoV RNA synthesis involves structurally different and functionally separable RNA synthesising complexes (*e.g.*, containing nsp12 or nsp(7+8)), each possessing their own dedicated RdRp characteristics and function in viral plus or

minus strand RNA synthesis. It will therefore be crucial to study whether these different polymerase activities are part of the same enzyme complex and, if so, whether they can influence each other's activity or are subject to additional control mechanisms.

Material and methods

Cloning, mutagenesis and expression

For SARS-CoV nsp7-nsp8 expression, the sequence encoding amino acids 3837-4117 of the SARS-CoV replicase pp1a was amplified by reverse transcription-polymerase chain reaction (RT-PCR) from the genome of SARS-CoV isolate Frankfurt-1 (Genbank accession number AY291315). The primers used were SAV704 and SAV429 (Table S1). For nsp8 expression, the sequence encoding pp1a residues 3920 to 4117 was amplified by RT-PCR using SAV428 and SAV429 as primers (Table S1). Both PCR products were digested with *Sac*II and *Bam*HI, and ligated into expression vector pASK3-Ub-CHis$_6$ [154]. This vector was originally derived from the pET26-Ub-CHis$_6$ vector [277], but drives expression of N-terminally ubiquitin-tagged and C-terminally His$_6$-tagged fusion proteins via a tetracyclin-inducible promoter, to rule out the potential T7 polymerase contaminations that are known to cause false positive results when using T7 promoter-driven systems for recombinant RdRp expression. All described nsp8 mutants were engineered via site-directed mutagenesis according to the QuikChange protocol (Stratagene) using the primers listed in Table S2.

For nsp7-8 or nsp8 expression, *Escherichia coli* C2523 cells (New England Biolabs) were transformed with the plasmids pASK3-Ub-nsp7-8-CHis$_6$ or pASK3-Ub-nsp8-CHis$_6$ together with the Ubp1 protease expression plasmid pCG1 [277]. Routinely, 50 ml of Luria Broth, containing ampicillin (50 µg/ml) and chloramphenicol (34 µg/ml), was inoculated 1:1000 with o/n precultures, and cells were grown to OD_{600} >0.8 at 37°C. Subsequently, the cells were slowly cooled to 20°C, followed by induction with anhydrotetracycline (Fluka) at a final concentration of 200 ng/ml for 16 h. Expression at 20°C was, however, only crucial for the preparation of certain nsp8 mutants and similar yields of active wild-type protein could be obtained by expression at 37°C for 3-4 h. Cells were harvested by centrifugation and stored at 20°C until protein purification was started.

The expression of SARS-CoV nsp7 with a C-terminal His$_6$-tag (nsp7-His) was achieved from plasmid pDEST14-nsp7-His$_6$ according to the protocol previously described for EAV nsp9 [325]. SARS-CoV nsp5-His$_6$ (nsp5-His) was expressed as a self-cleaving maltose binding protein (MBP)-fusion protein and was purified via its C-terminal His$_6$-tag [326]. The pASK3-His-nsp8 plasmid for expression of the N-terminally His$_6$-tagged nsp8 was kindly provided by Dr. Imbert and Dr. Canard (University of Marseille, France).

Purification of SARS-CoV nsp8, nsp7-8 and nsp7

Bacterial pellets were thawed on ice, resuspended in buffer A (20 mM Hepes pH 7.4, 10 mM imidazole, 0.05% Tween-20, 5 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche)) containing 500 mM NaCl, and lysed by sonication. The supernatant was cleared by ultracentrifugation at 20,000 *g* for 30 min and subsequently incubated with Talon beads (Clontech) for 2 h at 4°C. The beads were washed four times 15 min with 20 volumes of binding buffer. Ultimately, the C-terminally His_6-tagged proteins were eluted with 150 mM imidazole in buffer A containing 150 mM NaCl, or cleaved off of the column during a 3-h digestion with SARS-CoV nsp5 in the presence of 4 mM $MgCl_2$.

The eluates were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) and typically found to be >90% pure. Elution fractions containing nsp8-, nsp7-8, or nsp7 were subsequently pooled, dialysed, stored and analysed as described previously for SARS-CoV nsp12 [154].

Chemical cross-linking

To study SARS-CoV nsp(7+8) complex formation, different nsp8:nsp7 ratios were mixed in binding buffer (20 mM Hepes pH 7.5, 50 mM NaCl, 5% glycerol, 0.1% Triton X-100, and 1 mM DTT) to give a final reaction volume of 10 µl. The proteins were pre-incubated for 10 min at 20°C, after which cross-linking was initiated through the addition of 0.5 µl of a freshly prepared 2.5% glutaraldehyde solution. The reactions were incubated for a further 5 min at 30°C and then terminated with 1 µl 1 M Tris pH 8.0. Analysis of complex formation was performed on SDS-PAGE gels, which were stained with Coomassie G-250 dye.

Template binding assays

A dilution series of 0-5 µM SARS-CoV nsp8 in storage buffer was incubated for 10 min at 20°C with 0.2 nM of ^{32}P-labelled duplex RNA. Subsequently, samples were directly loaded onto 8% polyacrylamide gels containing 5% glycerol and 0.5x TGE (25 mM Tris, 190 mM glycine and 10 mM EDTA) buffer and run at 150 V for 1 h at 4°C. Gels were dried on Whatman filter paper and bands were quantified by phosphorimaging using a Typhoon variable mode scanner (GE Healthcare) and ImageQuant TL 7.0 software (GE Healthcare) as described elsewhere [154]. Using the Matlab 2009a Curve Fitting Toolbox, the percentage of bound RNA was fit to the Hill equation, which is defined as:

$RNA_{bound} = b \times [nsp8]^n / (K_d^n + [nsp8]^n)$.

Here *b* is the upper binding limit, *[nsp8]* the nsp8 concentration, *n* the Hill coefficient and K_d the dissociation constant.

Polymerase activity assays

The oligoribonucleotide substrates used for polymerase assays are listed in Table 1 and were prepared as described previously [154]. Primer-extension assays for nsp8, the nsp7-8 polyprotein, and the nsp(7+8) complex were essentially performed as described previously for SARS-CoV nsp12 [154,325]. In each primer-extension reaction, typically 1 μM wild-type or mutant nsp8 was incubated with 4 mM $MgCl_2$, 50 μM GTP, 50 μM ATP, 0.17 μM [α-^{32}P]ATP, 1mM DTT, 0.1% Triton X-100, 10 mM KCl and 20 mM Tris (pH 9.5). At most 10 mM NaCl and 5% glycerol were introduced with the nsp8 storage buffer. Gels were run and analysed as described previously [154]. To convert the phosphorimager signal into the amount of [α-^{32}P]AMP incorporated, a 10^{-2} to 10^{-5} dilution series of the [α-^{32}P]ATP stock was spotted in triplicate on Whatman filter paper and exposed alongside the PAGE gel. The amount of incorporated label was ultimately corrected for the concentration of competing, unlabelled nucleotides present in the reaction mixture.

De novo initiation assays were essentially performed as described by Imbert *et al.* [155], with small modifications for optimisation. Briefly, 1 μM wild-type or mutant nsp8 was incubated with 4 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM GTP, 5 μM ATP, 0.17 μM [α-^{32}P]ATP and 1 μM of oligo AFMB131.

Sequence alignment

Alignments of nsp8 sequences were made using Muscle [327]. Sequences used included the alphacoronaviruses human CoV 229E (NC_002645), human CoV NL63 (NC_005831), and bat CoV HKU8 (NC_010438); the betacoronaviruses SARS-CoV Frankfurt-1 (AY291315), mouse hepatitis virus A59 (MHV, NC_001849), and human CoV OC43 (NC_005147); and the gammacoronaviruses beluga whale CoV SW1 (NC010646), turkey CoV (NC_010800), and avian infectious bronchitis virus (IBV, AJ311317).

Table 1: Oligoribonucleotides used for activity assays

RNA oligo	Purpose	Sequence
SAV555	$(UC)_{10}$ template	5′-UUUUUUUUUUUUUUUUUUUUAUAACUUAAUCUCACAUAGC-3′
SAV556	$(U)_{20}$ template	5′-UCUCUCUCUCUCUCUCUCUCAUAACUUAAUCUCACAUAGC-3′
SAV557	primer	5′-GCUAUGUGAGAUUAAGUUAU-3′
AFMB131	*de novo* assay template	5′-UAUAAUCCAAAA-3′

Acknowledgements

The authors thank Dr. Danny Nedialkova, Lorenzo Subissi, Dr. Isabelle Imbert, Dr. Bruno Canard, and Dr. Alexander Gorbalenya for stimulating discussions; Linda Boomaars-van der Zanden and Dr. Clara Posthuma for assistance with nsp5 purification; Puck van Kasteren and Dr. Marjolein Kikkert for providing the EAV nsp2 protease; and Jos van Vugt for his initial work on nsp8 in our lab. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037 and ECHO grant 700.55.002 from the Council for Chemical Sciences (NWO-CW).

CHAPTER 5 - SUPPLEMENTAL INFORMATION

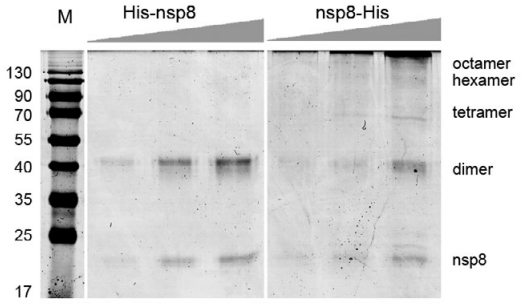

Figure S1:Analysis of nsp8 multimerisation via chemical cross-linking. Cross-linking analysis of N-terminally His_6-tagged nsp8 and C-terminally His_6-tagged nsp8. Increasing concentrations of nsp8 were incubated in Hepes buffer (pH 7.5) in the presence of 0.12% glutaraldehyde for 5 min. Subsequent SDS-PAGE analysis and staining with Coomassie G-250 dye shows that only the C-terminally tagged protein with native N-terminus forms higher order multimers, whereas the N-terminally tagged protein reveals solely mono- and dimers.

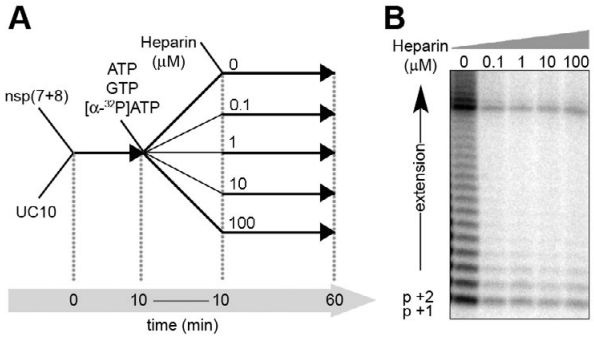

Figure S2: Inhibition of nsp(7+8) activity with heparin. (**A**) Schematic presentation of the single-cycle reaction. Template and nsp(7+8) complex were pre-incubated for 10 min before nucleotides were added. The mixture was then rapidly split into equal aliquots that were immediately mixed with different concentrations of heparin. (**B**) Samples were taken after 60 min and resolved on 20% PAGE/7M Urea.

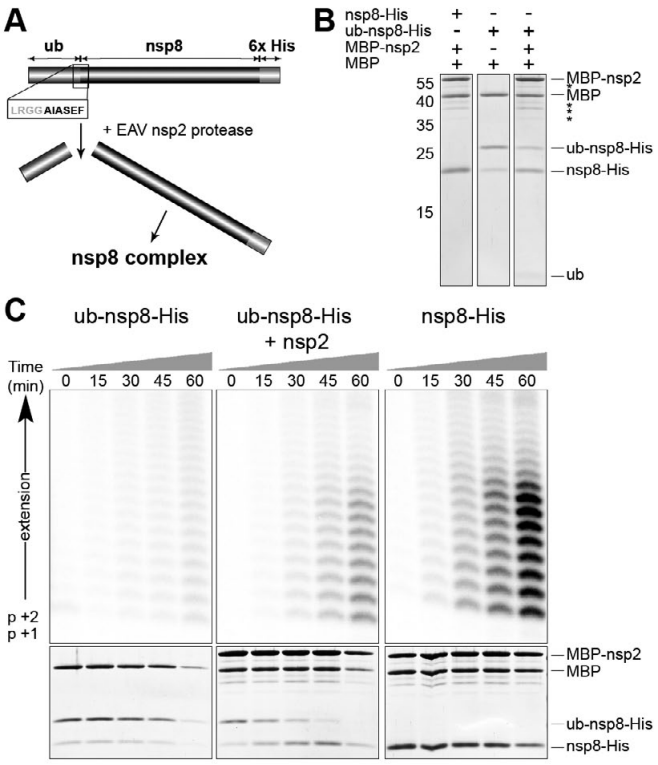

Figure S3: Activation of nsp8 RdRp activity by removal of an N-terminal fusion partner. (**A**) To study whether the interference by non-nsp7 N-terminal extensions was reversible, we purified nsp8 with an N-terminal ubiquitin extension. Addition of either purified EAV nsp2 or Ubp1 would subsequently result in hydrolysis of the fusion protein C-terminal of the LRGG site. (**B**) Analysis of ub-nsp8-His cleavage products by SDS-PAGE and Coomassie G-250 staining. Asterisks indicate unspecific bands (**C**) Time-course of [α-^{32}P]AMP incorporation of ub-nsp8-His, ub-nsp8-His in the presence of EAV nsp2, or nsp8-His that was *in vivo* cleaved by Ubp1. Lower panels demonstrate the stability or cleavage of ub-nsp7-8-his over time through SDS-PAGE analysis. In all reactions, MBP was added an independent loading control.

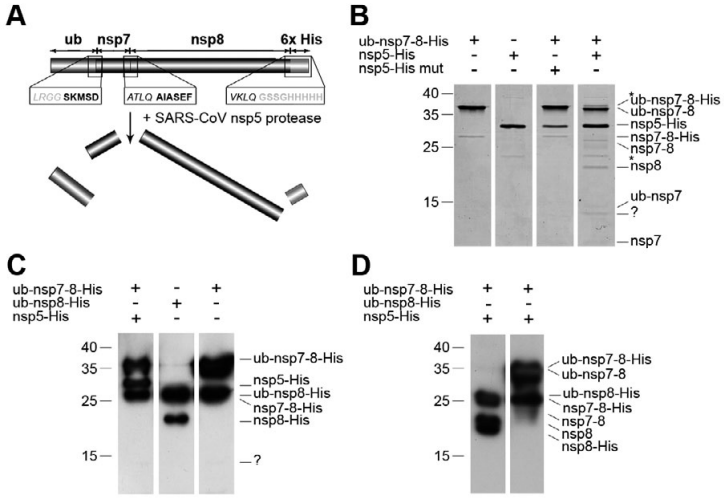

Figure S4: Processing of ub-nsp7-8-His by the SARS-CoV nsp5 main protease. (**A**) Due to the presence of the natural nsp7-8 cleavage sites in the ub-nsp7-nsp8-His construct, the polyprotein could be processed into mature subunits by the SARS-CoV nsp5 main proteinase. The question mark indicates an unidentified cleavage product. (**B**) SDS-PAGE analysis followed by Coomassie G-250 dye staining demonstrates that addition of nsp5 to purified ub-nsp7-nsp8-His results in cleavage at the sites indicated in Fig. 2C. Asterisk indicates non-specific band. The question mark indicates an unidentified cleavage product. (**C**) Western blot analysis of the protein samples used in Fig. 2D using an anti-His_6 monoclonal antibody and (**D**) an anti-nsp8 monoclonal antibody.

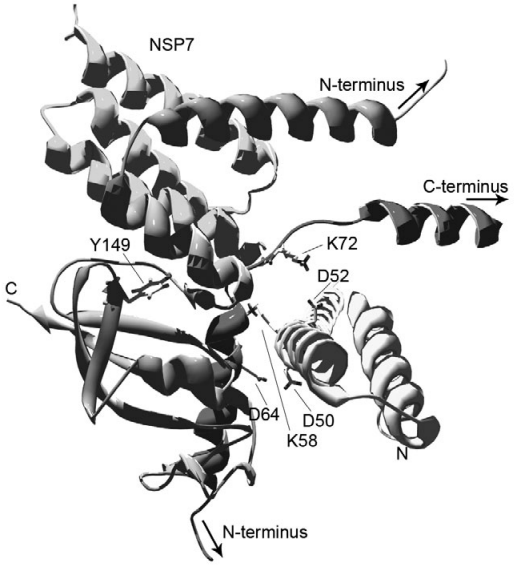

Figure S5: Position of (putative) critical residues in the published nsp(7+8) structure. Position of the critical residues, including D50, D52 in the nsp(7+8) hexadecamer. For reference, the conserved residues K58A, D62, K72A, and Y149, and the location of the termini are also indicated. The residues of the different termini are hidden to simplify the image. The nsp7 subunit that is in direct contact with the three nsp8 subunits is shown in the top left part of the structure.

Table S1: DNA primers that were used for cloning of nsp8 and nsp7-8

Primer	Sequence
SAV704	5′-GCGGGTACCCCGCGGTGGATCTAAAATGTCTGACGTAAAGTGCACA-3′
SAV429	5′-GCGCGATCGGGATCCCTGTAGTTTAACAGCT-3′
SAV428	5′-GCGGGTACCCCGCGGTGGAGCTATTGCTTCAGAAT-3′

Table S2: DNA primers that were used for mutagenesis of nsp8

nsp8 mutation	PCR primers	Sequence
D50A	SAV574	5′-GCTAAATCTGAGTTTGCCCGTGATGCTGCCATG-3′
	SAV575	5′-CATGGCAGCATCACGGGCAAACTCAGATTTAGC-3′
D52A	SAV590	5′-TCTGAGTTTGACCGTGCTGCTGCCATGCAACGC-3′
	SAV591	5′-GCGTTGCATGGCAGCAGCACGGTCAAACTCAGT-3′
K58A	SAV402	5′-GCCATGCAACGCGCTTTGGAAAAGATGG-3′
	SAV403	5′-CCATCTTTTCCAAAGCGCGTTGCATGGC-3′
K127A	SAV501	5′-GACTACAGCAGCCGCACTCATGGTTGTTG-3′
	SAV502	5′-CAACAACCATGAGTGCGGCTGCTGTAGTC-3′
D161A	SAV503	5′-CCAGCAAGTTGTTGCTGCGGATAGCAAGA-3′
	SAV504	5′-TCTTGCTATCCGCAGCAACAACTTGCTGG-3′
D163A	SAV505	5′-AGTTGTTGATGCGGCTAGCAAGATTGTTC-3′
	SAV506	5′-GAACAATCTTGCTAGCCGCATCAACAACT-3′