Biochemical characterization of the nidoviral replicase complex

Inauguraldissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
des Fachbereichs 08 - Biologie und Chemie
der Justus-Liebig-Universität Gießen

vorgelegt von
Ganesh Bylapudi
Master of Science (M.Sc.)

Gießen, Juli 2021

Die vorliegende Arbeit wurde am Institut für Medizinische Virologie des Fachbereiches 11 (Medizin) der Justus-Liebig-Universität Gießen in der Zeit von September 2014 bis Juli 2020 unter der Leitung von Prof. Dr. John Ziebuhr angefertigt.

1. **Gutachter:** Prof. Dr. Peter Friedhoff
   Institut für Biochemie
   Justus-Liebig-Universität Gießen

2. **Gutachter:** Prof. Dr. John Ziebuhr
   Institut für Medizinische Virologie
   Justus-Liebig-Universität Gießen

Abstract

Nidoviruses are a monophyletic group of RNA viruses with exceptionally large genomes ranging from 12 - 41 kb. Despite their widely varying genome sizes, members of the order *Nidovirales* share common characteristic features, such as an array of conserved domains in their replicase genes and the expression of structural and accessory proteins from a nested set of sub-genomic mRNAs. The replicase gene occupies the 5'-terminal two-thirds of the genome and encodes for nonstructural proteins (nsp) that form active replication-transcription complexes (RTC) and have other important functions in the viral replication cycle. Previously, attempts to develop *in vitro* polymerase assays for coronaviruses and a few other nidoviruses resulted in contradictory results. Reverse genetics and biochemical studies have suggested that, in coronaviruses, nsp7, nsp8 and nsp12 are the key factors required for RNA synthesis *in vitro*. In this work, recombinant and tag-less forms of nsp7, nsp8 and the pre-assembled nsp(7+8) complex were produced and used to investigate the polymerase activity of nsp12, a two-domain protein harboring the canonical RNA-dependent RNA polymerase (RdRp) domain in its C-terminal region. The studies revealed that a pre-assembled coronavirus nsp(7+8)+nsp12 complex was required for efficient elongation activity from partially double-stranded ('primed') RNA substrates in the presence of Mg^{2+} or Mn^{2+}. In the absence of nsp12, human coronavirus 229E (HCoV-229E) nsp8 and a pre-assembled nsp(7+8) complex exhibited RNA 3'-terminal adenylyltransferase activity on different RNA substrates but failed to produce complementary copies of heteropolymeric template RNAs. Production and characterization of nsp7/8/12 complexes of representative alpha- and betacoronaviruses revealed that all these complexes were able to initiate RNA synthesis *de novo* from single-stranded template RNAs in the presence of Mn^{2+} but not Mg^{2+}, suggesting that coronavirus replicase complexes do not strictly depend on the presence of (RNA or protein) primers to initiate

RNA synthesis, at least *in vitro*. With very few exceptions, RdRp activity of nsp12 was revealed to require the presence of the cognate nsp(7+8) complex, while nsp(7+8) complexes from other coronavirus species failed to attain RdRp activity, suggesting important functional and/or structural constraints in the formation of active nsp7/8/12 complexes. Amino acid substitutions in the conserved RdRp motifs A and C, respectively, abolished both *de novo* and primer-dependent activity. At physiological Mg^{2+} or Mn^{2+} concentrations, the HCoV-229E nsp(7+8)+nsp12 complex was shown to differentiate between rNTPs and dNTPs and to incorporate the correct nucleotides during RNA elongation. Further investigations using a range of alpha- and betacoronavirus nsp(7+8) complexes produced in *E. coli* provided evidence that feline infectious peritonitis virus (FIPV) and transmissible gastroenteritis virus (TGEV) nsp(7+8) predominantly form heterotrimeric complexes (2:1), while HCoV-229E, PEDV, SARS-CoV and MERS-CoV produce predominantly heterotetrameric nsp(7+8) complexes (2:2). In all these cases, nsp(7+8) heterodimeric (1:1) complexes were detected, suggesting that they form the building block for larger complexes in all coronaviruses.

In another part of the study, the arterivirus equine arteritis virus (EAV) nsp9 (harboring the RdRp domain) was characterized. Although the present study, in line with previous studies, failed to provide conclusive evidence for nsp9-mediated RdRp activity, the study showed that the protein has RNA 3'-terminal adenylyltransferase activity and employs its N-terminal NiRAN domain to transfer UMP and GMP, respectively, to the N-terminus of the arteriviral nsp7 or nsp7α. Mass spectrometry analyses confirmed that nsp9 mediates the transfer of one molecule of NMP to the acceptor protein.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. John Ziebuhr for giving me this excellent opportunity, guiding me throughout my doctoral thesis and introducing me to the field of virology and biochemistry at the Institute of Medical Virology at the Justus Liebig University Giessen. I will forever be grateful for the time he spent discussing the data and giving me suggestions for improvement. I also thank him for his excellent teaching, inspiration and showing me that a person can be a successful scientist and thank him for always being there when I had questions or moments of doubts. He supported me not only in my research but also with personal decisions, for which I am grateful. Finally, I am lucky to have him as my doctoral mentor.

I must show gratitude to my daily supervisor Dr. Ramakanth Madhugiri for his support, ideas, spending time throughout data discussion, practice talks and problem solving during my doctoral studies.

It has been a great honor for me to be a member of the DFG-supported IRTG1384 of the Universities of Giessen and Moscow and to receive extensive training and financial support for the initial two years of my doctoral studies. I would like to thank Professors Pingoud and Friedhoff for coordinating this PhD training program. I would also like to thank SFB 1021 for financial support for the last four years. I would also express my gratitude to Prof. Dr. Peter Friedhoff for his interest in my work and for helpful advice and discussions.

I would like to thank former and present members of the laboratory Nadja Karl, Ulrike Wend, Karin Schultheiß, Marc Sauerwald, Dr. Alexandra Friedrich, Dr. Christin Müller, Dr. Heiko Slanina, Dr. Jana Tvaragová and Dr. Sandra Blanck of the Institute of Medical Virology for all their help and support.

I would like to thank Dr. Boris Krichel, Dr. Charlotte Uetrecht for analyzing the coronaviruses proteins by native mass spectrometry (Heinrich Pette Institute, Hamburg).

Also, I would like to thank Dr. Uwe Linne and Dr. Heiko Slanina for analyzing equine arteritis virus proteins by mass spectrometry.

Really special thanks to Prof. Dr. Vikas Jain and Prof. Dr. R. Mahalakshmi, my family and friends for supporting and helping me during my post-MSc and PhD studies.

Abbreviations

$3CL^{pro}$	3C-like protease
6-TG	6-thioguanine
Ala	Alanine
AMP	Adenosine-5’-monophosphate
ATP	Adenosine-5'-triphosphate
BCoV	Bovine coronavirus
bp	base pair
CavV	Cavally virus
cDNA	Complementary DNA
CoV	Coronavirus
CPE	cytopathic effect
CTP	Cytidine-5'-triphosphate
CV	column volumes
Da	Dalton
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytosine-5'-triphosphate
DEPC-H_2O	Diethylpyrocarbonate water
ddH_2O	Double deionized water
dGTP	2'-deoxyguanosine-5'-triphosphate
DMV	Double-membrane vesicle
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside-5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DTT	Dithiothreitol

E	Envelope
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
ER/GIC	Endoplasmic reticulum/Golgi intermediate compartment
ExoN	Exoribonuclease
FCoV	Feline coronavirus
GTP	Guanosine-5'-triphosphate
HCoV	Human coronavirus
HEL	Helicase
IBV	Infectious bronchitis virus
ICTV	International Committee on Taxonomy of Viruses
LB	Luria-Bertani
M	Membrane
MERS	Middle East respiratory syndrome
MHV	Mouse hepatitis virus
min	Minute
M^{pro}	Main protease
mRNA	Messenger RNA
MT	Ribose-2'-*O* methyltransferase
N7-MTase	guanosine-N7 methyltransferase
N	Nucleocapsid
NendoU	Nidoviral endoribonuclease (uridylate specific)

Ni-NTA	Nickel-nitrilotriacetic acid
NiRAN	Nidovirus RdRp-associated nucleotidyltransferase
NMP	Nucleoside monophosphate
NMR	Nuclear magnetic resonance
nsp	Nonstructural protein
nt	Nucleotide
NTase	Nucleotidyltransferase
NTP	Nucleoside-5'-triphosphate
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
PL^{pro}	Papain-like protease
PNK	Polynucleotide kinase
pp	Polyprotein
PRF	Programmed ribosomal frameshifting
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
rNMP	Ribonucleoside 5'-monophosphate
RTC	Replicase/transcriptase complex
S	Spike
SARS	Severe acute respiratory syndrome
SDS	Sodium dodecyl sulfate
sec	Seconds
sg	Subgenomic
ss	Single stranded

TAE	Tris-acetate-EDTA
TATase	Terminal adenylyltransferase
TEMED	Tetramethylethylenediamine
TGEV	Transmissible gastroenteritis virus
TM	Transmembrane
TNTase	Terminal nucleotidyltransferase
Tris	Tris-(hydroxymethyl)-aminomethane
TRS	Transcription-regulating sequence
Ubl	Ubiquitin-like domain
Ubp1	Ubiquitin-specific protease 1
UTP	Uridine-5'-triphosphate
UTR	Untranslated region
UV	Ultraviolet
v/v	Volume per volume
w/v	Weight per volume
wt	wild type

Table of Contents

1. Introduction

1.1 Nidoviruses

Nidoviruses are a genetically diverse virus group with complex genome organizations and the largest known genomes amongst RNA viruses. They are enveloped viruses with single-stranded, positive-sensed RNA (+ssRNA) genomes ranging from 12-41 kb (Saberi et al., 2018). Although considerably different in genome size, nidoviruses are similar in their genome organization. They possess a linear 5'-capped, non-segmented RNA, with the two large open reading frames (ORFs) 1a and 1b located in the 5'-proximal region of the genome encoding for most of the nonstructural proteins (nsp) (Cowley et al., 2000; den Boon et al., 1991; Gorbalenya et al., 2006; Gorbalenya et al., 1989b; Snijder et al., 2003). These nsps have a varity of enzymatic activities that, to a large extent, are involved in viral RNA replication and transcription. Members of the order *Nidovirales* employ special mechanisms in their RNA replication and transcription strategies (Lai and Cavanagh, 1997; Ziebuhr, 2004). These include the generation of a nested set of subgenomic (sg) mRNAs, from which the name *Nidovirales* (the Latin word *nidus* = nest) was derived. The structural proteins are expressed from this 3'-coterminal nested set of sg mRNAs. Several viruses from this order have attracted much attention in the past and present for being the causative agents of serious diseases in humans and animals. In 2003 and 2012, respectively, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) have been shown to be associated with severe respiratory disease and high fatality rates in humans (Song et al., 2019) and 2019 the outbreak of Covid-19 caused by SARS-CoV-2 (Coronaviridae Study Group of the International Committee on Taxonomy of, 2020; Zhu et al., 2020). Studies indicate that bats are the natural reservoirs and are likely to play an important role in the introduction of coronaviruses

to other species as well as the evolution and dissemination of coronaviruses (Tong et al., 2009). The first arterivirus to be isolated was the Equine arteritis virus in 1953, known to infect horses and donkeys. Arterivirus infections can vary from acute symptomatic or asymptomatic infection, abortion or lethal haemorrhagic fever and persistent infection.

1.2 Taxonomy

The taxonomic organization of the order *Nidovirales* was revised in 2018 and 2019 by the International Committee on the Taxonomy of Viruses (ICTV). According to these recent revisions, the order *Nidovirales* currently comprises 14 virus families, which are the *Abyssoviridae*, *Arteriviridae*, *Cremegaviridae*, *Gresnaviridae*, *Olifoviridae*, *Coronaviridae*, *Medioniviridae*, *Mesoniviridae*, *Mononiviridae*, *Nangoshaviridae*, *Nanhypoviridae*, *Euroniviridae*, *Roniviridae*, and *Tobaniviridae* (Siddell et al., 2019; Walker et al., 2020).

The *Coronaviridae* family is divided into the subfamilies *Letovirinae* and *Orthocoronavirinae*, with the latter being further subdivided into four genera (*Alpha-*, *Beta-*, *Gamma-* and *Deltacoronavirus*) and numerous subgenera according to the degree of sequence similarity in the 5 most conserved domains of the viral replicase genes (Coronaviridae Study Group of the International Committee on Taxonomy of, 2020). The first described coronavirus was the avian infectious bronchitis virus (IBV) in the 1930's (Hudson and Beaudette, 1932) and up to date seven human coronaviruses have been identified (HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, SARS-CoV-2 and MERS-CoV) (Zaki et al., 2012).

According to latest classification, the *Arteriviridae* family consists of the six subfamilies *Equarterivirinae*, *Variarterivirinae*, *Simarterivinae*, *Heroarterivirinae*,

Crocarterivirinae and *Zealarterivirinae* (Siddell et al., 2019). Equine arteritis virus (EAV) belongs to the subfamily *Equarterivirinae* and genus *Alphaarterivirus* (Table 1).

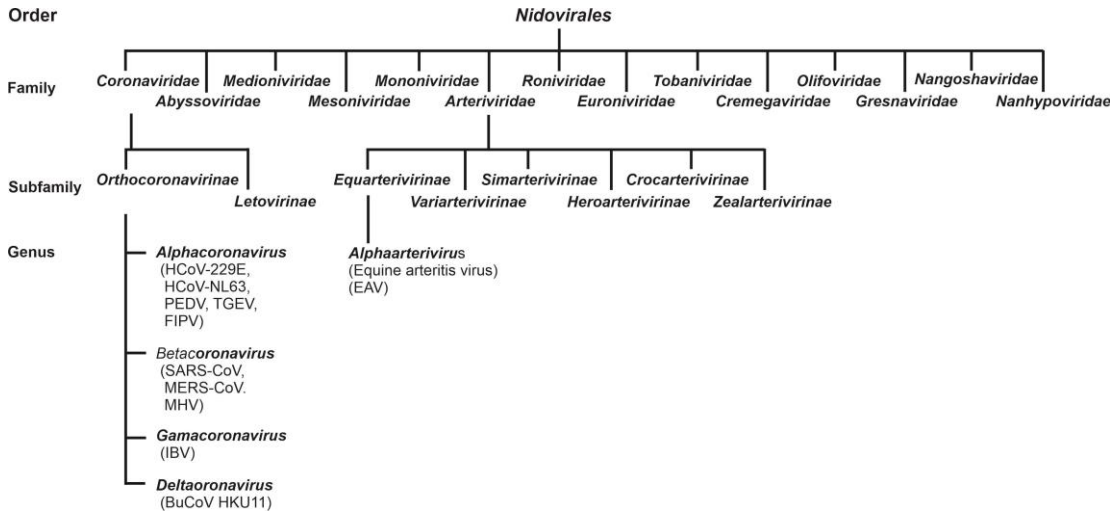


Table 1. Schematic representation of the taxonomy of the order *Nidovirales*. Family *Coronaviridae*; subfamily *Orthocoronavirinae*; Virus name abbreviations used: HCoV-229E, human coronavirus 229E; HCoV-NL63, human coronavirus NL63; PEDV, porcine epidemic diarrhea virus; TGEV, transmissible gastroenteritis virus; FIPV, feline infectious peritonitis virus; SARS-CoV, severe acute respiratory syndrome coronavirus; MERS-CoV, Middle East respiratory syndrome coronavirus, MHV, mouse hepatitis virus; IBV, avian infectious bronchitis virus; BuCoV-HKU11, bulbul coronavirus HKU11. Family *Arteriviridae*; Subfamily *Equaarterivirinae*, Genus *Alphaarterivirus*, Species *Alphaarterivirus equid*. Abbreviation used: EAV, equine arteritis virus.

1.3 Morphology of corona- and arteriviruses

1.3.1 Coronaviruses

Coronavirus virions range from 60-220 nm in diameter and have a pleomorphic, although generally spherical shape (Siddell et al., 1982, 1983). The virus contains four to five structural proteins named spike protein (S), membrane protein (M), envelope protein (E), nucleocapsid protein (N) and, in some betacoronaviruses, a virus-encoded hemagglutinin-esterase protein (HE) is part of the virus particle (de Haan et al., 2005) (Figure 1).

The viral genome associates with the N protein as a helical nucleocapsid which interacts with the viral membrane protein M (Stadler et al., 2003). M and E proteins are

transmembrane proteins that assist in viral morphogenesis, assembly, and are essential for the budding of the virion (Boscarino et al., 2008; Neuman et al., 2011; Opstelten et al., 1995; Westerbeck and Machamer, 2015). Only small amounts of the highly hydrophobic envelope proteins are present in virions. It has been suggested that the E protein induces membrane curvature (Kuo et al., 2007; Kuo and Masters, 2003), possesses ion channel activity required for optimal virus replication (Wilson et al., 2004) and appears to play a role in the induction of apoptosis (An et al., 1999). The minor envelope protein (E) is necessary for virus particle assembly (Stadler et al., 2003).

The membrane protein (M) is anchored in the lipid envelope and has two conformations, elongated and compact, which give the coronavirus its spherical shape (Neuman et al., 2011).

The S protein appears like the spikes on a crown in electron microscopy, giving coronaviruses their name (lat. *corona* = crown). The spike protein interacts with the cell receptor and is comprised of three structural domains: a large N-terminal domain that is further split into the S1 and S2 domains, a transmembrane domain and a small C-terminal (internal) domain. These extermal domains are cleaved during virus maturation by one or more cellular proteases. S1 is responsible for receptor binding and its sequence is more variable than S2. The S2 subunit is critically involved in structural rearrangements during the fusion process of the viral and cellular membranes and is the more conserved region of the protein (Belouzard et al., 2012).

1.3.2 Arteriviruses

The particles of arteriviruses are spherical with a diameter of 40 to 60 nm and feature a 25-to-35 nm icosahedral core surrounded by a relatively smooth envelope that lacks large projections. The lipid bilayer surrounding the nucleocapsid and six envelope proteins: E, GP2, GP3, GP4, GP5 and M (de Vries et al., 1992; Hyllseth, 1973; Zeegers

et al., 1976) (Figure 1). Unlike coronaviruses, arteriviruses do not possess a spike protein but encode the E protein and a GP2–GP3–GP4 heterotrimer which may play a role in receptor binding or virus entry (Van Breedam et al., 2010a; Van Breedam et al., 2010b). Only the nucleocapsid protein (N) along with viral RNA and GP5/M dimer are required for budding. The E protein appears to be an ion-channel protein that may play a role in the uncoating process during virus entry.

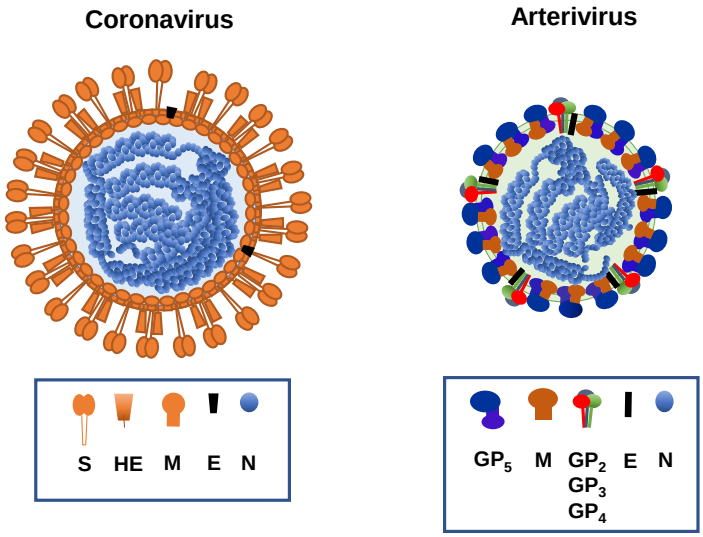


Figure 1: Schematic representation of a corona- and arterivirus particle. S = spike protein; HE = hemagglutinin-esterase; M = membrane protein; E = envelope protein; N = nucleocapsid protein; GP = glycoprotein. Modified from (de Groot, 2012).

1.4 The nidoviral life cycle

The virus entry into the host cell occurs either via direct fusion with the plasma membrane or by endocytosis and subsequent fusion of the viral and endosomal memebrane (Figure 2). Cell entry is initiated by interactions of the receptor-binding protein interacts with its specific host cell receptor (Gallagher and Buchmeier, 2001). After entry, the viral nucleocapsid disassembles and the released genome with its 5'-methyl cap structure and 3'-poly A tail, mimicking eukaryotic mRNA, is translated. In this process, the first two open reading frames present on the viral genome (ORFs 1a and 1b) are translated into two large polyproteins (Sawicki et al., 2005; Weiss and

Navas-Martin, 2005). The replicase ORF1a is translated into the polyprotein 1a (pp1a) while translation of ORF1b involves a -1 ribosomal frameshift just upstream of the ORF1a termination codon which then results in the synthesis of polyprotein 1ab (pp1ab) (Bredenbeek et al., 1990; Brierley and Dos Ramos, 2006). The two replicase polyproteins are processed by viral proteases into 15 to 16 non-structural proteins (nsps) in coronaviruses and 12 to 13 (major) nsps in arteriviruses (Liu and Brown, 1995; Tibbles et al., 1996; Ziebuhr and Siddell, 1999; Ziebuhr et al., 2000). These matured non-structural proteins assemble into replication/transcription complexes (RTC) which are anchored onto modified ER-derived double-membrane vesicles (DMVs) in the cytoplasm which is where viral transcription/replication takes place (Angelini et al., 2013; Gosert et al., 2002; Knoops et al., 2008; Oudshoorn et al., 2017; Snijder et al., 2006; van Hemert et al., 2008a).

The ORFs located in the 3'-proximal genome region encode for the viral structural and accessory proteins which are translated from a set of subgenomic mRNAs. In all nidoviruses, full-length minus-strand RNA is used as template for the production of progeny genome RNAs (= RNA replication), while the production of sg mRNAs (= transcription) is templated by a set of subgenome-length, negative-strand RNAs that are produced in a process called discontinuous RNA synthesis. In corona- and arteriviruses, these subgenomic negative-strand RNAs contain at their 3’ end a copy of the 5’-leader present on the viral genome (Pasternak et al., 2006; Sawicki et al., 2007; van Marle et al., 1999). The initiation of minus-strand RNA synthesis occurs after recognition of RNA signals close to the 3'-end of the viral genome by the RTC. The resulting minus-strand sg RNAs then serves as template for the transcription into the corresponding sg mRNAs. These sg mRNAs express the structural proteins. In some cases, the expression of one or more additional ORFs from a single sg mRNA has been described.

After translation of the structural proteins, the genomic RNA and the N protein form the nucleocapsid. The main location of coronavirus particle formation is the Golgi complex, although during early and late infection budding can also occur from the ERGIC and rER (Salanueva et al., 1999). The virus particles are transported through the Golgi compartment to secretory vesicles for subsequent release of the virions at the cell surface through exocytosis.

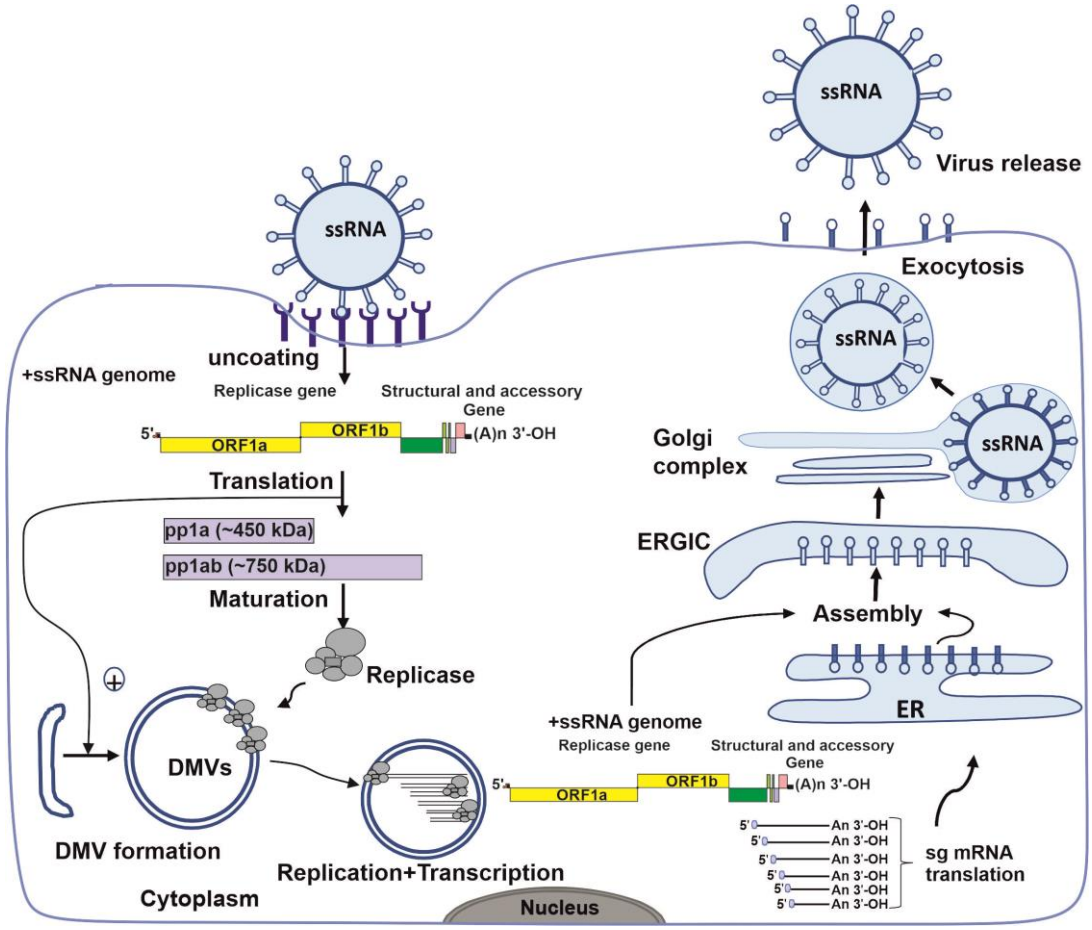


Figure 2: Schematic overview of the nidoviral life cycle. The nidoviral life cycle starts with the attachment of the virion to the cellular receptor. After fusion of the viral and the host cell membrane, the nucleocapsid disassembles releasing its genomic RNA into the host cytoplasm. The genomic RNA acts as mRNA for the expression of the replicase gene comprised of ORFs 1a and 1b. ORF1a/ORF1b are translated into two large polyproteins (pp1a/pp1ab). pp1ab is produced via a programmed -1 ribosomal frame shift occurring just upstream of the ORF1a stop codon. pp1a/pp1ab are processed by viral proteases into individual nonstructural proteins (nsps). Some of the cleaved products are inserted into the endoplasmic reticulum membrane where they induce the formation of double membrane vesicles (DMV) on which the replication-translation complexes (RTC) assemble. The RTC produces negative-stranded genomic and subgenomic RNAs that are used as templates for the synthesis of (plus-stranded) genomic and subgenomic RNAs. The structural proteins are translated from the nested set of subgenomic mRNAs. The nucleocapsid protein encapsidates the genomic RNA to form the nucleocapsid which interacts with the viral M protein. Both the M protein and the other structural proteins (S,

E in coronaviruses) are inserted into membranes of the ER-Golgi intermediate compartment (ERGIC). Induction of membrane curvature and budding into the ERGIC eventually results in the formation of viral particles. Finally, the virions are transported out of the cell via the cellular secretory pathway. Modified from (de Haan and Reggiori, 2008).

1.5 Genome organization

Coronavirus and arterivirus genomes contain a 5' leader sequence which is part of the 5’ untranslated region (5'UTR). The 3'-end also contains an untranslated region (3'UTR) followed by a poly(A) tail (Figure 3). The first gene (replicase) comprises two-thirds of the genome with the translation of the two large ORFs (ORF 1a and ORF 1b) resulting in polyprotein 1a and polyprotein 1ab via a -1 frame shifting mechanism involving a hairpin pseudoknot structure formed by the genomic RNA (Bredenbeek et al., 1990; Brian and Baric, 2005; Brierley and Dos Ramos, 2006; Lee et al., 1991; Namy et al., 2006; Ziebuhr, 2005). The 3’-terminal one-third of the genome contains the ORFs coding for the structural and accessory proteins.

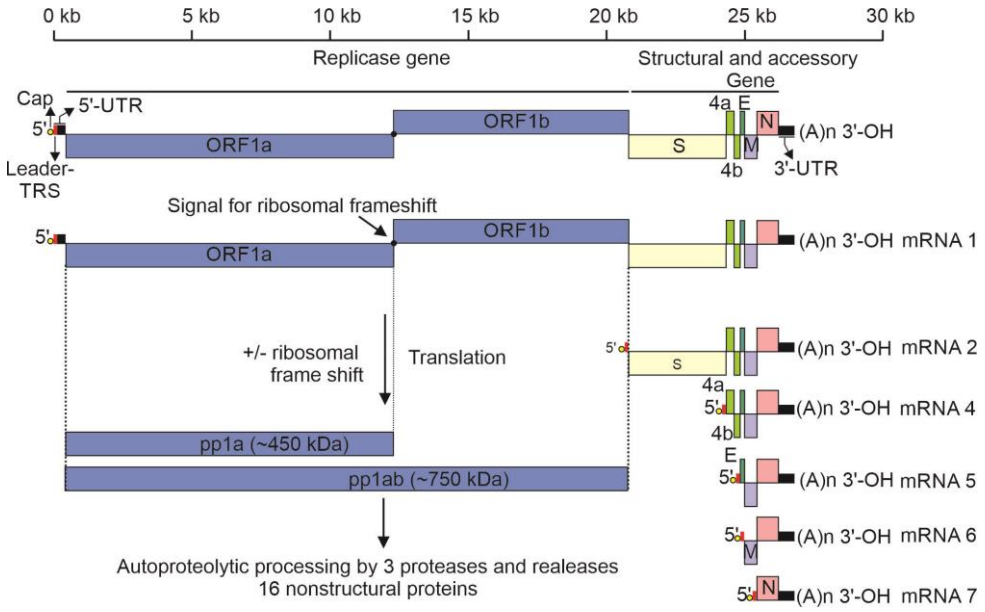


Figure 3: Coronavirus genome organization using HCoV-229E as an example. Open reading frames are indicated with boxes. The small red box indicates the 5'-leader sequence present on each of the viral plus-strand RNAs produced in infected cells. The 3'-terminal third of the genome encodes the structural and accessory proteins. The replicase gene is comprised of the two large open reading frames ORF 1a and 1b, encoding the polyproteins pp1a and pp1ab that are proteolytically cleaved into individual nonstructural proteins (nsps). Expression of ORF1b sequences requires a programmed -1 ribosomal frameshift (ribosomal frameshift site is indicated by black circle). The 5'-leader sequence (L), 5'-cap structure, 3'-poly(A) tail (A_n) are indicated.

1.6 Nonstructural proteins

The nonstructural proteins required for the replication/transcription complex (RTC) are encoded by the replicase gene, which is expressed from two overlapping ORFs. The first N-terminal proteins of pp1a and pp1ab are released by one or two papain-like proteinase(s) (PL^{pro}) whereas the main proteinase (called M^{pro} or $3CL^{pro}$) cleaves out the C-terminal non-structural proteins (Ziebuhr et al., 2007; Ziebuhr et al., 2000) (Figure 4). ORF1a-encoded enzymes include viral proteases, several anti-host activities, membrane-anchoring domains, RNA-binding, and RNA-modifying activities while ORF1b mainly encode enzymes that are more directly involved in RNA replication, transcription, strand separation, proofreading and capping activities (Decroly et al., 2008).

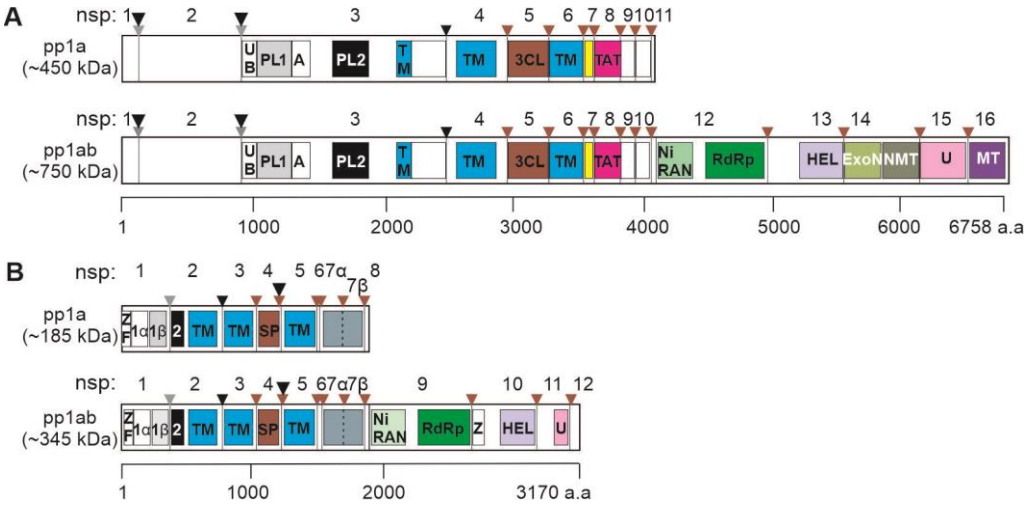


Figure 4: Domain organization of polyprotein 1a/1ab of the coronavirus HCoV-229E (A) and the arterivirus EAV (B).
(A) Polyprotein cleavage sites are depicted with arrowheads matching the color of the proteinase involved (papain-like protease 1 [$PL1^{pro}$, grey arrowheads]; papain-like protease 2 [$PL2^{pro}$: black arrowheads]). Brown arrowheads indicate $3CL^{pro}$ cleavage sites. Proteolytic processing by the viral proteases results in a total of 16 nonstructural proteins (nsp1 to 16) for HCoV-229E. Previously characterized functions of individual nsps are indicated as follows: UB = ubiquitin-like domain; $PL1^{pro}$ and $PL2^{pro}$ = papain-like protease 1 and 2; A = ADP-ribose-1''-phosphatase (also called macrodomain); TM = transmembrane domain; 3CL = 3C-like main protease ($3CL^{pro}$); TAT = terminal adenylyltransferase (TATase); NiRAN = nidovirus RdRp-associated nucleotidyltransferase; RdRp = RNA-dependent RNA polymerase; HEL = NTPase/helicase; ExoN = exoribonuclease; NMT = guanosine N7-methyltransferase; NendoU = nidovirus uridylate-specific endoribonuclease; MT = ribose 2'-O methyltransferase. **(B)** Polyprotein cleavage sites are depicted with arrowheads matching the color of the proteinase

involved. Proteolytic processing by the viral proteases results in a total of 12 nonstructural proteins (nsp1 to 12), the numbers of which are indicated at the top. Abbreviations: ZF = zinc finger; 1α (PLPα1 or PCPα) and 1β (PLPβ or PCPβ) = papain-like cysteine protease; 2 (PLP2 or CP) = papain-like cysteine protease 2; SP (also called M^{pro}) = chymotrypsin-like serine proteinase; TM = transmembrane domain; NiRAN = nidovirus RdRp-associated nucleotidyltransferase; RdRp = RNA-dependent RNA polymerase; Z = zinc-binding domain; HEL = NTPase/helicase; U = nidovirus uridylate-specific endoribonuclease (NendoU).

1.6.1 Transmembrane proteins

All nidoviruses encode several transmembrane proteins in their ORF1a. Coronaviral nsp3, nsp4 and nsp6 mediate the rearrangement of host cell membranes required for the formation of viral RTCs and are involved in inducing membrane curvature and the formation of subcellular membrane organelles such as convoluted membranes, tubular bodies, vesicles, and zippered ER spherules (Knoops et al., 2008; Maier et al., 2013). Nsp3 is the largest coronavirus-encoded protein multiple functional domains that interact with several proteins involved in the viral replication and transcription processes (Kanjanahaluethai et al., 2007; Lei et al., 2018; von Brunn et al., 2007). Nsp4 contains four transmembrane helices and a carboxy-terminal domain and possesses at least one predicted glycosylation site (Oostra et al., 2008). Crystal structures suggested a novel fold for the C-terminal domain that was proposed to engage in protein-protein interactions (Angelini et al., 2013; Oudshoorn et al., 2017). Nsp6 contains six transmembrane helices in almost all viruses and a highly conserved C-terminus (Baliji et al., 2009).

In arteriviruses, three transmembrane domains have been identified in nsp2, nsp3 and nsp5. Nsp2 contains two hydrophobic regions spanning the lipid bilayer several times (Snijder et al., 2001). However, the exact topology of this protein needs further characterization. The highly hydrophobic nsp3 is presumed to be a tetra-spanning transmembrane protein and thereby playing a key role in the interaction between the arteriviral RTC and the host cell membrane (van Hemert et al., 2008a). Nsp5 is involved

in the membrane association of the EAV replication complex (van der Meer et al., 1998). The presence of the hydrophobic nsp5 at the N-terminus of several polyprotein processing intetrmediates, such as nsp5 to nsp8, nsp5 to 7 and even larger intermediates like nsp5 to nsp12, supports an association of other pp1a- and pp1ab-derived proteins with intracellular membranes (van der Meer et al., 1998).

1.6.2 Papain-like proteases

The N-terminal regions of most nidoviral polyproteins 1a and 1ab are proteolytically processed by papain-like cysteine proteases (PLPs). In all coronaviruses this function is resides in the largest nonstructural protein, nsp3 (Kanjanahaluethai et al., 2007; Lei et al., 2018; von Brunn et al., 2007). However, there are different mechanisms of processing for different coronaviruses (Gadlage and Denison, 2010; Ziebuhr et al., 2007; Ziebuhr et al., 2001). Some viruses possess only one PL^{pro} (PLP1) while others have two active PL^{pro}s with overlapping functions (Baker et al., 1993; Bonilla et al., 1997; Graham and Denison, 2006; Ziebuhr et al., 2001). The active site of PL^{pro} contains a canonical Cys-His-Asp catalytic triad (Ratia et al., 2006).

Arteriviruses possess 2-3 papain-like cysteine proteases (PLP) (EAV contains 2 PLPs and PRRSV contains 3 PLPs) located in nsp1 (PCPα and PCPβ) and nsp2 (PLP2). PCPβ is responsible for cleavage between nsp1↓nsp2 (Snijder et al., 1992) whereas PLP2 is responsible for the cleavage between nsp2↓nsp3 (Snijder et al., 1995). Additionally, PLP2 acts as a cofactor for the processing between nsp4↓5 by the main protease (Snijder et al., 1995).

1.6.3 Main protease

The main protease (M^{pro}) plays a key role in the processing of the replicase polyproteins pp1a and pp1ab (Gorbalenya et al., 1989b; Gorbalenya et al., 1991; Ziebuhr and Siddell, 1999; Ziebuhr et al., 2000). The main difference between the arteriviral and coronaviral

main protease lies in the catalytic center. While the coronavirus nsp5 (also called 3C-like protease) has a chymotrypsin-like fold but (unlike chymotrypsin) employs a Cys-His catalytic dyad (Anand et al., 2002), the arterivirus main protease nsp4 (nsp4, chymotrypsin-like serine protease) combines a chymotrypsin-like fold with the typical Ser-His-Asp catalytic triad of serine proteases (Barrette-Ng et al., 2002).

The following model has been proposed for the maturation or auto-release of $3CL^{pro}$ (Chen et al., 2010; Hsu et al., 2005; Muramatsu et al., 2016). Initially, two immature monomeric polyprotein proecursors form a dimer. Followed by a domain switch where the N-terminus region of the polyprotein immediately upstream of $3CL^{pro}$ interjects itself into the active site of the opposing monomer, an immature dimer is formed. Next, the polyprotein is auto-proteolytically cleaved at its N-terminal cleavage site. The unprocessed C-terminus of the dimer is fixed in the active site of another dimer liberating an active mature dimer in a concentration dependent manner (Grum-Tokars et al., 2008). Finally, the formed, active dimer can process other cleavage sites in the polyprotein containing a conserved glutamine residue at the P1 position (Herold et al., 1993; Ziebuhr et al., 1995).

Two cleavage pathways have been described for the arteriviral main protease, nsp4 (Gorbalenya et al., 1991; Snijder et al., 1996). In the major pathway, nsp2 acts as cofactor for nsp4, triggering nsp4-mediated cleavage of nsp3 to nsp8. Afterwards, nsp5-8 is cleaved at the nsp7↓8 site while the nsp5↓6 and nsp6↓7 sites are inaccessible to the protease. In the minor proteolytic pathway, the nsp4↓5 site remains intact while the nsp5↓6 and nsp6↓7 sites are cleaved (van Aken et al., 2006; Wassenaar et al., 1997). The cleavage sites for nsp4 are Glu↓Gly, Glu↓Ser, Gln↓Ser. These sites are conserved in all arteriviruses but, in some cases, Ala or Lys are found at the P1' position (Wassenaar et al., 1997).

1.6.4 Small nonstructural proteins

Small proteins encoded by ORF1a have been identified as cofactors of other viral nonstructural proteins, which are not identical for corona- and arteriviruses.

1.6.4.1 Coronaviruses

The molecular masses of nsp7 and nsp8 are approximately 9 and 22 kDa, respectively. A crystal structure reported for SARS-CoV nsp7 and nsp8 suggested that the two proteins interact to form large complexes (Zhai et al., 2005). Imbert et al. analyzed a recombinant form of SARS-CoV nsp8 and obtained evidence to suggest that the protein is able to synthesize short oligonucleotides in a Mn^{2+}-dependent manner using specific coronaviral RNA sequences as template (Imbert et al., 2006). These oligonucleotide products were speculated to be subsequently utilized by the RdRp (nsp12) as a primer in primer-dependent initiation of RNA synthesis, implying that nsp8 acts as a virus-encoded primase. In another study, genetic evidence was obtained to suggest that SARS-CoV nsp8 interacts with the 3'-UTR where it may have a role in the initiation of minus-strand RNA synthesis (Züst et al., 2008). Primer-dependent polymerase activity and *de novo* RNA synthesis activity were reported for both nsp8 and nsp(7+8) complex in the presence of Mg^{2+} using primed (partial-duplex) substrate RNAs (te Velthuis et al., 2012). FCoV nsp8 and the uncleaved HCoV-229E nsp7-8-9-10 polyprotein precursor were reported to synthesize short oligonucleotides of up to 6 nucleotides (Xiao et al., 2012). Xiao et al. also suggested that the FCoV nsp8-mediated polymerase activity could be enhanced by the presence of nsp7 in which case a robust RdRp activity resulting in a 67-nucleotide product was observed (Xiao et al., 2012). In another study, Subissi *et al.* suggested that nsp7+8 have a critical role in recruiting nsp12 to RNA, thereby possibly promoting an essential step in nsp12-mediated RNA synthesis (Subissi et al., 2014). Overall, these studies provided a rather inconsistent and partly

contradictory picture on the potential roles of nsp8 in viral RNA synthesis, ranging from (i) primase activity to (ii) a „noncanonical“ *de novo* and primer-dependent RdRp activity and (iii) a role as an RdRp processivity and/or recruitment factor. In contrast to these studies and using additiona controls including RNA substrates with 3’ biotin-blocked ends, Tvarogová *et al*. could show that (at least some of the previously published data) provide evidence for an nsp8-mediated 3’-terminal nucleotidyltransferase rather than RdRp activity. A more detailed characterization of HCoV-229E and SARS-CoV nsp8 revealed that the proteins preferentially transfer AMP to the 3'-end of single- and double stranded RNA substrates (Tvarogova et al., 2019). Most prominent TATase was detected with partially double-stranded RNA containing a 5'-oligo(U) overhang, while little TATase activity was observed if partial-duplex RNAs with 5’ overhangs with heteropolymeric or other homopolymeric sequences were used as substrates. TATase activity was demonstrated for SARS-CoV, FIPV and HCoV-229E, suggesting that nsp8-mediated TATase activity is conserved among coronaviruses.

Different crystal structures have been reported for coronavirus nsp7+8 complex. For the betacoronavirus SARS-CoV nsp7+8, a hollow, cylindrical hexadecameric supercomplex was reported, with eight molecules of nsp7 acting as "mortar" stabilizing a complex of eight molecules of nsp8 (8:8) with two slightly different conformations. The inner cavity of this complex was lined with positively charged amino acid residues and presumably capable of binding RNA, similar to the eukaryotic PCNA (proliferating cell nuclear antigen) sliding clamp and the DNA polymerase III β-subunit of *E. coli* (Zhai et al., 2005) (Figure 5A). Largely based on this structure, the nsp7+8 supercomplex was proposed to be a processivity factor for nsp12 (RdRp). In the crystal structure of this complex, nsp7 was found to adopt a globular structure (Figure 6 D) while nsp8 had a golf club-like structure adopting either a straight (nsp8I) (Figure 6 A)

or bent (nsp8II) conformation (Figure 6 B). In a subsequent study, other trimeric and truncated heterodimeric SARS-CoV nsp7+8complexes were reported, which could potentially be incorporated into the nsp7+8 supercomplex, thereby, most likely, interfering with its activity (Li et al., 2010). In addition to the hexadecameric SARS-CoV nsp7+8 structure, heterotetrametric nsp7+8 (2:2 ratio) complexes were reported for SARS-CoV-2 (Konkolova et al., 2020) (Figure 5 B) and SARS-CoV (this study) (Krichel et al., 2020).

Conversely, the crystal structure analysis of the FCoV nsp7+8 complex revealed a heterotrimeric stoichiometry (Figure 5 C) involving two molecules of nsp7 and one molecule of nsp8 (Xiao et al., 2012). The structure was described to resemble a pistol-like structure (nsp8) with a magazine (2× nsp7), with the α-helices of the two nsp7 molecules adopting different conformations, nsp7I (Figure 6 E) and nsp7II (Figure 6 F).

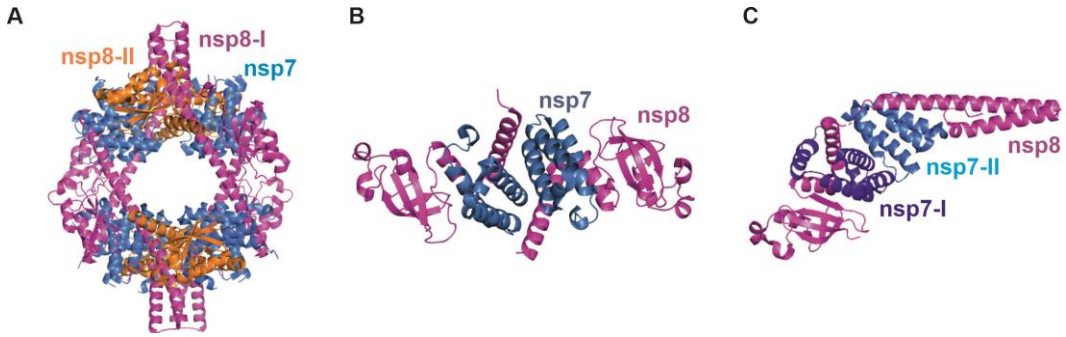


Figure 5: Ribbon representation of the crystal structures reported for the nsp7+8 complex of SARS-CoV (A), SARS-CoV-2 (B) and FCoV (C). The diagram illustrates the hexadecameric SARS-CoV ns7+8 crystal structure (Zhai et al., 2005), the heterotetrameric SARS-CoV-2 nsp7-8 crystal structure (Konkolova et al., 2020) and the heterotrimeric structure reported for nsp7 and nsp8 of FCoV (Xiao et al., 2012). PDB codes: SARS-CoV 2AHM, SARS-CoV-2 6YHU and FCoV 3UBO. The presentation was generated using PyMol.

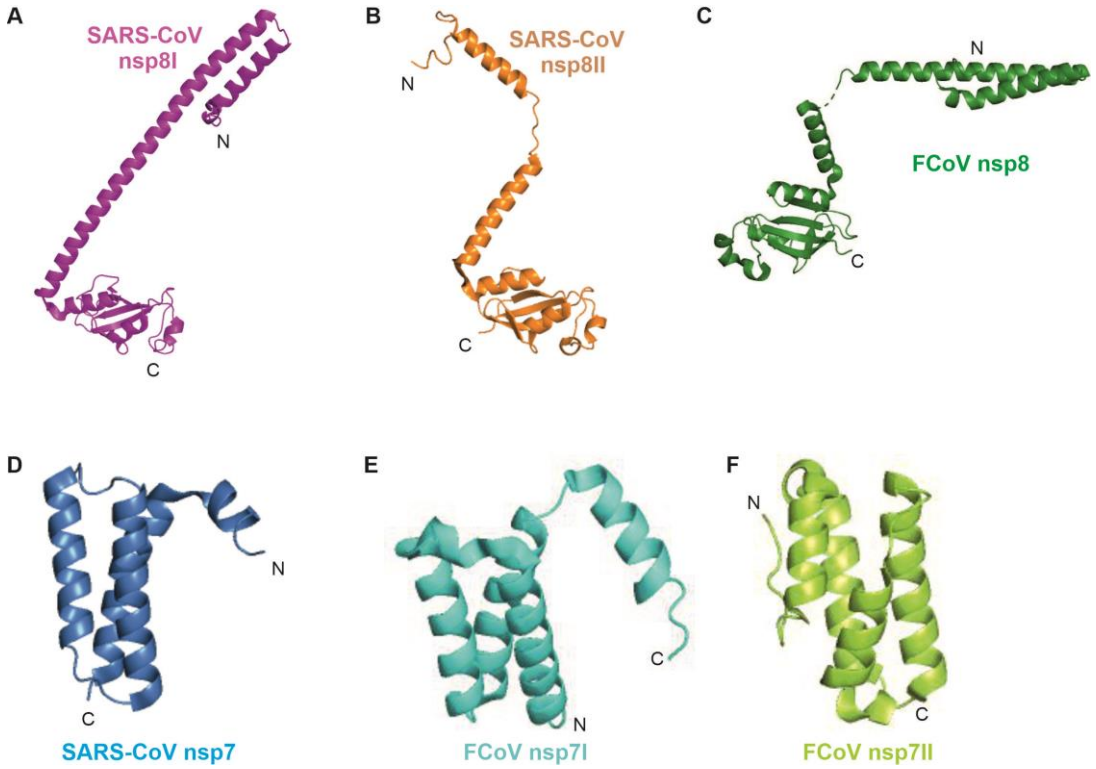


Figure 6: Ribbon representation of the crystal structures reported for nsp7 and nsp8 of SARS-CoV and FCoV. SARS-CoV nsp8I conformation (purple) and nsp8II conformation (orange) **(A and B)** (Zhai et al., 2005); FCoV nsp8 (green) (Xiao et al., 2012) **(C)**. SARS-CoV nsp7 (skyblue) (Zhai et al., 2005); FCoV nsp7I (cyan) **(E)** and nsp7II (chartreuse) **(F)** (Xiao et al., 2012). PDB codes: SARS-CoV 2AHM and FCoV 3UBO.

Coronavirus nsp9 has a mass of approximately 12 kDa. Structural and RNA binding experiments concluded that it is possibly involved in single-stranded RNA binding, most likely by associating with other proteins (Egloff et al., 2004; Sutton et al., 2004; Zeng et al., 2018). RNA-protein shift assays and biophysical analysis suggested that monomers bind to ssRNA and ssDNA less efficiently than dimers. Analytical centrifugation experiments provided evidence for an interaction between nsp9 and nsp8, and suggested that nsp9 may stabilize nsp8 (Sutton et al., 2004). Multiple groups have solved crystal structures for HCoV-229E, PEDV, PDCoV, IBV, SARS-CoV and SARS-CoV-2 nsp9 and an additional disulfide bond stabilizing the interface of the HCoV-229E nsp9 dimer was the only major difference between the structures determined for the

SARS-CoV and HCoV-229E proteins (Egloff et al., 2004; Littler et al., 2020; Ponnusamy et al., 2008; Sutton et al., 2004; Zeng et al., 2018).

Nsp10 of coronaviruses plays an important role in the synthesis of viral RNA (Donaldson et al., 2007b) and polyprotein processing (Donaldson et al., 2007a), most importantly by acting as a cofactor for the methyltransferase activity of nsp16 and the exonuclease avtivity of nsp14. Sequence alignments and biochemical studies suggest that nsp10 is a highly conserved component of the coronavirus RTC and has been shown to bind RNA and DNA (Joseph et al., 2006). Nsp10 was found to form trimers which are assembled into a higher-order dodecameric structure (Su et al., 2006). Co-crystal structural analysis showed for nsp14 and nsp16 that they interact with (nearly) the same surface of nsp10 (Chen et al., 2013; Decroly et al., 2011a).

Nsp11 is a short peptide of highly variable sequence which is produced if no frameshift event takes place in the ORF1a/b overlap region. So far, no function has been identified for nsp11.

1.6.4.2 Arteriviruses

The relatively small nsp6 and nsp8 proteins encoded by arteriviruses consist of only 20 and 50 amino acid residues, respectively, and their functions in viral replication remain yet to be characterized (Li et al., 2012; Li et al., 2015; Snijder et al., 1994).

An additional internal cleavage site was identified in nsp7 and shown to be cleaved by the nsp4 protease resulting in the two products, nsp7α (~13 kDa) and nsp7β (~13 kDa) (van Aken et al., 2006). Nsp7α contains four hydrophobic regions which are required for the correct folding of the protein; it has a unique conformation consisting of three α-helices packed against a mixed β-sheet. There is no clear evidence for a specific function of the uncleaved form of nsp7, while there are some reports suggest that nsp7α is critical for RNA synthesis (Chen et al., 2017; Manolaridis et al., 2011).

1.6.5 Polymerase

Nidoviral RNA-dependent RNA polymerases (RdRp) are the central enzymes catalyzing viral genome replication. Nidoviral polymerase proteins have two enzymatic activities that are mediated by two independent domains: the NiRAN domain located at the N-terminus and the RdRp domain occupying the C-terminal two-thirds of the protein.

The Nidovirus RdRp-associated nucleotidyl transferase (NiRAN) domain is conserved across all currently known members of the order *Nidovirales*. The domain was named after its *in vitro* enzymatic activity confirmed for the equine arteritis virus (EAV) nsp9 (which is a homologue of the coronavirus nsp12). Based on sequence alignments, it was suggested that an uncharacterized third domain (50 to 170 amino acid) connects the NiRAN and RdRp domains (Lehmann et al., 2015) (Figure 7A).

The nidoviral RdRp domain, first identified by comparative amino acid sequence analysis, revealed phylogenetic relationships to the RdRps of other plus-stranded RNA viruses (Boursnell et al., 1987; Gorbalenya et al., 1989b; Gorbalenya et al., 2002). Homology modelling of nsp12 with other known positive-stranded viral RdRps identified seven conserved motifs (A to G) resembling a cupped right hand, consisting of fingers, palm, and thumb subdomains (Figure 7A), which was later confirmed by crystal structure analyses (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019). In contrast to the usual GDD signature sequence of most positive-stranded viral RdRps, the palm domain contains the signature sequence SDD, which is conserved across the *Nidovirales* order and forms the catalytic center for RNA polymerase activity (Xu et al., 2003). In addition, a G motif (SXGXP followed by a basic residue) was identified in the nidoviral RdRp domain and suggested to be involved in primer-dependent RNA synthesis of this group of RdRps. Using cryo-electron microscopy, the

structures were solved for SARS-CoV and SARS-CoV-2 nsp12 in combination with nsp7+8 (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019). Kirchdoerfer *et al*. determined for SARS-CoV that one copy of a nsp7+8 heterodimeric complex is bound to the thumb domain of one molecule of nsp12 (Figure 7 B), thereby facing the NTP entry channel. The authors hypothesized that, upon binding of the nsp7+8 heterodimer, the RdRp index finger loop is stabilized to allow binding of the template. The second subunit of nsp8 is bound to the interface domain without nsp7 close to the finger domain and RNA-binding channel of the polymerase. A similar nsp7-nsp8-nsp12 architecture was also determined for SARS-CoV-2 (Figure 7C) (Gao et al., 2020). At the same time, Hillen et al. determined a pre-initiation form of the replicase complex that consisted of a nsp12+nsp8-nsp7+RNA duplex (Figure 7 D). The conserved residues of the nsp12 active site cleft were bound to the first turn of RNA while two copies of nsp8 were shown to bind to the opposite side of the nsp12 active-site cleft and second turn of RNA. Therefore, it was suggested that the N-terminal long helical extensions of nsp8 slides over the RNA like a clamp preventing the premature termination of RNA synthesis (Hillen et al., 2020). In earlier studies, a model of an active replicase complex was proposed which contained a hexadecameric nsp7+8 complex associated with RNA and synthesizing primers by *de novo* mechanism. Also, Cheng et al. reported a primer-dependent polymerase activity for a N-terminally GST-fused unstable nsp12 protein using a 12-18 nt double-stranded homopolymeric poly(A)/oligo(dT) RNA substrate and Mg^{2+} (Cheng et al., 2005). Non-processive primer-dependent polymerase activity was found in the presence of Mg^{2+} and nonspecific nucleotide addition by SARS-CoV nsp12 by using the SARS 3'-UTR as a substrate (Ahn et al., 2012; te Velthuis et al., 2010a). The report described that the presence of an N-terminal tag on nsp12 and high concentration of Mn^{2+} stimulate *de*

novo initiation of RNA synthesis (Ahn et al., 2012). Subissi et. al found *de novo* RNA synthesis using SARS-CoV UTRs and primer-dependent RNA polymerase activity only with a complex of nsp7L8 (hexahistidine physical linker connecting nsp7 and nsp8) and nsp12. The authors suggested that the nsp7L8 complex confers the RNA–binding ability of nsp12 with support of extensive replacements of conserved amino acid residues of nsp7 and nsp8 (Subissi et al., 2014).

The 78 kDa EAV nsp9 contains the RdRp domain at the C-terminus and directs viral RNA synthesis in conjunction with other viral and host cellular proteins.

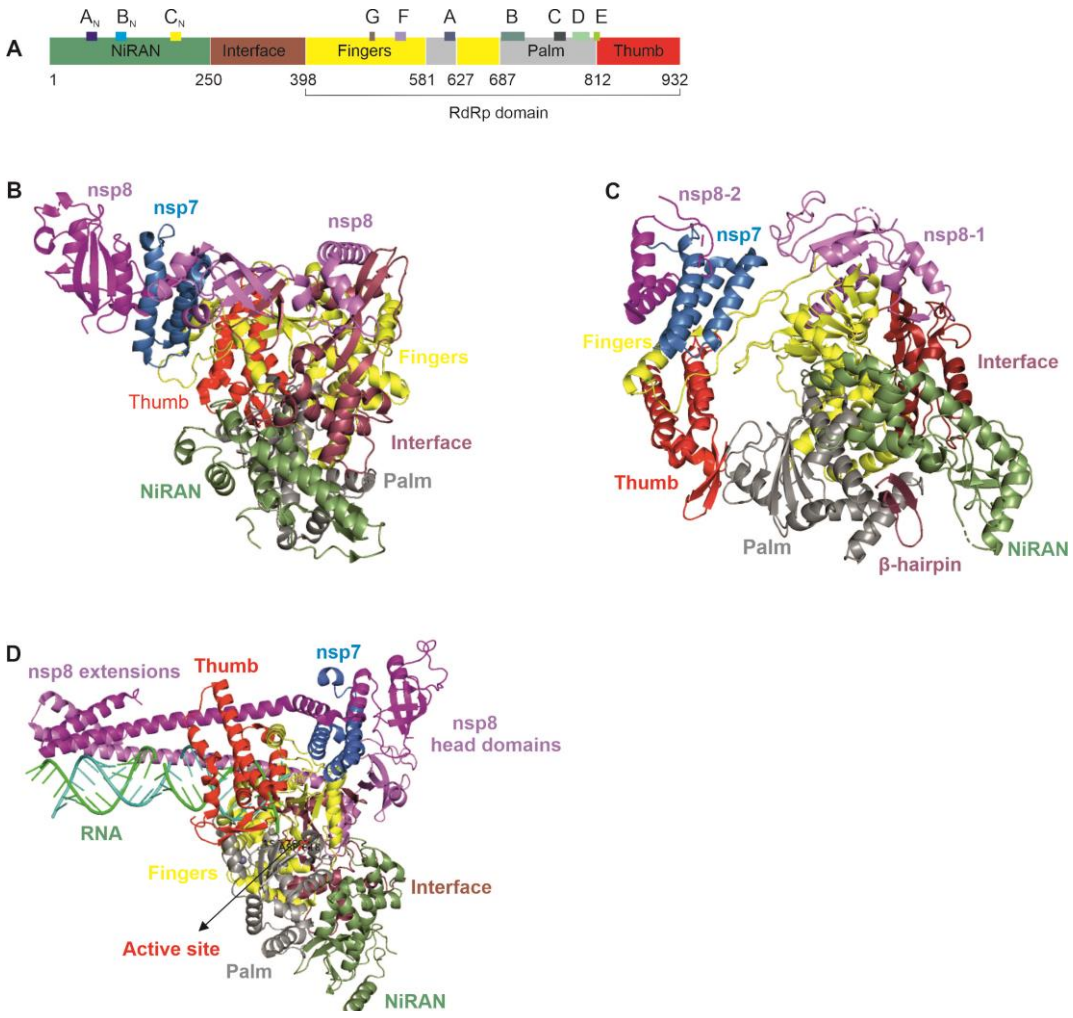


Figure 7: Cryo-EM tructures of SARS-CoV and SARS-CoV-2 nsp7-nsp8-nsp12 complexes. (A) Schematic representation of domains, conserved motifs A-G and regions observed in nsp12. **(B, C and D)** Ribbon representation of the electron microscopic structure reported for SARS-CoV and SARS-CoV-2 nsp7-nsp8-nsp12 complexes. Nsp12 contains the NiRAN domain (green), interface (brick) and RdRp domain (yellow, gray, and red). **(B)** A nsp7+8 heterodimer in complex with nsp12 and an additional nsp8 reported to interact with the N-

terminal extension of nsp12 (Kirchdoerfer and Ward, 2019). **(C)** Ribbon representation of a SARS-CoV-2 nsp7-nsp8-nsp12 complex with a nsp7-nsp8 heterodimer and an additional monomeric nsp8 (Gao et al., 2020). **(D)** Ribbon representation of the SARS-CoV-2 nsp7-nsp8-nsp12+RNA complex. The active site cleft of nsp12 is bound to the first turn of RNA. Two copies of nsp8 with conformation a and b, respectively, are bound to opposite sides of the active-site cleft and help to position the second turn of the RNA. Nsp7 interacts with conformation b of nsp8. The nsp8 long helical extensions were found to protrude along the exiting RNA (Hillen et al., 2020)

1.6.6 Helicase

The helicase is one of the most conserved replicase subunits of the nidoviral RNA synthesis machinery (Gorbalenya et al., 1989a; Seybert et al., 2000). Coronaviral nsp13/arteriviral nsp10 are multi-functional proteins containing a zinc-binding domain (ZBD) at the N-terminus and a helicase domain at the C-terminus (Gorbalenya and Koonin, 1989; Gorbalenya et al., 1988). Biochemical studies revealed similar enzymatic properties for corona- and arterivirus helicases, including nucleic acid-stimulated ATPase and 5'-3' duplex unwinding activities on both RNA and DNA substrates containing 5'-single-stranded regions (Ivanov et al., 2004b; Ivanov and Ziebuhr, 2004; Seybert et al., 2000). It has been also shown that coronaviral nsp13 exhibit a 5'-RNA triphosphatase (5'-RTPase) activity *in vitro* and concluded that this activity may be involved in RNA capping by generating 5'-ppN-RNA. The conserved residues (13 conserved Cys and His) in the ZBD region are involved in coordinating zinc ions and are critical for the ATPase and duplex-unwinding activities (Gorbalenya et al., 1989b; van Dinten et al., 2000). Mutations of conserved Cys/His residues impaired interactions between the ZBD and helicase, resulting in the blocking of viral RNA synthesis and virion biogenesis (Seybert et al., 2005; van Dinten et al., 2000). Adedeji *et al*. has shown a 2-fold stimulation of helicase activity in the presence of SARS-CoV nsp12 (Adedeji et al., 2012). Nidoviral helicases belong to the SF1 helicase superfamily and share structural features with the eukaryotic Upf1 helicase, a key factor involved in nonsense-mediated mRNA decay in cells (Deng et al., 2014).

1.6.7 Exoribonuclease

Nsp14 contains an N-terminal 3'-5' exonuclease domain (ExoN). Due to three conserved motifs (motif I (DE), II (E), and III (D)), nsp14-ExoN belongs to the DE(D/E)D superfamily of exonucleases (Snijder et al., 2003). Coronavirus replication fidelity is thought to involve the 3'-5' exoribonuclease which was suggested to excise mismatched nucleotides incorporated into the product strand (Ferron et al., 2018; Ma et al., 2015; Snijder et al., 2003). Inactiviation of the ExoN catalytic center in genetically engineered alpha- and betacoronaviruses resulted in reduced abundance of MHV sgmRNA2, reduced replication and a mutator phenotype (Eckerle et al., 2007) and severe replication and transcription defects in the case of HCoV-229E (Minskaia et al., 2006). Subissi *et al*. proposed that the nsp10/14 complex can correct mismatched nucleotides generated during RNA synthesis (Subissi et al., 2014). Studies proposed that nsp10 stimulates the ExoN activity of nsp14 (Bouvet et al., 2012; Decroly et al., 2011b; Subissi et al., 2014). The exoribonuclease is capable of hydrolyzing ssRNAs, dsRNA and nonmethylated RNA substrates and acts as a proofreading system that is generally accepted to be involved in improving the replication fidelity of the large coronavirus genome (Lauber et al., 2013). The smaller arteriviruses lack a homologue to the coronaviral nsp14.

1.6.8 Endoribonuclease

The highly conserved uridine-specific endoribonuclease is a multifunctional protein with a crucial role in viral RNA synthesis and is considered a genetic marker of the order *Nidovirales* (Guarino et al., 2005; Ivanov et al., 2004a; Nedialkova et al., 2009; Posthuma et al., 2006). The exact function of this protein in the context of viral replication and the observed stimulatory effect of Mn^{2+} ions on the endonucleolytic activity of nsp15 remain to be characterized. Nsp15 forms a hexamer with the active site at the C-terminus (Guarino et al., 2005; Ricagno et al., 2006; Snijder et al., 2003;

Zhang et al., 2018). EAV nsp11 is the homolog to the coronaviral nsp15. In contrast to the coronaviral nsp15, metal ion-dependent uridylate-specific endoribonuclease activity, EAV nsp11 was found to possess metal ion-independent activity with a less constrained specificity for pyrimidines (Nedialkova et al., 2009).

1.6.9 Methyltransferase

Nsp10 acts as a cofactor for the 2'-O methyltransferase activity of nsp16 for the maturation of viral RNA 5' cap structures (Bouvet et al., 2010). So far, two methyltransferases have been identified in coronaviruses. The nsp14 C-terminal domain harbors the guanosine N7-methyltransferase activity and nsp16 has ribose 2'-O methyltransferase activity. The N7-methyltransferase domain of nsp 14 is separated form the ExoN domain by a flexible hinge region consisting of a loop and three strands. This region allows the rotational movements of two domains to coordinate the two different enzymatic activities (Ogando et al., 2019). Two clusters of residues have been shown to be key for the N7-MTase activity of SARS-CoV nsp14/10 (Chen et al., 2009; Ma et al., 2015). Cluster one is a canonical SAM-binding motif I (DxGxPxG/A), where SAM is the methyl donor for the N7-methyltransferase reaction catalyzed by nsp14. The second cluster forms a pocket that holds the GTP of the mRNA cap structure in proximity of the methyl donor SAM (Chen et al., 2009; Chen et al., 2013).

The 2'-O-methyltransferase plays a key role in the coronavirus life cycle by preventing virus detection by the cell's innate immune response (Chen et al., 2011; Decroly et al., 2008). The catalytic site of nsp16 contains the conserved K-D-K-E catalytic tetrad typical for SAM-dependent 2'-O-methyltransferases. Further crystallography studies determined a heterodimeric complex for nsp10/nsp16 and suggested that, by interacting with nsp16, nsp10 helps preserve the SAM-binding pocket of nsp16 in the active conformation (Decroly et al., 2011a). Nsp16 has been proposed to be involved in cap

synthesis by adding the final methyl group to complete the cap structure formation (Bouvet et al., 2010; Chen et al., 2011).

Capping involves three sequential enzymatic activities: First, an RNA triphosphatase (RTPase) removes the γ-phosphate group from the 5'-triphosphate end (pppN) of the nascent mRNA chain to generate a diphosphate 5'-ppN (possibly coronaviral nsp13/arteriviral nsp10) (Ivanov et al., 2004b; Ivanov and Ziebuhr, 2004). Second, a guanylyltransferase (GTase) transfers a GMP to the 5'-diphosphate ends to yield the cap core structure (GpppN) (this activity has not yet been identified). Last, the cap structure is methylated at the N7 and 2'-O position of guanosine and the ribose of the first nucleotide by a N7- and 2'-O-methyltransferase (MTase) to produce a cap-0 and cap-1 structure ($m7GpppN_{2'\text{-}Om}$).

The following tables summarizes the approximate protein sizes and previously established enzymatic activities associated with corona- and arterivirus nonstructural proteins.

Table 2: Characteristics of coronavirus replicase gene-encoded nsps.

Protein	MW (kDa)	Activities[1]
nsp1	12-27	Host mRNA degradation, translational inhibition, cell cycle arrest, inhibition of IFN signaling
nsp2	65-87	Unknown
nsp3	177-222	Papain-like proteases ($PL1^{pro}$, $PL2^{pro}$), poly (ADP-ribose) binding, de-ADP-ribosylation, DMV formation, IFN antagonist, RNA binding, deubiquitinating activity
nsp4	54-56	DMV formation (TM2)
nsp5	33-34	Main protease (M^{pro}, $3CL^{pro}$)
nsp6	32-34	DMV formation (TM3)

nsp7	9-10	Forms complex with nsp8
nsp8	22-23	TATase activity, forms multimeric complexes in with nsp7
nsp9	12	Single-stranded RNA binding
nsp10	14-15	Cofactor for nsp14 and nsp16
nsp11	1-2	Unknown
nsp12	105-106	NiRAN and RdRp activities
nsp13	66-67	5'-3' helicase, NTPase and RNA 5'-triphosphatase
nsp14	58-60	3'-5'-exoribonuclease and N7-methyltransferase
nsp15	38-42	Hexameric uridylate-specific endoribonuclease
nsp16	33-34	RNA cap1 formation (2′-O-methyltransferase)

Table 3: Characteristics of arteriviral replicase gene-encoded nsps

Protein	**MW (kDa)**	**Activities**[1]
nsp1	28.5	Papain-like cysteine protease(s), RNA and protein-protein interactions, activation of sg RNA synthesis
nsp2	61.5	Transmembrane domain (TM1), cysteine protease, co-factor for nsp4 to process nsp4↓5 site
nsp3	25	DMV formation (TM2)
nsp4	22	Main protease (M^{pro}, chymotrypsin-like serine protease)
nsp5	18.2	DMV formation (TM3)
nsp6	2.4	Unknown
nsp7	26	Unknown
nsp8	5.5	Unknown
nsp9	78	NiRAN and RdRp activities
nsp10	50	Helicase (5'-3') and NTPase
nsp11	24	Endoribonuclease (NendoU) (Cleaves after uridylte or cytidylate nucleotides)
nsp12	12.5	Unknown

2 Aim of this study

2.1 Goal I

Although different laboratories have tried to determine the functional roles of nsp7, nsp8 and nsp12 in coronaviral viral RNA synthesis, our understanding of the functions of these protein remains limited and partially contradictory results have been published for recombinant protein constructs. Therefore, one goal of this study was to generate tag-free nsp7 and nsp8 proteins, pre-assembled nsp(7+8) complexes and nsp12 to investigate their biochemical properties in, for example, primed RNA synthesis (elongation), *de novo* initiation of RNA synthesis, nucleotide selection and fidelity, and the role of metal ions in the polymerase activities.

So far, different conformations and structures of nsp7+8 complexes have been reported and it remained unclear whether a common subunit composition and stoichiometry exists for nsp7+8 complexes of all coronaviruses. To determine coronavirus nsp(7+8) stoichiometries more systematically, a set of alpha- and betacoronavirus nsp(7+8) complexes produced from nsp7-8 and nsp7-8-9 polyprotein precursors by M^{pro}-mediated proteolytic cleavage were planned to be analyzed using gel filtration chromatography and native mass spectrometry to identify common and distinct characteristics of these complexes.

2.2 Goal II

A second major goal of this study was to characterize the role of arteriviral replicase proteins in RNA synthesis. Previously published data for the EAV polymerase (nsp9) proposed *de novo* RNA synthesis using polyU and polyC templates (Beerens et al., 2007), another study only detected primer extension activity on homopolymeric duplex RNA substrates (Lehmann et al., 2016), whereas more recent studies identified self-

nucleotidylation (NiRAN) properties for the N-terminal domain of EAV nsp9 but failed to confirm the previously reported polymerase activity (Lehmann et al., 2016; Lehmann et al., 2015). Also, a possible acceptor protein (target) of the NiRAN-mediated nucleotidylation activity had not been identified. While previous studies reported that the coronavirus nsp(7+8) complex acts as processivity factor for nsp12, no such cofactor has been identified for the homologous arterivirus polymerase protein (nsp9). Therefore, the work focused on EAV nsp9 and the identification of EAV-encoded proteins potentially involved in RdRp and/or NMPylation activities of nsp9. To this end, EAV nsp6, nsp6-7-8, nsp7, nsp9, nsp10 and nsp12 were produced and used in polymerase and protein nucleotidylation assays.

3 Materials and Methods

3.1 General chemicals and reagents

All chemicals were purchased either from Roth or Sigma at the highest available purity unless stated otherwise.

3.2 Radioactive nucleotides

All [^{32}P]-labeled nucleotides were purchased from Hartmann Analytic at a specific activiyt of 3000 Ci/mmol, 10 mCi/ml.

3.3 Enzymes

Name	**Supplier**
T4 polynucleotide kinase	Thermo Scientific
Proteinase K	Invitrogen
RNAseOUT	Invitrogen
Reverse transcriptase (SuperScript III)	Invitrogen
Restriction endonucleases	Thermo Scientific/NEB
Phusion High-Fidelity DNA polymerase	Thermo Scientific

3.4 Antibiotics

Name	**Stock solution**	**Final concentration**
Ampicillin	100 mg/ml	100 µg/ml
Anhydrotetracycline	2 mg/ml	200 ng/ml
Chloramphenicol	34 mg/ml	34 µg/ml

3.5 Plasmids

Name	**Antibiotic resistance**	**Promoter**
pASK-IBA3plus	Ampicillin	tet
pMAL-c2X	Ampicillin	tac

pCGI	Chloramphenicol	GAL10

3.6 Bacterial stains

Name	Genotype
BL21 (DE3)	F– *ompT gal dcm lon hsdS$_B$* (r_B^- m_B^-) λ (DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])
TOP10F'	F'[*lacI^q* Tn10(*tet^R*)] *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 deoR nupG recA1 araD139* Δ(*ara-leu)7697 galU galK rpsL* (*Str^R*) *endA1* λ$^-$
TB1	F$^-$ *ara* Δ*(lac-proAB) [Φ80dlac* Δ*(lacZ)M15] rpsL* (*Str^R*) *thi hsdR*
Rosetta (DE3)	*F- ompT hsdS$_B$ (r_B- m_B-) gal dcm (DE3) pRARE (CamR)*

3.7 Oligonucleotides

All oligonucleotides were purchased from IDT.

Single-stranded RNA substrates	
KR01	5'-GCUAUGUGAGAUUAAGUUAU-3'
KR01 **-b**	5'-GCUAUGUGAGAUUAAGUUAU-3'-**Bio**
KR07	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3'
KR07 **-b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3'-**Bio**
KR07 **A$_4$**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC**AAAA**-3'
KR07 **A$_4$ -b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC**AAAA**-3'-**Bio**
KR07 **A$_{10}$**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC**AAAAAAAAAA**-3'
KR07 **A$_{10}$ -b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC**AAAAAAAAAA**-3'-**Bio**

Double-stranded RNA substrates	
KR01 / U_{20} JTR1 -comp **-b**	5'-GCUAUGUGAGAUUAAGUUAU-3' ||||||||||||||||||| **Bio-**3'-CGAUACACUCUAAUUCAAUA**UUUUUUUUUUUUUUUUUUUU**-5'
KR01 / JTR1 **-b**	5'-GCUAUGUGAGAUUAAGUUAU-3' ||||||||||||||||||| **Bio-**3'-CGAUACACUCUAAUUCAAUA**CUCUCUCUCUCUCUCUCUCU**-5'
KR07 / **U$_{10}$** KR07 -comp **-b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3' |||||||||||||||||||||||||| **Bio-**3'-AUUACCUUGCCAAAGCUAUACCUAUGUG**UUUUUUUUUU**-5'

KR07 / **A_{10}** KR07 -comp -**b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3' |||||||||||||||||||||||||| **Bio-**3'-AUUACCUUGCCAAAGCUAUACCUAUGUG**AAAAAAAAAA**-5'
KR07 / **C_{10}** KR07 -comp -**b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3' |||||||||||||||||||||||||| **Bio-**3'-AUUACCUUGCCAAAGCUAUACCUAUGUG**CCCCCCCCCC**-5'
KR07 / **$(CU)_5$** KR07 -comp -**b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3' |||||||||||||||||||||||||| **Bio-**3'-AUUACCUUGCCAAAGCUAUACCUAUGUG**CUCUCUCUCU**-5'
KR07 / **$(UC)_5$** KR07 -comp -**b**	5'-UAAUGGAACGGUUUCGAUAUGGAUACAC-3' |||||||||||||||||||||||||| **Bio-**3'-AUUACCUUGCCAAAGCUAUACCUAUGUG**UCUCUCUCUC**-5'
KR07 40 -b / KR07 10-comp	5'-**GUUAGUAACUGCUAAUGGAACGGUUUCGAU**AUGGAUACAC-3'**-Bio** |||||||||| 3'-UACCUAUGUG-5'
KR07 40 -b / KR07 10-comp **-b**	5'-**GUUAGUAACUGCUAAUGGAACGGUUUCGAU**AUGGAUACAC-3'**-Bio** |||||||||| **Bio-**3'-UACCUAUGUG-5'

Primers used for cloning	
HCoV-229E	
Fw_nsp7_1	FW: 5'-CTCCGCGGTGGATCTAAATTGACTGATCTTAAG-3'
Rev_nsp7_1	RV: 5'-GAGTGCGGCCTCAGTGGTGGTGATGATGGTGATG-3'
Fw_nsp7_2	FW: 5'-CAGTCAATTTAGATCCACCGCGGAGACGCAGGAC-3'
Rev_nsp7_2	RV: 5'-TCCATTTTGCAACATCACCATCATCACCACCACTAAT-3'
Fw_nsp8_1	Fw: 5'-CTCCGCGGTGGAAGTGTTGCATCTTCTTTTGTTG-3'
Rev_nsp8_1	RV: 5'-GAGTGCGGCCTCATTAGTGGTGGTGATGATGGTGATGCTGCAATTTAACGACACGTTC-3'
Fw_nsp8_2	FW: 5'-GAAGATGCAACACTTCCACCGCGGAGACGCAGGAC-3'
Rev_nsp8_2	RV: 5'-GTTAAATTGCAGCATCACCATCATCACCACCACTAATGAGGCCGCACTCGAGGTCGAC-3'
Fw_nsp7-L_1	FW: 5'-CTCCGCGGTGGATCTAAATTGACTGATCTTAAG-3'
Rev_nsp7-L_1	RV: 5'-GATGATGGTGATAGATGCAACACTTTGCAAAATGGAGTCGTTCTCAAAATAAGAATCG-3'
Fw_nsp7-L_2	FW: 5'-GAGAACGACTCCATTTTGCAAAGTGTTGCATCTATCACCATCATCACCACCACTGAGGCCGCAC-3'
Rev_nsp7-L_2	Rv: 5'-CAGTCAATTTAGATCCACCGCGGAGACGCAGGAC-3'
Fw_nsp8-L_1	FW: 5'-CTCCGCGGTGGAAGTGTTGCATCTTCTTTTGTTG-3'
Rev_nsp8-L_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGTATTTCATTGTTCTGCAATTTAACGACACG-3'
Fw_nsp8-L_2	FW: 5'-GTTAAATTGCAGAACAATGAAATACATCACCATCATCACCACTGAGGCCGCACTCGAG-3'
Rev_nsp8-L_2	RV: 5'-GAAGATGCAACACTTCCACCGCGGAGACGCAGGAC-3'
Fw_nsp7-8-L_1	FW: 5'-CTCCGCGGTGGATCTAAATTGACTGATCTTAAG-3'
Rev_nsp7-8-L_1	RV: 5'-CCTCAGTGGTGATGATGGTGATGACCAGATCCAGACTGCAATTTAACGACACGTTCAC-3'
Fw_nsp7-8-L_2	FW: 5'-GTTAAATTGCAGTCTGGATCTGGTCATCACCATCATCACCACTGAGGCCGCACTCGAG-3'
Rev_nsp7-8-L_2	RV: 5'-TCCATTTTGCAACATCACCATCATCACCACCACTAAT-3'
Fw_nsp7-8-9_1	FW: 5'-CTCCGCGGTGGATCTAAATTGACTGATCTTAAG-3'
Rev_nsp7-8-9_1	RV: 5'-GAGTGCGGCCTCAGTGGTGGTGATGATGGTGATG-3'
Fw_nsp7-8-9_2	FW: 5'-CAGTCAATTTAGATCCACCGCGGAGACGCAGGAC-3'
Rv_nsp7-8-9_2	RV: 5'-GTGAGATTGCAACATCACCATCATCACCACCACTAAT-3'
Rv_nsp5His_6_1	RV: 5'-GACGGCCAGTGCTTAATGGTGATGATGGTGATGTTGCAGGTTAACACCAAACATTTG-3'

Fw_nsp5His$_6$_2	FW: 5'- GGTGTTAACCTGCAACATCACCATCATCACCATTAAGCACTGGCCGTCGTTTTAC-3'
Fw_nsp12_SAA	FW: 5'-GATGATTCTTTCTGCTGCTAGTGTTGTGTGC-3'
Rv_nsp12_SAA	RV: 5'-CAACACTAGCAGCAGAAAGAATCATCATAG-3'

PEDV	
Fw_nsp7-8-L_1	FW: 5'-CTCCGCGGTGGATCTAAACTGACTGATATTAAGTG-3'
Rv_nsp7-8-L_1	RV: 5'-CCTCAGTGGTGATGATGGTGATGACCAGATCCAGACTGGAGCTTAACAATACGCTCAC-3'
Fw_nsp7-8-L_2	FW: 5'-GTTAAGCTCCAGTCTGGATCTGGTCATCACCATCATCACCACTGAGGCCGCACTCGAGG-3'
Rv_nsp7-8-L_2	RV: 5'-TCAGTTTAGATCCACCGCGGAGACGCAGGACCAGA-3'
Fw_nsp12_1	Fw: 5'-CTCCGCGGTGGAAGCACTGATATGGCTTATTTAAACGAG-3'
Rv_nsp12_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGTTGCAAAACTGCAGATTTCTCATACATG-3'
Fw_nsp12_2	FW: 5'-GAAATCTGCAGTTTTGCAACATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGAC -3'
Rv_nsp12_2	RV: 5'-CCATATCAGTGCTTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp12-FS-del	FW: 5'-CACTGATTATGGCTTATTTAAACGAGTACGGGGCTC-3'
Rv_nsp12-FS-del	RV: 5'-GTTTAAATAAGCCATAATCAGTGCTTCCACCGCGGAG-3'
Fw_nsp12_SAA	FW: 5'-GATGATTCTTTCTGCTGCTGGCGTTGTTTGTTACAACAATG-3'
Rv_nsp12_SAA	Rv: 5'-CAAACAACGCCAGCAGCAGAAACGATCATCATTGAAAAATG-3'

TGEV	
Fw_nsp7-8-L_1	FW: 5'-TCTCCGCGGTGGATCAAAACTTACAGAGATGAAATGTA-3'
Rv_nsp7-8-L_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGACCAGATCCAGACTGAAGCTTTGTGGTTC-3'
Fw_nsp7-8-L_2	FW: 5'-CAAAGCTTCAGTCTGGATCTGGTCATCACCATCATCACCACTGAGGCCGCACTCGAG-3'
Rv_nsp7-8-L_2	RV: 5'-TAAGTTATCCACCGAGGAGACGCAGGACCAGA-3'
Fw_nsp12_1	Fw: 5'-CTCCGCGGTGGAAGTTTTACTGTTGATCAAAGTTATTTAAAC-3'
Rv_nsp12_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGTTGCAAGACAGTGGACTTTTCATAGAG-3'
Fw_nsp12_2	FW: 5'-GTCCACTGTCTTGCAACATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGACCTG-3'
Rv_nsp12_2	RV: 5'-CAACAGTAAAACTTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp12-FS-del	FW: 5'-GTTGATCAAAAGTTATTTAAACGAGTGCGGGGTTC-3'
Rv_nsp12-FS-del	RV: 5'-GTTTAAATAACTTTTGATCAACAGTAAAAC-3'
Fw_nsp12_SAA	FW: 5'-GATGATTTTATCTGCTGCTGGAGTTGTGTGCTACAAC-3'
Rv_nsp12_SAA	Rv: 5'-CACAACTCCAGCAGCAGATAAAATCATCATAGAAAAG-3'

FIPV	
Fw_nsp7-L_1	FW: 5'-TCTCCGCGGTGGATCTAAACTTACAGAGATGAAGTGT-3'
Rv_nsp7-L_1	RV: 5'-CAGTGGTGATGATGGTGATGAGATGCTACACTCTGTAAAATAGTGGTGTTCTCAA-3'
Fw_nsp7-L_2	FW: 5'-ACTATTTTACAGAGTGTAGCATCTCATCACCATCATCACCACTGAGGCCGCACTCGAGG-3'
Rv_nsp7-L_2	RV: 5'-TGTAAGTTTAGATCCACCGCGGAGACGCAGGACC-3'

Fw_nsp8-L_1	FW: 5'-CGTCTCCGCGGTGGAAGTGTAGCATCTGCTTATGCTGCTTTG-3'
Rv_nsp8-L_1	RV: 5'-CTCAGTGGTGATGATGGTGATGGATTTCATTATTCTGCAACTTTGTGGTTCTCTC-3'
Fw_nsp8-L_2	FW: 5'-CCACAAAGTTGCAGAATAATGAAATCCATCACCATCATCACCACTGAGGCCGCACTCG-3'
_nsp8-L_2	RV: 5'-GCAGATGCTACACTTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp7-8-L_1	FW: 5'-TCTCCGCGGTGGATCTAAACTTACAGAGATGAAGTGT-3'
Rv_nsp7-8-L_1	RV: 5'-CCTCAGTGGTGATGATGGTGATGACCAGATCCAGACTGCAACTTTGTGGTTCTCTC-3'
Fw_nsp7-8-L_2	FW: 5'-CACAAAGTTGCAGTCTGGATCTGGTCATCACCATCATCACCACTGAGGCCGCACTC-3'
Rv_nsp7-8-L_2	RV: 5'- TGTAAGTTTAGATCCACCGCGGAGACGCAGGACC -3'
Fw_nsp7-8-9_1	FW: 5'-TCTCCGCGGTGGATCTAAACTTACAGAGATGAAGTGT-3'
Rv_nsp7-8-9_1	RV: 5'- GTGCGGCCTCAGTGGTGATGATGGTGATGTTGCAGACGAACCGTTGCACCGATATAAC-3'
Fw_nsp7-8-9_2	FW: 5'-TCGTCTGCAACATCACCATCATCACCACTGAGGCCGCACTCGAGG-3'
Rv_nsp7-8-9_2	RV: 5'-TGTAAGTTTAGATCCACCGCGGAGACGCAGGACC-3'
Rv_nsp5His$_6$_1	RV: 5'-GACGGCCAGTGCTTAATGGTGATGATGGTGATGACCACTCTGAAGATTAACACCATAC-3'
Fw_nsp5His$_6$_2	FW: 5'-GTTAATCTTCAGAGTGGTCATCACCATCATCACCATTAAGCACTGGCCGTCGTTTTAC-3'
Fw_nsp12_1	Fw: 5'-CGTCTCCGCGGTGGAGGCACCACTATTGATCAGAGTTATTTAAAC-3'
Rv_nsp12_1	RV: 5'-GTGCGGCCTCAGTGGTGGTGATGATGGTGATGTTGCAAAACAGTGGATTTTTCATAGAG-3'
Fw_nsp12_2	FW: 5'-CCACTGTTTTGCAACATCACCATCATCACCACCACTGAGGCCGCACTCGAGGTCG-3'
Rv_nsp12_2	RV: 5'-CAATAGTGGTGCCTCCACCGCGGAGACGCAGGACCAGATG-3'
Fw_nsp12-FS-del	FW: 5'-CTAGTGCAGCTCGACTTGGAACCCTGTAATGGTACTGATCCAG-3'
Rv_nsp12-FS-del	RV: 5'-GTACCATTACAGGGTTCCAAGTCGAGCTGCACTAGAACCCCGCAC-3'
Fw_nsp12_SAA	FW: 5'-GATGATTTTATCTGCTGCTGGAGTTGTGTGCTACAACAAAG-3'
Rev_nsp12_SAA	Rev: 5'-CACAACTCCAGCAGCAGATAAAATCATCATAGAAAAGTGTTTTC-3'

SARS-CoV	
Fw_nsp7-L_1	FW: 5'-TCCGCGGTGGATCTAAAATGTCTGCTGACGTAAA-3'
Rv_nsp7-L_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGTGAAGCAATAGCCTGAAGAGTAGCAC-3'
Fw_nsp7-L_2	FW: 5'-CTACTCTTCAGGCTATTGCTTCACATCACCATCATCACCACTGAGGCCGCACTCGAG-3'
Rv_nsp7-L_2	RV: 5'-GACATTTTAGATCCACCGCGGAGACGCAGGAC-3'
Fw_nsp7-8-L_1	FW: 5'-TCCGCGGTGGATCTAAAATGTCTGCTGACGTAAA-3'
Rv_nsp7-8-L_1	RV: 5'-CTCAGTGGTGATGATGGTGATGACCAGATCCAGACTGTAGTTTAACAGCTGAGTTGGCTC-3'
Fw_nsp7-8-L_2	FW: 5'-GTTAAACTACAGTCTGGATCTGGTCATCACCATCATCACCACTGAGGCCGCACTCGAG-3'
Rv_nsp7-8-L_2	RV: 5'- GACATTTTAGATCCACCGCGGAGACGCAGGAC-3'
Fw_nsp7-8-9_1	FW: 5'-TCCGCGGTGGATCTAAAATGTCTGACGTAAA-3'
Rv_nsp7-8-9_1	RV: 5'-CTCAGTGGTGATGATGGTGATGCTGAAGACGTACTGTAGCAGCTAA-3'
Fw_nsp7-8-9_2	FW: 5'-ACGTCTTCAGCATCACCATCATCACCACTGAGGCCGCACTCGAGGTC-3'
Rv_nsp7-8-9_2	RV: 5'-GACATTTTAGATCCACCGCGGAGACGCAGGAC-3'

MERS-CoV	
Fw_nsp7-8-9_1	FW: 5'-GTCTCCGCGGTGGATCTAAACTTACAGATCTTAAATGCACATC -3'
Rv_nsp7-8-9_1	RV: 5'- GTGCGGCCTCAGTGGTGATGATGGTGATGTTGCAATCTAACAGTCGCAGCAATGTGCCCT -3'
Fw_nsp7-8-9_2	FW: 5'- CTGCGACTGTTAGATTGCAACATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGAC -3'
Rv_nsp7-8-9_2	RV: 5'- GATCTGTAAGTTTAGA TCCACCGCGGAGACGCAGGACCAGATG -3'
Fw_nsp7-8-9-10-11_1	FW: 5'- GTCTCCGCGGTGGATCTAAACTTACAGATCTTAAATGCACATC-3'
Rv_nsp7-8-9-10-11_1	RV: 5'- GTGCGGCCTCAGTGGTGATGATGGTGATGGTTTAAAAAATTGGAATCTTTAGATTG-3'
Fw_nsp7-8-9-10-11_2	FW: 5'-GATTCCAATTTTTTAAAC CATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGAC-3'
Rv_nsp7-8-9-10-11_2	RV: 5'-GATCTGTAAGTTTAGA TCCACCGCGGAGACGCAGGACCAGATG-3'
Fw_nsp12_1	FW: 5'-CTCCGCGGTGGATCTAAAGATTCCAATTTTTTAAACGAG-3'
Rv_nsp12_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGCTGCAAAGTGGTAGGCGAACTATAGAGATC-3'
Fw_nsp12_2	FW: 5'-GTTCGCCTACCACTTTGCAGCATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGAC-3'
Rv_nsp12_2	RV: 5'-GGAATCTTTAGATCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp12-FS-del	FW: 5'-GATTCCAATTTTTTAAACAGAGTCCGGGGTTCTATTG-3'
Rv_nsp12-FS-del	RV: 5'-GAACCCCGGACTCTGTTTAAAAAATTGGAATCTTTAG-3'
Fw_nsp12_SAA	FW: 5'-CTATGATGATACTGTCTGCTGCCGGTGTCGTTTGCTATAATAG-3'
Rv_nsp12_SAA	RV: 5'-CAAACGACACCGGCAGCAGACAGTATCATCATAGAAAAGTG-3'

EAV	
Fw_nsp4_1	FW: 5'-GATCGAGGGAAGGGGGCTATTCAGGTCACCGAAGG-3'
Rv_nsp4_1	Rv: 5'- GACGGCCAGTGCTTAGTGATGGTGATGATGGTGATGCTCTCTATTGGATAAGCCATCAAT-3'
Fw_nsp4_2	FW: 5'-CAATAGAGAGCATCACCATCATCACCATCACTAAGCACTGGCCGTCGTTTTACAA-3'
Rv_nsp4_2	RV: 5'-GACCTGAATAGCCCCCTTCCCTCGATCCCGAGGTTGTTG-3'
Fw_nsp6_1	FW: 5'-CGTCTCCGCGGTGGAGGAGGAGTGAAAGAGAGTGTC-3'
Rv_nsp6_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGTGCAGTGAGACTCTCCTGGGTAATTGG-3'
Fw_nsp6_2	FW: 5'-GGAGAGTCTCACTGCACATCACCATCATCACCACTGAGGCCGCACTCGAGGTCG-3'
Rv_nsp6_2	RV: 5'-CTTTCACTCCTCCTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp7-L_1	FW: 5'-CGTCTCCGCGGTGGAAGTCTCACTGCAACATTAGCTGCCCTC-3'
Rv_nsp7-L_1	RV: 5'-GTGCGGCCTCAGTGGTGATGATGGTGATGCTGATCTAGGCCTTCATAGCTCCCCTTGC-3'
Fw_nsp7-L_2	FW: 5'-CTATGAAGGCCTAGATCAGCATCACCATCATCACCACTGAGGCCGCACTCGAGGTCG-3'
Rv_nsp7-L_2	RV: 5'-GTTGCAGTGAGACTTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp7α_1	FW: 5'-CGTCTCCGCGGTGGAAGTCTCACTGCAACATTAGCTGCCCTC-3'
Rv_nsp7α_1	RV: 5'-CTCGAGTGCGGCCTCAGTGGTGATGATGGTGATGTTCTTCCTTAACGACGTCACAAATG-3'
Fw_nsp7α_2	FW: 5'-GTCGTTAAGGAAGAACATCACCATCATCACCACTGAGGCCGCACTCGAGGTCGACCTG-3'
Rv_nsp7α_2	RV: 5'-GTTGCAGTGAGACTTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp7β-L_1	FW: 5'-CTGCGTCTCCGCGGTGGAGCCAATGACACCCCAGTTAAGCCAATG-3'
Rv_nsp7β-L_2	RV: 5'-GAGTGCGGCCTCAGTGGTGATGATGGTGATGGTTTAACTGATTCACTGCCTCAACTG-3'
Fw_nsp7β-L_1	FW: 5'-GAGGCAGTGAATCAGTTAAACCATCACCATCATCACCACTGAGGCCGCACTCGAGGTC-3'

Rv_nsp7β-L_2	RV: 5'-CTGGGGTGTCATTGGCTCCACCGCGGAGACGCAGGACCAGATGTAAG-3'
Fw_nsp6-7-8_1	FW: 5'-CGTCTCCGCGGTGGAGGAGGAGTGAAAGAGAGTGTC-3'
Rv_nsp6-7-8_1	RV: 5'-GAGTGCGGCCTCAGTGGTGATGATGGTGATGGTTTAACTGATTCACTGCCTCAACTG-3'
Fw_nsp6-7-8_2	FW: 5'-GAGGCAGTGAATCAGTTAAACCATCACCATCATCACCACTGAGGCCGCACTCGAGGTC-3'
Rv_nsp6-7-8_2	RV: 5'-CTTTCACTCCTCCTCCACCGCGGAGACGCAGGACCAG-3'
Fw_nsp9_1	FW: 5'-CTCCGCGGTGGAGGCCTAGATCAGGACAAAGTG-3'
Rv_nsp9_1	RV: 5'-GAGTGCGGCCTCAGTGGTGGTGATGATGGTGATGCTCATACTGCTTGGTGCGGAAG-3'
Fw_nsp9_2	FW: 5'-CAAGCAGTATGAGCATCACCATCATCACCACCACTGAGGCCGCACTCGAGGTCG-3'
Rv_nsp9_2	RV: 5'-CCTGATCTAGGCCTCCACCGCGGAGACGCAGGAC-3'
Fw_nsp9_FS-Del	FW: 5'-CAGTGAATCAGCTTAACCTGAGAGCGCCCCACATCTTTCC-3'
Rv_nsp9_FS-Del	RV: 5'-GTGGGGCGCTCTCAGGTTAAGCTGATTCACTGCCTCAACTG-3'
Fw_nsp10_1	FW: 5'-CGTCTCCGCGGTGGAAGTGCCGTGTGCACAGTTTG-3'
Rv_nsp10_1	RV: 5'-GTGCGGCCTCAGTGGTGGTGATGATGGTGATGAATTTTGTTGGATTGCTTTTCCCAGC-3'
Fw_nsp10_2	FW: 5'-GCAATCCAACAAAATTCATCACCATCATCACCACCACTGAGGCCGCACTCGAGG-3'
Rv_nsp10_2	RV: 5'-CTGTGCACACGGCACTTCCACCGCGGAGACGCAGG-3'
Fw_nsp12_1	FW: 5'-CTGCGTCTCCGCGGTGGAGGTGTTGATGCAGTTACATCAGCAC-3'
Rv_nsp12_1	RV: 5'-GAGTGCGGCCTCAGTGGTGGTGATGATGGTGATGGGGCCCAATGACTGAACCAATTTG-3'
Fw_nsp12_2	FW: 5'-GGTTCAGTCATTGGGCCCCATCACCATCATCACCACCACTGAGGCCGCACTCGAGGTC-3'
Rv_nsp12_2	RV: 5'-GTAACTGCATCAACACCTCCACCGCGGAGACGCAGGACCAGATG-3'

Primers used for site-directed mutagenesis	
Amino acid substitution	
Fw_nsp9_K94A	FW: 5'-CTTTTTATTTAACATCGCATTTGTGTGCGACGAAGAGTTC-3'
Rv_nsp9_K94A	Rv: 5'-CGTCGCACACAAATGCGATGTTAAATAAAAAGTCC-3'
Fw_nsp9_K106A	FW: 5'-CACAAAGACCCCAGCAGACACACTGCTTGGGTACG-3'
Rv_nsp9_K106A	RV: 5'-GCAGTGTGTCTGCTGGGGTCTTTGTGAACTCTTC-3'
Fw_nsp9_R124A	FW: 5'-CTGGTTTATTTTCGCTCGTACGCACCGGTCGCTG-3'
Rv_nsp9_R124A	RV: 5'-GGTGCGTACGAGCGAAAATAAACCAGTAACCAG-3'
Fw_nsp9_S129A	FW: 5'-CGTACGCACCGGGCGCTGATTGATGCATACTGG-3'
Rv_nsp9_S129A	RV: 5'-CATCAATCAGCGCCCGGTGCGTACGACGGAAAATAAAC-3'
Fw_nsp9_D132A	FW: 5'-GGTCGCTGATTGCTGCATACTGGGACAGTATGG-3'
Rv_nsp9_D132A	RV: 5'-GTCCCAGTATGCAGCAATCAGCGACCGGTGCGTACGAC-3'
Fw_nsp9_D165A	FW: 5'-CGAGCGATGGGCTTTTGAATCTCCCGGAGGAGG-3'
Rv_nsp9_D165A	RV: 5'-GGAGATTCAAAAGCCCATCGCTCGCCCGTCACTG-3'
Fw_nsp9_F166A	FW: 5'-GAGCGATGGGATGCTGAATCTCCCGGAGGAGGCCGTG-3'
Rv_nsp9_F166A	RV: 5'-CTCCGGGAGATTCAGCATCCCATCGCTCGCCCGTCAC-3'
Fw_nsp9_D445A	FW: 5'-CCTTGAAACAGCCCTGGAGAGTTGTGATCGCTC-3'

Rv_nsp9_D445A	RV: 5'-CAACTCTCCAGGGCTGTTTCAAGGCAGTACTTG-3'

HCoV-229E; GenBank accession number AF304460, FIPV strain 79/1146; GenBank accession number DQ010921, SARS-CoV strain Frankfurt-1, Genbank accession number AY291315, PEDV strain CV777; accession number NC_003436, TGEV strain Purdue; accession number NC_038861, MERS-CoV strain HCoV-EMC; accession number NC_019843 and EAV Bucyrus strain; accession number NC_002532.

3.8 Bacterial culture

3.8.1 Bacterial growth medium and *E. coli* genotypes used

E. coli cells were grown in LB medium or LB-agar. LB consisted of 1 % [w/v] tryptone, 1 % [w/v] NaCl and 0.5 % [w/v] yeast extract in ddH_2O. LB-agar contained 1.5 % [w/v] agar in LB medium. Prior to use, the media were sterilized by autoclaving. *E. coli* Top10F' cells were used for cloning, whereas *E. coli* TB1 cells were used for heterologous protein expression.

3.9 Molecular cloning

3.9.1 First-strand cDNA synthesis

cDNA was synthesized by using SuperScript III (Invitrogen) according to the manufacturer's instructions. The reactions were stored at -20°C until further use.

3.9.2 Polymerase chain reaction (PCR)

Appropriate sense and antisense primers suitable to amplify the region of interest were used in the subsequent PCR as follows:

PCR reaction (total volume of 50 µl)	**Volume**
5 x HF-buffer or GC-buffer	10 µl
dNTPs (4 mM)	2.5 µl
Primer forward (Fw) and reverse (Rv) (10 pmol/µl)	2.5 µl each

DMSO (final concentration of 5%)	1-2 µl
Template (DNA or cDNA)	50-100 ng
Phusion high-fidelity DNA polymerase	1 unit
DNase-RNase free water	up to 50 µl

For PCR, the following conditions were used:

Cycle conditions			
	Cycle(s)	**Temperature**	**Time**
Initial denaturation	1	98°C	5 min
Denaturation	30-35	98°C	1 min
Primer annealing		50°C	45 sec
Extension		72°C	15–30 seconds/ kb
Final extension	1	72°C	5 min
Hold		4°C	∞

3.9.3 Agarose gel electrophoresis

To analyze the PCR product, the samples were separated on a 1 % [w/v] agarose gel (containing 0.1 µg/ml ethidium bromide) in 1x TAE-buffered (0.1 M Tris-Cl, 0.08 M boric acid and 1 mM EDTA; pH 8.2) agarose gel. Approximately 4 µl of sample was mixed with 1x loading dye (10 mM Tris-Cl, 0.03 % [w/v] bromophenol blue, 40 % [w/v] sucrose, 75 mM EDTA, pH 7.5), loaded and the agarose gel was run at a constant voltage of 100 V. The size of the PCR product was determined using a 1 kb DNA ladder (NEB or Fermentas).

3.9.4 Restriction- and ligation-free cloning (RF and LF)

As reported previously (Li and Breaker, 2001; van den Ent and Lowe, 2006; Yao et al., 1992) restriction- and ligation-free cloning was performed (Figure 8).

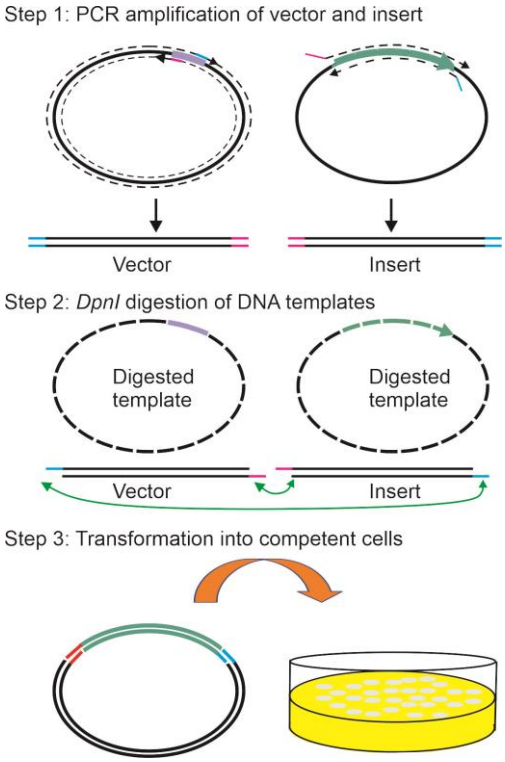


Figure 8: Restriction- and ligation-free cloning. Amplification of vector and insert by PCR and transformation into competent *E. coli* cells. Step 1 - PCR amplification of vector and insert DNA. Primers were designed to create overlapping tail regions (with identical sequence) of approximately 25-50 bp in the resulting insert and vector amplicons. Step 2 – DpnI digest. DpnI only cleaves the parental, methylated plasmid DNA. Step 3 - Transformation of PCR fragments into chemically competent *E. coli* (Yao et al., 1992).

3.9.5 Preparation of competent *E. coli* cells

Chemically *E.coli* competent cells were made using the following protocol (Lorsch, 2013). A single colony from an LB-agar plate was inoculated into 5 ml of LB medium and grown at 37°C (~ 225 rpm) overnight. The following day, 250 ml of fresh LB medium containing 20 mM $MgSO_4$ was inoculated with the 5 ml overnight culture and grown at 37°C at 225 rpm until OD_{600} 0.4-0.6. After that cell were pelleted by centrifugation at 4,500 x g for 5 minutes at 4°C. The cell pellet was gently resuspended in 0.4 volume (based on the original culture volume) of filter sterilized ice-cold TFB1 (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 1000 mM RbCl, 15 % [v/v] glycerol; pH 5.8 adjusted with acetic acid) and incubated on ice for 5 minutes. The cells were pelleted by centrifugation at 4,500 x g for 5 minutes at 4°C and gently reuspended

in 1/25th of the original culture volume in steril-filtered ice-cold TFB2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15 % [v/v] glycerol; pH 6.5 with KOH). The cells were incubated on ice for 60 minutes before 100 µl aliquots were shock-frozen and stored at -80°C until further use.

3.9.6 Transformation

The aliquoted 100 µl chemically competent *E. coli* cells were thawed on ice before the appropriate DNA was added to the cells and mixed gently. The mixture was incubated on ice for 30 min. Next, a heat shock was carried out for 45 sec at 42°C and the mixture was placed on ice for 5 min. Thereafter, 900 µl LB medium was added, mixed very gently, and incubated at 37°C for 1 h at 225 rpm. Finally, 100 µl of cells were streaked on a LB-agar plate containing the appropriate antibiotic(s) and incubated overnight at 37°C.

3.9.7 Plasmid mini preparation

A single *E. coli* colony was used to inoculate 5 ml of LB medium containing the appropriate antibiotic(s) and the culture was incubated at 37°C overnight at 225 rpm. The next day, the entire culture was pelleted down by centrifugation for 2 min at 6,000 x g. The supernatant was discarded, the cell pellet resuspended in 200 µl buffer 1 (50 mM Tris-HCl pH 7.5, 10 mM EDTA and 0.1 µg/ml RNase A) and incubated on ice for 2 minutes. Then, 200 µl of freshly prepared buffer 2 (200 mM NaOH, 1 % [w/v] SDS) was added, mixed gently, and incubated for 2 minutes for cell lysis. Subsequently, 200 µl of buffer 3 (3 M potassium acetate, pH 5.5) was added and mixed. The precipitate was pelleted by centrifugation at 21000 x g for 20 min at 4°C. The DNA was precipitated by adding 200 µl of isopropanol and pelleted by centrifugation at 21000 x g for 20 minutes at 4°C. The supernatant was discarded, the pellet washed with 200 µl

of 70 % [v/v] ethanol and air dried. Finally, the DNA pellet was dissolved in 100 µl RNAase/DNAase-free H_2O.

To verify the integrity, identity, and quality of the purified plasmid DNA, a restriction digest was carried out using suitable restriction enzymes. 200-300 ng of plasmid DNA was digested for 1 h at 37°C in a 10 µl reaction volume containing the appropriate 1 x reaction buffer and 5-10 units of enzymes. Half of the reaction mixture was analyzed on a 1 % [w/v] agarose gel. The pattern of DNA fragments was compared with the DNA marker (Thermo fisher). The positive clones verified by restriction digests were subsequently subjected to sequence analysis.

3.9.8 Site-directed mutagenesis of plasmids

Site-specific mutations were introduced into the gene of interest based on the recombination of DNA fragments with homologous ends both containing the desired nucleotide substitution in *E. coli* (Jones and Howard, 1991; Yao et al., 1992). To reduce misincorporation of nucleotides the *Phusion* high-fidelity polymerase (Thermo Fischer) was used. After confirming a positive PCR result on an agarose gel, the products were purified using a PCR purification kit (Qiagen) and the (methylated) plasmid DNA template (with the parental sequence) was digested with *DpnI* overnight at 37°C. Next, the products were use to transform chemically competent *E. coli* Top10F' cells. After the isolation of plasmid DNA, the correct introduction of the site-specific mutation was verified by sequencing.

3.10 RNA molecular methods

3.11.1 5'-end labeling of RNA

Radioactively labeled 5'-[^{32}P]-RNA oligonucleotides were produced through phosphorylation with T4 polynucleotide kinase (PNK) using [γ-^{32}P]-adenosine 5'-

triphosphate. The 10 µl reaction was set up on ice. The reaction mixture is listed in the following table.

Reaction condition for RNA 5'-end labeling	
10x PNK forward buffer A (Fermentas)	1 µl
RNA oligo (10 pmol/µl)	3 µl
[γ-P^{32}]-adenosine 5'-triphosphate (3000 Ci/mmol, 10 mCi/ml)	3 µl
PNK (10 U/µl, Fermentas)	1 µl
DNase/RNase-free H_2O	2 µl

The reaction was initiated by adding T4 polynucleotide kinase enzyme and the reaction mixture was incubated at 37°C. After 1 hour, the reaction was terminated by the addition of 40 µl of STE buffer (10 mM Tris-Cl pH 8.0, 0.5 mM NaCl, 10 mM EDTA pH 8.0). Finally, unincorporated nucleotides were removed by loading the entire reaction onto a microspin G-25 column (GE Healthcare) and centrifuged for 2 min at 735 x g.

3.11.2 RNA-RNA hybrid formation

To characterize primer dependent polymerase activity, partially double-stranded RNA substrates were prepared. For unlabeled substrates, 10 µM RNA and 13 µM complementary RNA were used; for 5'-[^{32}P]-labeled substrates 600 nM of the raiolabeled RNA was mixed with 780 nM unlabeled complementary RNA. All RNA substrates were prepared at a 1:1.3 ratio in STE buffer and incubated at 95°C for 5 min. Then, the heat block was switched off and the reaction was slowly cooled to room temperature.

3.11.3 Electrophoretic mobility shift assay

Protein-RNA binding assays were performed in a 10 µl volume containing 20 mM HEPES-KOH/pH 8.0, 30 mM KCl, 5 % [v/v] glycerol, 1.5 mM β-mercaptoethanol, 1

mM DTT, 0.2 % [v/v] Triton X-100, 60 nM duplex RNA substrate and 1 µM protein (or 1 µM of each of the proteins included) . The reaction was initiated by the addition of 1 mM Mg^{2+} and incubated for 1 h on ice. Finally, the samples were analyzed on a gradient (4 % and 12 %) native-PAGE run at a constant 200 V for 4 h at 4°C buffered in 0.5 x TBE. The gel was dried in a vacuum drier at 80°C for 1 h and exposed for 12 h to an imaging screen. The protein-RNA complexes were visualized by phosphor imaging (Typhoon 9200, Amersham).

3.11.4 *In vitro* RNA polymerase assays

RNA polymerase assays were carried out in a 10 µl volume containing 20 mM HEPES-KOH/pH 8.0, 30 mM KCl, 5 % [v/v] glycerol, 1 mM β-mercaptoethanol, 1 mM DTT, 0.001 % [v/v] Triton X-100, 1 µM RNA substrate, NTPs [(100 µM ATP,100 µM GTP, 100 µM CTP and 75 µM UTP) for *de novo* initiation, (100 µM NTPs, 0.17µM [α-^{32}P]-NTPs) for elongation, 1 µM of individual or a 1:1 ratio of enzyme. The reaction was initiated by adding either 1 mM Mg^{2+} or 1 mM Mn^{2+} for elongation, varying concentrations (10, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 µM for the metal ion titration or 0.24 mM Mg^{2+} or 0.1 mM Mn^{2+} for the nucleotide selection assay. The reaction was terminated after 1h at 30°C by the addition of 0.3 mM sodium acetate (pH 5.3) and 20 volumes of absolute ethanol. The reaction products were precipitated overnight at -20°C and subsequently pelleted for 1h at 4°C and 20854 x g. The supernatant was discarded, and the pellet was air dried followed by suspension in 5 µl of a 1:10 dilution of proteinase K and incubation at 55°C for 30 min. Following proteinase K treatment, 5 µl FU-MIX (80% [v/v] deionized formamide, 1x TBE, 0.1% [w/v] bromophenol blue, 0.1% [w/v] xylene cyanol and 6 M urea) was added and the RNA was denatured at 65°C for 10 mins. The products were separated for 1.5 h at 50W

in a 12% 7 M urea polyacrylamide gel in 1x Tris-Borate-EDTA buffer and analyzed using a Typhoon 9200 imager (GE healthcare) and Quantity One software (Bio-Rad).

Table 4: Metal ion and nucleotide concentrations used in different assays.

Activity assay	**Mg^{2+} or Mn^{2+} (mM)**	**NTPs, [α-^{32}P]-NTPs (µM)**
De novo initiation	1	100 µM ATP, 100 µM GTP, 100 µM CTP and 75 µM UTP), 0.17µM [α-^{32}p]-UTP
Elongation and nucleotide selection and fidelity	1	100 µM NTPs, 0.17µM [α-^{32}P]-NTPs
nucleotide selection and fidelity (reduced concentration)	0.24 or 0.1	100 µM NTPs, 0.17 µM [α-^{32}P]-NTPs
metal ion titration	0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	100 µM ATP, 100 µM GTP, 0.17 µM [α-^{32}P]-GTP

3.11.9 Nucleotidylation assays

Nucleotidylation assays were carried out in a 10 µl volume containing 20 mM HEPES-KOH/pH 8.0, 30 mM KCl, 5 % [v/v] glycerol, 0.2 % [v/v] Triton X-100, 1.5 mM β-mercaptoethanol, 2 mM DTT, NTPs (25 µM ATP or 25 µM GTP or 25 µM CTP or 25 µM UTP), 0.17 µM [α-^{32}P]-ATP or GTP or CTP or UTP (1 µCi/µl), 1 µM of individual or a 1:1 ratio of enzyme. The reaction was initiated by the addition of 1 mM Mg^{2+} or Mn^{2+} and incubated at 30°C for 1h and terminated by adding 1 x Laemmli loading buffer (50 mM Tris-HCl/pH 6.8, 2.5 % [w/v] SDS, 10 % [v/v] glycerol, and 0.01 % [w/v] bromophenol blue). The samples were denatured at 100°C for 10 min and analyzed on a 12 % Tris-tricine-PAGE. The gels were stained with 0.05 % [w/v] Coomassie brilliant blue G-250, 50 % [v/v] methanol and 10 % [v/v] acetic acid for 30 min and destained in 20 % [v/v] acetic acid. The destained gels were exposed to a phosphor imager screen

for 12 h and the radiolabeled products were visualized by phosphorimaging (Typhoon 9200, Amersham).

3.12 Heterologous protein expression in *E. coli*

3.12.1 Purification of main proteases

3.12.1.1 Expression

pMal-c2X-EAV-nsp4-H_6, pMal-c2X-HCoV-229E-nsp5-H_6 and pMal-c2X-FIPV-nsp5-H_6 were transformed into *E. coli* TB1 and grown on LB agar plates containing 100 µg/ml ampicillin. For the pMal-c2X-SARS-CoV-nsp5-H_6 construct, *E. coli* Rosetta cells were used. 2 liters of LB with 100 µg/ml Ampicillin were inoculated with 10 ml of preculture and grown at 37°C by shaking at 225 rpm. After reaching the OD_{600} of 0.4, the cells were cooled at 4°C for 30 minutes. Then, protein production was induced by adding 0.3 mM isopropyl β-D-thiogalactopyranoside (IPTG) and cells were grown at 18°C with shaking at 180 rpm for 16 hours.

3.12.1.2 Purification

The cells were pelleted at 3447 x g and resuspended in 20 ml lysis buffer per liter culture volume (20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 10 mM imidazole, 10 mM β-mercaptoethanol). The resuspended cells were incubated with a final concentration of 0.1 mg/ml of lysozyme for 30 min, rotating at 4°C. The cells were lysed by sonication (20 x 10 sec pulses, 10 sec on/off), cell debris was cleared by centrifugation for 50 min at 25200 x g at 4°C. The supernatant was collected and mixed with 2 ml Ni-NTA agarose equilibrated with 20 CV of lysis buffer. The lysate and Ni-NTA mixture was incubated for 3 h at 4°C. The matrix with the bound protein was washed with 20 CV of wash buffer (20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 30 mM imidazole, 10 mM β-

mercaptoethanol). Finally, the proteins were eluted with 20 ml elution buffer (20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 200 mM imidazole, 10 mM β-mercaptoethanol) and 5 µl of the elution fraction was analyzed on a 12 % Tris-tricine-PAGE gel.

3.12.1.3 Factor Xa cleavage

MBP-SARS-nsp5-His_6 auto-proteolytically releases itself from its N-terminal MBP tag during expression in *E. coli*. Therefore, the purified protein was directly dialyzed after the Ni-NTA affinity chromatography with 3 buffer exchanges in buffer containing 20 mM Tris-Cl/pH 7.4, 200 mM NaCl, 2 mM DTT. The dialyzed protein was stored at -80°C in 500 µl aliquots until further use.

The eluted MBP-EAV-nsp4-H_6, MBP-HCoV-229E-nsp5-H_6 and MBP-FIPV-nsp5-H_6 was dialyzed with 3 buffer exchanges in dialysis buffer (20 mM Tris-Cl/pH 7.4, 200 mM NaCl, 1 mM $CaCl_2$, 2 mM DTT). To seperate the proteases from their MBP tag, the fusion proteins were cleaved with factor Xa. The purified and dialyzed 20 ml of fusion protein was diluted with 20 ml of dialysis buffer, 20-30 µl of factor Xa (1 mg/ml) added and incubated at 4°C. After 2-3 days the percentage of cleavage was checked by loading 10 µl reaction mixture on a 12% Tris-tricine-PAGE. If the cleavage was not sufficient, an additional 10 µl of factor Xa was added to the reaction and incubated at 4°C further. The cleaved MBP+EAV-nsp4-H_6, MBP+HCoV-229E-nsp5-H_6 and MBP+FIPV-nsp5-H_6 mixtures were passed 5 times through 5-10 ml of amylose matrix (New England Biolabs) that had been pre-equilibrated with dialysis buffer to remove MBP and any unprocessed MBP-protease-H_6 fusion protein. As the last purification step, the collected EAV-nsp4-H_6, HCoV-229E-nsp5-H_6 or FIPV-nsp5-H_6 flow through from the amylose column was bound to 1 ml of Ni-NTA matrix for 30 min at 4°C on a rotating wheel and thereafter washed with 10 CV of wash buffer containing 20 mM

Tris-Cl/pH 7.4, 200 mM NaCl and 10 mM β-mercaptoethanol. The protease-H_6 was eluted with 10 ml elution buffer (20 mM Tris-Cl/pH 7.4, 200 mM NaCl, 200 mM imidazole and 10 mM β-mercaptoethanol). The eluted proteins were dialyzed in 3 buffer exchanges with buffer containing 20 mM Tris-Cl/pH 7.4, 200 mM NaCl and 2 mM DTT and 500 µl aliquots of the purified protein were prepared and stored at -80ºC until further use.

3.12.2 Nonstructural protein expression using the pASK-Ub system

3.12.2.1 Expression

All proteins except for the proteases were expressed as fusion proteins with an N-terminal ubiquitin tag and a C-terminal hexahistidine tag (His_6) using the pASK3-Ub expression system (Gohara et al., 1999; Lehmann et al., 2016; te Velthuis et al., 2010a; Tvarogova et al., 2019). The appropriate pASK-Ub-nsp-H_6 plasmid was transformed into *E. coli* strain TB1 containing the Ubp1 protease expression plasmid pCG1 and incubated over night at 37°C on LB plates containing 100 µg/ml Ampicillin and 34 µg/ml chloramphenicol. A preculture of 10-20 ml LB medium also containing ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) was inoculated with 1 colony and incubated at 37°C, 225 rpm overnight. This preculture was used to inoculate 2-4 liters of LB medium (5 ml of preculture/liter) (coronaviral nsp12 was grown in 10 liters culture) containing ampicillin (100 µg/ml) and chloramphenicol (34 ng/ml) and the culture was grown at 37°C at 225 rpm until an OD_{600} of 0.4 was reached. Then, the cells were cooled at 4°C for 30 min and protein expression was induced with anhydrotetracycline (final concentration 200 ng/ml) at 18°C, 180 rpm. After 16 h, the bacterial cells were harvested by centrifugation at 3447 x g for 15 min at 4°C and the pellet was resuspended in 20 ml cold lysis buffer (per liter bacterial culture) containing

20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 10 mM imidazole and 10 mM β-mercaptoethanol.

3.12.2.2 Purification

The cell suspension was treated with 0.1 mg/ml of lysozyme (Roth) together with EDTA-free protease inhibitor cocktail (Roche) for 30 min at 4°C. The cells were lysed by sonication (20 x 10 sec pulses,10 sec on/off) and the cell debris was cleared by centrifugation at 25200 x g for 50 min at 4°C. The cleared supernatant was incubated with 2 ml (per liter culture volume) of Ni-NTA matrix that had been pre-equilibrated with lysis buffer on a rotating platform for 3 h at 4°C. The Ni-NTA matrix with the bound protein was washed with 20 CV wash buffer containing 20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 30 mM imidazole and 10 mM β-mercaptoethanol. Finally, the proteins were eluted with 10-15 ml elution buffer containing 20 mM Tris-Cl/pH 8, 300 mM NaCl, 5 % [v/v] glycerol, 0.05 % [v/v] Tween-20, 200 mM imidazole and 10 mM β-mercaptoethanol. The purity was determined by loading 10 µl of protein on a Tris-tricine-PAGE and the subjected to a further purification step by anion-exchange chromatography (see below). Prior to the anion exchange chromatography, the samples were diluted 5-fold with elution buffer to prevent aggregation due to high protein concentration (except for coronavirus nsp12).

Table 5: pASK-UB-virus-nsp-His6 constructs used to express viral proteins

Plasmid construct used to produce viral proteins	Recombinant protein
pASK-Ub-HCoV-229E-nsp7-His_6	HCoV-229E nsp7-His_6
pASK-Ub-HCoV-229E-nsp8-His_6	HCoV-229E nsp8-His_6
pASK-Ub-HCoV-229E-nsp7-L-His_6	HCoV-229E nsp7-L-His_6
pASK-Ub-HCoV-229E-nsp8-L-His_6	HCoV-229E nsp8-L-His_6
pASK-Ub-HCoV-229E-nsp7-8-L-His_6	HCoV-229E nsp7-8-L-His_6
pASK-Ub-HCoV-229E-nsp7-8-9-His_6	HCoV-229E nsp7-8-9-His_6
pASK-Ub-HCoV-229E-nsp12-His_6	HCoV-229E nsp12-His_6
pASK-Ub-HCoV-229E-nsp12_mut-His_6	HCoV-229E nsp12_mut-His_6

pASK-Ub-PEDV-nsp7-8-L-His_6	PEDV nsp7-8-L-His_6
pASK-Ub-PEDV-nsp12-His_6	PEDV nsp12-His_6
pASK-Ub-PEDV-nsp12_mut-His_6	PEDV nsp12-mut-His_6
pASK-Ub-TGEV-nsp7-8-L-His_6	TGEV nsp7-8-L-His_6
pASK-Ub-TGEV-nsp12-His_6	TGEV nsp12-His_6
pASK-Ub-TGEV-nsp12_mut-His_6	TGEV nsp12-mut-His_6
pASK-Ub-FIPV-nsp7-L-His_6	FIPV nsp7-L-His_6
pASK-Ub-FIPV-nsp8-L-His_6	FIPV nsp8-L-His_6
pASK-Ub-FIPV-nsp7-8-L-His_6	FIPV nsp7-8-L-His_6
pASK-Ub-FIPV-nsp7-8-9-His_6	FIPV nsp7-8-9-His_6
pASK-Ub-FIPV-nsp12-His_6	FIPV nsp12-His_6
pASK-Ub-FIPV-nsp12_mut-His_6	FIPV nsp12-mut-His_6
pASK-Ub-SARS-CoV-nsp7-L-His_6	SARS-CoV nsp7-L-His_6
pASK-Ub-SARS-CoV-nsp8-L-His_6	SARS-CoV nsp8-L-His_6
pASK-Ub-SARS-CoV-nsp7-8-L-His_6	SARS-CoV nsp7-8-L-His_6
pASK-Ub-SARS-CoV-nsp7-8-9-His_6	SARS-CoV nsp7-8-9-His_6
pASK-Ub-SARS-CoV-nsp12-His_6	SARS-CoV nsp12-His_6
pASK-Ub-SARS-CoV-nsp12_mut-His_6	SARS-CoV nsp12_mut-His_6
pASK-Ub-MERS-CoV-nsp7-8-9-His_6	MERS-CoV nsp7-8-9-His_6
pASK-Ub-MERS-CoV-nsp7-8-9-10-11-His_6	MERS-CoV nsp7-8-9-10-11-His_6
pASK-Ub-MERS-CoV-nsp12-His_6	MERS-CoV nsp12-His_6
pASK-Ub-MERS-CoV-nsp12_mut-His_6	MERS-CoV nsp12_mut-His_6
pASK-Ub-EAV-nsp6-His_6	EAV nsp6-His_6
pASK-Ub-EAV-nsp7-L-His_6	EAV nsp7-L-His_6
pASK-Ub-EAV-nsp7α-His_6	EAV nsp7α-His_6
pASK-Ub-EAV-nsp7β-L-His_6	EAV nsp7β-L-His_6
pASK-Ub-EAV-nsp6-7-8-His_6	EAV nsp6-7-8-His_6
pASK-Ub-EAV-nsp9-His_6	EAV nsp9-His_6
pASK-Ub-EAV-nsp10-His_6	EAV nsp10-His_6
pASK-Ub-EAV-nsp12-His_6	EAV nsp12-His_6

3.12.3 Anion exchange chromatography

To largely remove RNA copurified from *E. coli*, anion exchange chromatography was performed using an ÄKTA prime plus (GE Healthcare). The eluates of the Ni-NTA chromatography purification were 10-fold diluted with buffer A (20 mM Tris-Cl/pH 8, 5 % [v/v] glycerol, 10 mM β-mercaptoethanol), loaded onto a HiTrap QHP column (1 ml; GE Healthcare). The column was washed with 30 CV wash buffer (20 mM Tris-Cl/pH 8, 30 mM NaCl, 5 % [v/v] glycerol, 10 mM β-mercaptoethanol). Finally, the

protein was eluted in 1.5 ml fractions using a continuous NaCl gradient from 30 mM to 1000 mM. The eluted fractions were analyzed by SDS-PAGE and proteins were visualized by Coomassie brilliant blue staining.

The unprocessed, coronaviral precursor proteins nsp7-L-H_6, nsp8-L-H_6, nsp7-8-9-H_6 and nsp7-8-9-10-11-H_6 were dialysed with 3 buffer exchanges into 20 mM Tris-Cl/pH 8, 200 mM NaCl and 10 mM β-mercaptoethanol for proteolytic cleavage.

The processed tag free nsp(7+8) complexes of coronavirus and EAV nsp7-L-H_6, nsp7α-H_6, nsp7β-L-H_6 and nsp6-7-8-H_6 were further purified by gel filtration chromatography. Eluate fractions of the respective coronavirus nsp12-H_6 proteins and equine arteritis virus nsp9-H_6, nsp10-H_6 and nsp12-H_6 were dialyzed into buffer containing 50 mM HEPES-KOH/pH 8, 150 mM KCl, 40 % [v/v] glycerol and 10 mM β-mercaptoethanol and the protein concentration was determined by Bradford assay.

3.12.4 Proteolytic processing of coronavirus nsp7-L-H_6, nsp8-L-H_6 and polyprotein precursor substrates using nsp5-H_6

To obtain tag-free nsp7, nsp8 and a nsp(7+8) complex, respectively, the nsp7-L-H_6, nsp8-L-H_6, nsp7-8-L-H_6, nsp7-8-9-H_6 and nsp7-8-9-10-11-H_6 precursor substrates were subjected to nsp5-mediated proteolysis. To this end, the anion exchange-purified protein (50-60 ml, with a concentration of approximately 2 mg/ml) was incubated with 0.2-0.3 mg/ml final concentration nsp5-H_6 in a buffer containing 20 mM Tris-Cl/ pH 8, 200 mM NaCl and 10 mM β-mercaptoethanol. After 48 h at 4ºC, 10 µl of the cleavage reaction was mixed with 4 x SDS-gel loading buffer, denatured for 5 min at 95°C and analyzed on a 12% Tris-tricine-PAGE to determine the percentage of cleavage. The gel was stained with 0.05 % [w/v] Coomassie brilliant blue G-250, 10 % [v/v] methanol and 5 % [v/v] acetic acid for 30 minutes and destained in 20 % [v/v] acetic acid solution. Once cleavage was complete, the reaction mixture was loaded onto a pre-equilibrated

Ni-NTA affinity column (20 mM Tris-Cl/pH 8, 200 mM NaCl, 10 mM β-mercaptoethanol) and incubated for 30-60 min before passing through the column. The generated nsp7, nsp8 proteins or nsp(7+8) complexes with their authentic N- and C-termini were in the flowthrough fractions whereas nsp5-H_6, L-H_6 and nsp9-H_6 remained bound to the Ni-NTA. The tag-free individual nsp7, nsp8 proteins and the pre-formed nsp(7+8) complexes were further purified to near homogeneity by ion exchange chromatography using a continuous NaCl gradient at 4°C.

3.13 Size-exclusion chromatography

3.13.1 Purification of proteins using size-exclusion chromatography

To further purify the proteins, an ÄKTA purifier (GE Healthcare) connected to a UV detector and fraction collector was used. The columns were rinsed with 2 CV of water and equilibrated with 3 CV of buffer at a flow rate of 0.5 ml/min before loading the sample. The UV absorbance was measured at 280 nm.

The anion exchange-purified proteins of HCoV-229E, FIPV and SARS-CoV that contained tag at C-terminus (nsp7-His_6 and nsp8-His_6) and tag-less at C-terminus of nsp7, nsp8 proteins and pre-assembled nsp(7+8) complexes obtained from nsp7-8-9-H_6 precursors are finally purified by using a prep grade HiLoad 16/60 (superdex 200 pg) size-exclusion chromatography column. The protein samples were concentrated to 5-10 ml, loaded onto the pre-equilibrated column (20 mM Tris-Cl/pH 8, 150 mM NaCl, 2 mM DTT) and 1 ml eluate fractions were collected.

For EAV, the nsp7-L-H_6, nsp7α-H_6, nsp7β-H_6 and nsp6-7-8-H_6 proteins (previously purified by anionexchange chromatography) were subjected to a final size-exclusion chromatography step using a prep grade HiLoad 16/60 Superdex 75 pg column. To this end, the samples were concentrated to 5-10 ml and loaded onto the column in buffer containing 20 mM Tris-Cl/pH 8.0, 200 mM NaCl, 2 mM DTT and 1 ml eluate fractions

were collected. The protein containing fractions were pooled and half of the protein sample was dialyzed in dialysis buffer (50 mM HEPES-KOH/pH 8, 150 mM KCl, 40 % [v/v] glycerol and 10 mM β-mercaptoethanol) and aliquots were stored at -20°C for subsequent activity assays. The remaining part was concentrated using Amicon Ultra 15 Centrifugal Filters (10,000 MWCO) (Millipore) and used to assess the elution profile of protein complexes using an analytical size-exclusion column.

A molecular weight marker mix (ribonuclease A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (44 kDa), conalbumin (75 kDa) and aldolase (158 kDa)) was used to assess the molecular weight of the peak fractions. The distribution coefficients (K_{av}) of known molecular weight standards were determined and a linear standard trendline was plotted by employing K_{av} values against the logarithmic values of the corresponding molecular weight protein. By employing peak retention volume values of nsp(7+8) complexes to the standard curve, the approximate molecular weights were determined.

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

K_{av} = gel phase distribution coefficient, V_e = elution volume, V_0 = void volume, V_c = column volume.

3.13.2 Analytical size-exclusion chromatography

An ÄKTA purifier (GE Healthcare) equipped with an UV detector and fraction collector was used for analytical gel filtration chromatography. A Superdex increase 200 10/300 GL column was rinsed with 2 CV of water followed by the equilibration with 3 CV of buffer at a flow rate of 0.15 ml/min, before 100 µl samples were injected onto the column. Elution fraction volume was 0.5 ml.

5 mg/ml of nsp7-H_6, nsp8-H_6 proteins, 10 mg/ml of nsp7-H_6+8-H_6 seperately purified and mixed proteins or 10 mg/ml of HCoV-229E, FIPV and SARS-CoV pre-assembled

nsp(7+8) complexes generated from nsp7-8-9-His_6 precursors were loaded onto the column and absorbance was detected at 280 nm. These experiments were performed in buffer containing 20 mM HEPES-KOH/pH 8, 150 mM KCl, 2 mM DTT.

The molecular weight of the peak fractions was calculated using a molecular weight marker as described in the chapter before.

Table 6: Columns used for purification and analysis of proteins.

Column	Company
HiLoad Superdex 75 pg 16/60	GE Healthcare
HiLoad Superdex 200 pg 16/60	GE Healthcare
Superdex increase 200 10/300 GL	GE Healthcare

3.14 Protein crosslinking

To determine the multimeric states of nsp7, nsp8, nsp7+8 (purified separately) and preformed nsp(7+8) complexes in solution, protein crosslinking was performed using either glutaraldehyde or BS^3. A standard reaction was set up on ice in a 10 µl volume containing 20 mM HEPES-KOH/pH 8, 30 mM KCl, 2 mM β-mercaptoethanol and 5 µg protein. Glutaraldehyde crosslinking was started by the addition of 0.5 µl glutaraldehyde (0.005 % [v/v] final concentration) and the reaction was incubated at 30°C for 10 minutes. BS^3 crosslinking was started by the addition of 2 µl of BS^3 (final concentration 10 µM). The samples were incubated at 37°C for 30 minutes by shaking at 500 rpm, quenched by the addition of 0.5 µl $(NH)_4HCO_3$ (final concentration 50 mM) and incubated for another 30 minutes by shaking at 37°C. For both crosslinking methods, the reactions were terminated by the addition of 4x Laemmli buffer and the proteins were separated on a 12% Tris-tricine SDS-PAGE. The gel was stained with 0.5 % [w/v] Coomassie brilliant blue G-250, 50 % [v/v] methanol and 10 % [v/v] acetic acid for 30 minutes and destained in 20 % [v/v] acetic acid solution.

3.15 Mass spectrometry

3.15.1 Native mass spectrometry

The native mass spectroscopy experiments were performed by Dr. Boris Krichel at the Heinrich Pette Institute, Hamburg, according to the following protocols. Samples for native MS measurements were buffer exchanged into an ESI compatible solution. Precursors were buffer exchanged into 300-500 mM ammonium acetate (99.99 % purity based on trace metals analysis, Merck Millipore), 1 mM DTT, pH 8 by five rounds of dilution and concentrated in centrifugal filter units (Amicon, 10,000 MWCO). Some precursors required an increase of the salt concentration to 500 mM AmAc to remain soluble. Concentrations were adjusted with theoretical extinction coefficient and absorption at 280 nM. To start the processing reactions, nsps and nsp5 were incubated (ratio ~1:5–1:10). Nano ESI capillaries were pulled in-house from borosilicate capillaries (1.2 mm outer diameter, 0.68 mm inner diameter, filament, World Precision Instruments) with a micropipette puller (P-1000, Sutter instruments) using a squared box filament (2.5 × 2.5 mm, Sutter Instruments) in a two-step program. Subsequently, capillaries 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. To acquire mass spectra time-points, sample aliquots of 1-3 µl were withdrawn by means of a microliter syringe (5 µL, Hamilton) with flexible fused silica tubing (Optronis) and loaded into in house fabricated nanoESI capillaries. Native MS was performed at an electrospray quadrupole time-of-flight (ESI-Q-TOF) instrument (Q-TOF2, Micro mass/Waters, MS Vision) modified for higher masses (van den Heuvel et al., 2006). Samples were ionized in positive ion mode with voltages applied at the capillary of 1300-1500 V and at the cone of 130-135 V. The pressure in the source region was kept at 10 mbar throughout all native MS experiments. For purpose of desolvation and

dissociation, the pressure in the collision cell was adjusted to 1.5×10^{-2} mbar argon. Native like spectra were obtained at accelerating voltage of 10-30 V while for collision-induced dissociation these voltages were increased to 30-200 V. In ESI-MS overview spectra for nsps, the quadrupole profile was 1-10,000 *m/z*. In tandem MS, for precursor selection, LMres and HMres were adjusted at 10-30 V collisional voltage until a single peak was selected, then dissociation was induced. To calibrate raw data, CsI (25 mg/ml) spectra were acquired, and calibration was carried out with Mass Lynx (Waters) software. To analyze the data the raw spectra were smoothed (2x5) in Mass lynx 4.1 (Waters) and the nsps were assigned to peak series according to their theoretical mass based on their amino acid sequence.

3.15.2 Measurement of intact protein masses of EAV proteins

These experiments were performed by Dr. Uwe Linne, Department of Chemistry, Philipps University of Marburg. Approximately 1-10 µl of buffered protein solution was desalted using a Waters ACQUITY H-Class HPLC-system equipped with a Mass Prep column (Waters). Desalted proteins were eluted into the ESI source of a Synapt G2Si mass spectrometer (Waters) by the following gradient of buffer A (0.05 % formic acid and H_2O) and buffer B (acetonitrile and 0.045 % formic acid) at a column temperature of 60°C and a flow rate of 0.1 ml/min: Isocratic elution with 5 % A for two minutes, followed by a linear gradient to 95 % B within 8 minutes and holding 95 % B for additional 4 minutes. Positive ions within the mass range of 500-5000 m/z were detected. Glu-Fibrinopeptide B was measured every 45s for automatic mass drift correction. Averaged spectra were deconvoluted after baseline subtraction and eventually smoothing using Mass Lynx instrument software with MaxEnt1 extension.

3.16 Bradford assay

A BSA standard was prepared in buffer containing 20 mM HEPES-KOH/pH 8, 150 mM KCl and 10 mM β-mercaptoethanol with final BSA concentrations of 1, 2, 4, 6 and 8 µg/µl of protein. Then, the 5x Bradford reagent was diluted with the same buffer 1:5. 1µl of BSA standard was added to 1 ml of 1x Bradford reagent and mixed. The absorbance was measured in triplicates at 595 nm using a spectrophotometer and a linear regression graph was generated for the BSA standards. To determine the unknown protein concentration, 2.5, 5 and 10 µl of protein sample was mixed with 1 ml of 1 x Bradford reagent and measured at OD_{595}. By applying these OD_{595} values to the BSA standard curve, the unknown protein concentration was determined.

3.17 Tris-tricine-PAGE

To analyze proteins or protein complexes by Tris-tricine SDS-PAGE (Haider et al., 2012; Schagger, 2006; Simpson, 2006) the samples were mixed with 4 x SDS gel-loading buffer (Laemmli, 1970) at a ratio of 1:1 and denatured at 100˚C for 10 min. The samples were loaded onto a 12% Tris-tricine-PAGE gel with Tricine-PAGE compatible buffer and run for 4 h at 35 mA. The gel was stained for 20-30 min in Coomassie brilliant blue solution (0.5 % [w/v] Coomassie brilliant blue G-250, 50 % [v/v] methanol, 10 % [v/v] acetic acid) and destained in 40 % [v/v] methanol and 5 % [v/v] acetic acid.

Resolving gel (12%)	Volume
3 M Tris-Cl/pH 8.45 & 0.3 % [w/v] SDS	3.32 ml
Rotiphorese® gel (30%)	Depending on gel percentage
100 % [v/v] glycerol	1 ml
10 % [w/v] APS	0.030 ml
TEMED	0.010 ml
H_2O	to 10 ml

Stacking gel (4%)	**Volume**
3 M Tris-Cl/pH 8.45 & 0.3 % [w/v] SDS	0.744 ml
Rotiphorese® gel (30%)	0.295 ml
10 % [w/v] APS	8.4 µl
TEMED	3 µl
H_2O	to 3 ml

Running buffer (cathode buffer)	**Concentration**
Tris-Cl/pH 8.9	25 mM
Tricine	25 mM
SDS	0.1%

Running buffer (anode buffer)	**Volume**
Tris-Cl/pH 8.9	25 mM

4 x SDS gel-loading buffer	**Concentration**
Glycerol	40 % [v/v]
SDS	8 % [w/v]
DTT	400 mM
Tris-Cl (pH 6.8)	200 mM
Bromophenol blue	0.4 % [w/v]

4 Results

4.1 Expression system to produce coronavirus nonstructural proteins

All proteins for this study (except coronavirus $3CL^{pro}$) were expressed by using a tetracycline inducible promoter system (pASK3) (Gohara et al., 1999) to exclude possible bacteriophage T7 polymerase contamination in subsequent polymerase assays. An earlier report from te Velthuis *et al.* has described the importance of a natural N-terminus for SARS-CoV nsp12 (te Velthuis et al., 2010a). Therefore, coronaviral nsps were expressed as N-terminal ubiquitin-tagged fusion proteins with a C-terminal His tag (Ub-nsp-H_6) (Figure 9). The Ub-nsp fusion protein is cleaved between the C-terminus of ubiquitin and the N-terminus of the nsp at the ubiquitin carboxyl-terminal hydrolase-1 recognition site LRGG|XXX, producing a C-terminal cleavage product containing the desired (authentic) N-terminus (e.g. Ser, Ala, Asn) and allowing purification via its C-terminal hexahistidine tag by Ni-IMAC affinity chromatography (Gohara et al., 1999).

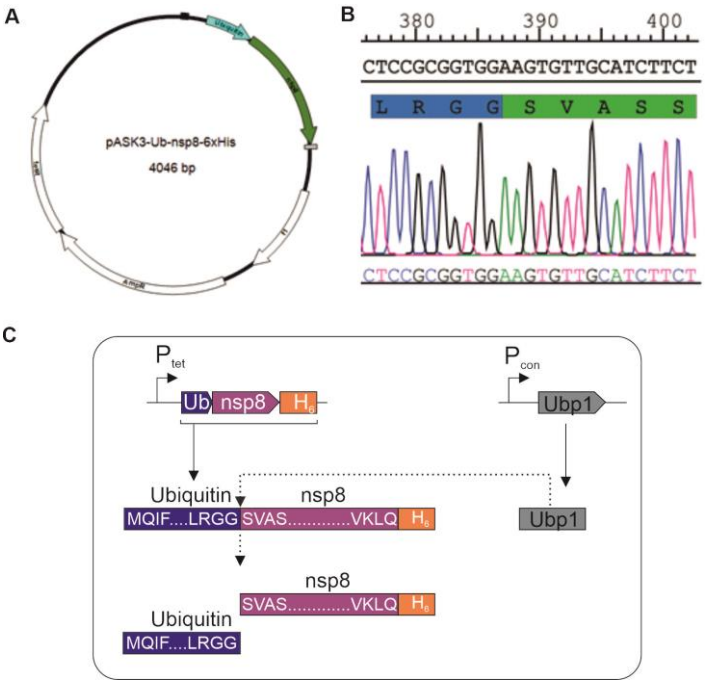


Figure 9: Schematic representation of the Ub-nsp8-H_6 construct which is shown as an example for co-expression of a Ub-tagged nidovirus protein with Ubp1 in *E. coli*. (A) The Ub-nsp8-H_6 coding sequence was inserted into pASK3, downstream of a *tet* promoter. The ubiquitin sequence is highlighted in blue and the nsp8 sequence is indicated in purple. **(B)** The DNA sequencing chromatogram shows amino acid residues corresponding to the C-terminus of ubiquitin (LRGG) and N-terminus of the HCoV-229E nsp8 (SVAS---). ↓ represents the cleavage site. **(C)** The ubiquitin-specific protease (Ubp1) is encoded by pCGI and recognizes the cleavage site LRGG↓XXX. After cleavage, nsp8-H_6 carries its authentic N-terminus.

4.2 Multimeric states of HCoV-229E nsp7-H_6 and nsp8-H_6

Previously, different (heterotrimeric versus hexadecameric) multimeric states have been reported for alpha- and betacoronavirus nsp7 and 8 complexes (Xiao et al., 2012; Zhai et al., 2005). Therefore, our first goal was to determine the nsp7 and nsp8 complex formation of HCoV-229E. The appropriate pASK-Ub-nsp-H_6 plasmid was transformed into *E. coli* TB1 [pCGI] cells. 2 liters of LB containing ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) were inoculated with 10 ml overnight pre-culture and cells were grown at 37°C until OD_{600} 0.4. At this stage, a 1 ml aliquot was taken as negative control ("non-induced") before adding anhydrotetracycline (200 ng/ml final concentration) for 16 h at 18°C, 180 rpm. 1 ml of post induced cells were collected as positive control ("induced"). The induced cells were harvested by centrifugation and Ni-IMAC purification was performed as described in Material and Methods. Figure 10 provides an example for the nsp7-H_6 and nsp8-H_6 expression and purification by Ni-IMAC chromatography. Eluate fractions containing nsp7-H_6 and nsp8-H_6 (Figure 10; lanes 7 and 15) were analyzed on a 12 % Tris-tricine-PAGE, along with total lysates of induced (Figure 10; lanes 2 and 9) and non-induced *E. coli* cells (Figure 10; lane 1 and 8), flow-through (Figure 10; lane 3 and 10) and wash fractions (Figure 10; lanes 4-6 and 11-14).

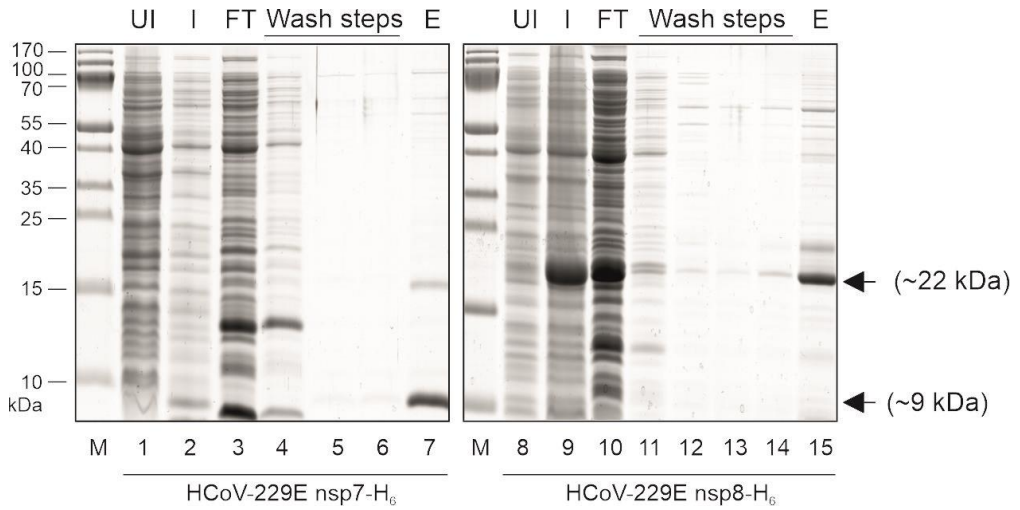


Figure 10: Tris-tricine-PAGE of HCoV-229E nsp7-H_6 and nsp8-H_6 purification by Ni-NTA affinity chromatography. TB1-pCGI cells were transformed with pASK-Ub-nsp7-H_6 and pASK-Ub-nsp8-H_6. Protein expression was monitored on a 12 % Tris-tricine-PAGE. Expression was induced with 200 ng/ml anhydrotetracycline for 16 h at 18°C. Following the confirmation of protein expression, nsp7-H_6 and nsp8-H_6 were purified by Ni-IMAC affinity chromatography. 5 µl of Ni-IMAC purified eluate of nsp7-H_6 and nsp8-H_6 were loaded. The arrows on the right side indicate purified HCoV-229E nsp7-H_6 and nsp8-H_6 with sizes that correspond with the calculated molecular weight of 9 and 22 kDa, respectively. Abbreviations: M – prestained page ruler; UI - uninduced (Lane 1 and 8); I - induced (Lane 2 and 9); FT - flow-through (Lane 3 and 10); wash (Lanes 4, 5, 6, 11, 12, 13, and 14); E - elution fractions (Lanes 7 and 15). Size markers are depicted on the left side in kDa.

To remove co-purified nucleic acids and several minor protein contaminants, the Ni-NTA eluate fractions were subjected to anion-exchange chromatography as described in Material and Methods. The main fractions of nsp7 and nsp8 were found to elute between 200 to 350 mM NaCl and dialyzed against buffer containing 50 mM HEPES-KOH/pH 8, 150 mM KCl, 10 mM β-mercaptoethanol. The Tris-tricine-PAGE analysis shows nsp7-H_6 and nsp8-H_6 with sizes of ~9 kDa and ~22 kDa (corresponding to their calculated molecular weights) (Figure 11; lanes 1 and 2).

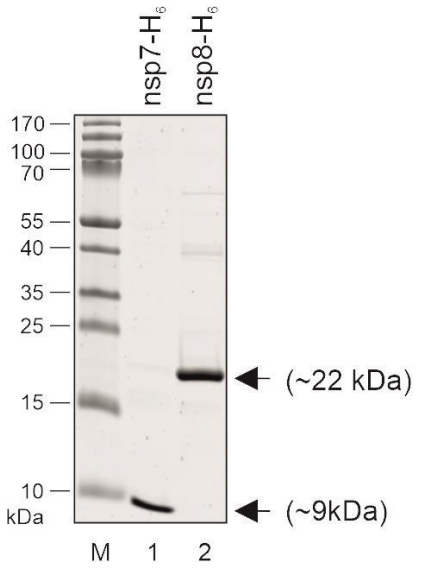


Figure 11: Tris-tricine-PAGE analysis of purified HCoV-229E nsp7-His$_6$ and nsp8-His$_6$ proteins after anion-exchange chromatography. 3 µg of anion-exchanged protein dialyzed against 20 mM HEPES-KOH/pH 8, 150 mM KCl, 10 mM β-mercaptoethanol were analyzed on a Tris-tricine-PAGE. Lane 1; nsp7 and Lane 2; nsp8. Abbreviations: M – prestained page ruler marker. Purified nsp7 and nsp8 with their predicted molecular masses of 9 kDa and 22 kDa are indicated with arrows on the right side of gel.

To determine oligomeric states of complexes formed by nsp7 and nsp8, 5 µg of HCoV-229E nsp7-H_6 and nsp8-H_6 were mixed and crosslinked using either 0.005 % [v/v] glutaraldehyde or 10 µM BS^3. As shown in Figure 12, both nsp7-H_6 and nsp8-H_6 yielded clear homodimer species with both crosslinking agents (Figure 12; lanes 2, 6 for glutaraldehyde crosslinking and lanes 8 and 12 BS^3 for crosslinking). Additionally, the formation of a homotrimer was detected for nsp7 with both crosslinking methods while nsp8 showed some higher-order species with poor resolution (Figure 12; lane 4, 6, 8, 12). Interestingly, the data suggested that a nsp7+8 heterodimeric stable complex (Figure 12; lane 4 and 10) was formed when nsp7 and nsp8 were mixed. The molecular mass of this complex (31 kDa) corresponded to the calculated molecular mass for a 1:1 dimeric complex of nsp7 and nsp8. This protein species was not detected when no crosslinking reagent was included in the reaction, suggesting that the detected nsp7+8

complex is truly a stable heterodimeric complex. Larger complexes could not be resolved clearly under the conditions used in our experiments.

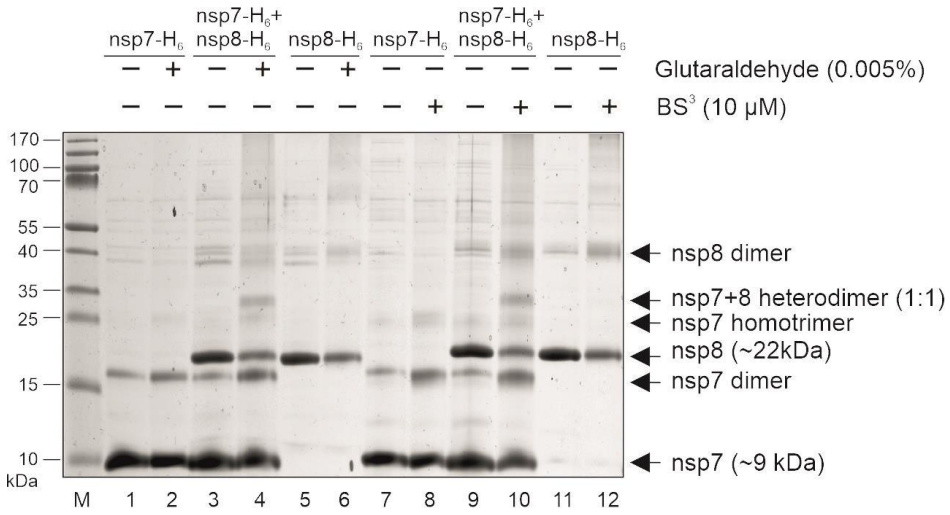
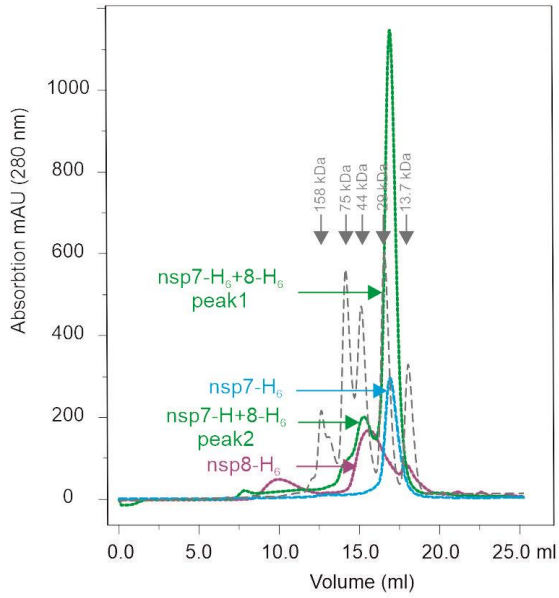


Figure 12: Tris-tricine-PAGE analysis of HCoV-229E nsp7-H$_6$ and nsp8-H$_6$ crosslinking. Crosslinking of 5 µg protein was carried out using either 0.005 % [v/v] glutaraldehyde or 10 µM BS^3. Crosslinking was performed at room temperature in buffer containing 20 mM HEPES-KOH/pH 8, 30 mM KCl and 10 mM β-mercaptoethanol. Glutaraldehyde crosslinking was carried out for 10 minutes whereas BS^3 crosslinking reactions were performed for 30 minutes, after which the reactions were quenched by adding 50 mM ammonium bicarbonate and incubated for another 30 minutes. All reactions were terminated by the addition of SDS loading buffer and analyzed on a 12 % Tris-tricine-PAGE. Lanes 1, 2, 3, 4, 5, 6 show glutaraldehyde crosslinking samples with controls and Lanes 7, 8, 9, 10, 11, 12 show BS^3 crosslinking samples with controls. Lanes 1, 7 non-crosslinked nsp7, Lanes 2, 8 crosslinked nsp7, Lanes 3, 9 non-crosslinked nsp7+nsp8, Lanes 4, 8 crosslinked nsp7+nsp8, Lanes 5, 11 non-crosslinked nsp8 and Lanes 5, 11 are crosslinked nsp8. Abbreviations: M – prestained page ruler, BS^3 – Bis [sulfosuccinimidyl]. The molecular weights of the crosslinked and non-crosslinked proteins are indicated with arrows on the right side of the gel.

In addition to the crosslinking experiments, 5 mg of nsp7-H$_6$ and nsp8-H$_6$ proteins were separated on an analytical size-exclusion column to assess their multimeric states. Consistent with the the crosslinking results, nsp7-H$_6$ eluted as a (presumed) homodimer indicated by its approximate molecular mass of ~18 kDa. For nsp8, a major peak at approximately ~44 kDa and several other minor peaks including peaks indicating the formation of higher-order oligomeric species were observed (Figure 13). The 44 kDa peak corresponds to the size expected for a nsp8-H$_6$ homodimer. To assess nsp7+8 complex formation in solution, a 1:1 ratio of separately purified nsp7-H$_6$+nsp8-H$_6$ (10 mg/ml each) were mixed and separated by analytical size-exclusion chromatography.

The elution profile obtained in this experiment indicated that nsp7-H_6 and nsp8-H_6 eluted separately, most likely, as homodimeric complexes rather than forming heterodimeric or higher-order complexes. A possible reason for the two proteins' inability to form complexes was the presence of hexahistidine tag at the C-terminus of both proteins which might have affected the folding of the proteins and, as a consequence, interactions that trigger 7+8 multimeric complex formation, as shown previously in studies of the poliovirus 3Dpol and SARS-CoV nsp12 (Gohara et al., 1999; te Velthuis et al., 2010a; te Velthuis et al., 2012). For example, te Velthuis et al. reported that the addition of a tag at either the N- or C-terminus altered the complex formation and affected the enzymatic ("non-canonical" polymerase) activity of nsp8 (te Velthuis et al., 2012).

A



B

	nsp7-H_6	nsp8-H_6	nsp7-H_6+8-H_6	
			peak 1	peak 2
V_e (ml)	16.86	15.27	16.86	15.27
K_{av}= V_e-V_o/V_c-V_o	0.542	0.431	0.542	0.431
Estimated MW	18.5 kDa	43.7 kDa	18.5 kDa	43.7 kDa

Figure 13: Size-exclusion chromatography analysis of the oligomerization states of C-terminally His_6 (C-H_6) tagged HCoV-229E nsp7, nsp8 and nsp7+nsp8. (A) Size-exclusion chromatograms. 5 mg/ml of individual protein or 10 mg/ml of a 1:1 ratio of nsp7-H_6+nsp8-H_6 was injected onto a Superdex 200 (10/300) analytical column. Prior to this analysis, all proteins

had been purified by Ni-NTA affinity and anion-exchange chromatography. nsp7-H_6: sky blue chromatogram, nsp8-H_6: pink chromatogram, nsp7-H_6+nsp8-H_6 complex: green chromatogram. Nsp7-H_6 eluted at the retention volume of 16.86 ml and nsp8-H_6 at 15.27 ml. The chromatogram showed that nsp7-H_6 and nsp8-H_6 eluted separately as individual proteins. X-axis represents elution fraction number, and the Y-axis represents UV absorbance at 280 nm. Molecular weight markers RNase A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (44 kDa), conalbumin (75 kDa) and aldolase (158 kDa) were analyzed on the the same column (gray chromatogram). **(B)** The sizes of nsp7-H_6, nsp8-H_6, nsp7-H_6+nsp8-H_6 proteins were estimated by using their retention elution volumes (V_e) and the void volume (V_o) marker (blue dextran 2000) and the column volume (V_c). Using these values, the gel phase diffusion coefficiency was determined by $K_{av}=V_e-V_o/V_c-V_o$. The retention volume (16.86 ml) of nsp7-H_6 corresponds to a molecular mass of approximately 18.5 kDa, indicating a homodimer, while nsp8-H_6 with its retention volume of 15.27 ml corresponded to an approximate molecular weight 43.7 kDa, indicating a nsp8 homodimer.

4.3 Scheme for the generation of nsp7, nsp8 and nsp(7+8) complexes with authentic N- and C-termini

To avoid potential deleterious effects of artificial residues at the N- and C-termini of the proteins, we modified our protein expression and purification strategy. As described above, the Ub-nsp fusion proteins expressed in *E. coli* were cleaved using the ubiquitin carboxyl-terminal hydrolase-1, resulting in proteins with authentic N-terminus. To ensure that the expressed nsp7 and nsp8 proteins also carried their authentic C-termini, the expression plasmids were engineered to encode a short four-amino-acid linker (L) between the respective protein's authentic C-terminus and the adjacent C-terminal H_6 tag. These additional residues, SVAS or NNEI, corresponded to the 4 N-terminal residues of nsp8 or nsp9, respectively, corresponding to the P1'-to-P4' positions of the nsp7|nsp8 and nsp8|nsp9 nsp5 (M^{pro}) cleavage sites (Figure 14). These proteins were purified by Ni-IMAC using their C-terminal H_6 tag and subsequently cleaved by using recombinant coronavirus nsp5 protease to remove any artificial amino acids from the C-terminus (↓L-H_6). To generate pre-assembled nsp(7+8) complexes, the nsp7-8-9-H_6 polyprotein was used as substrate for the nsp5 protease. Using this approach, nsp7 and nsp8 were found to form stable nsp(7+8) complexes. Using Ni-NTA affinity chromatography, L-H_6, nsp9-H_6 and nsp5-H_6 could be separated.

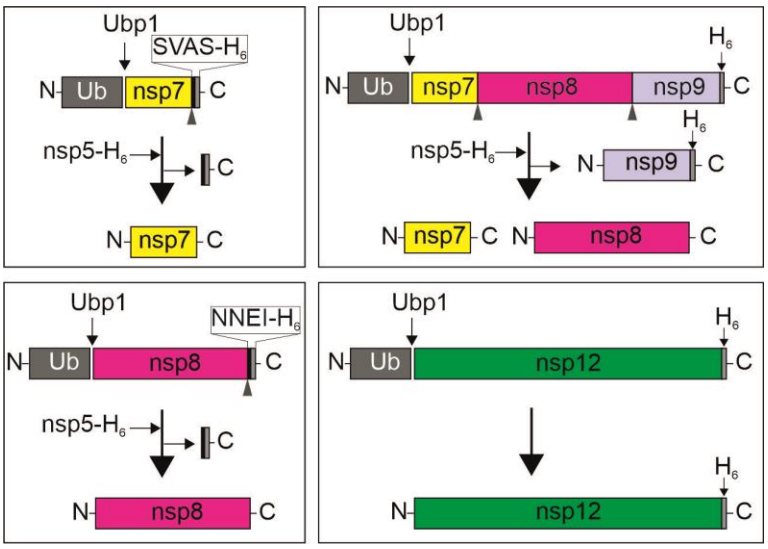


Figure 14: Experimental approach used to produce proteins with authentic N- and C-termini (nsp7, nsp8, nsp(7+8) complex) and nsp12-H_6 with its authentic N-terminus. Schematic representation of the production and purification of nsp7, nsp8, nsp(7+8) and nsp12-H_6. Nsp7-L-H_6, nsp8-L-H_6, nsp7-8-9-H_6 and nsp12-H_6 proteins were purified using Ni-IMAC and anion exchange chromatography. To liberate nsp7, nsp8 and nsp(7+8) with authentic N- and C-termini, the proteins were cleaved with 3C-like main protease (nsp5-H_6) and Ni-IMAC chromatography was employed to remove nsp9-H_6 and nsp5-H_6. Subsequently, nsp7, nsp8 and nsp(7+8) with authentic N- and C-termini were subjected to anion exchange chromatography and size-exclusion chromatography. The gray box indicates Ub; Ubp1 between Ub and nsp7 or nsp8 or nsp12 indicates that ubiquitin carboxyl-terminal hydrolase-1 cleaves specifically between Ub and the respective coronavirus nsp. Nsp7, nsp8, and nsp9 are indicated by yellow, pink and light blue boxes, respectovely. Gray arrowheads indicate cleavage sites processed by recombinant nsp5 (M^{pro}), the green box indicates nsp12. H_6 indicates the hexahistidine tag.

4.4 Expression and purification of nsp5-H_6

Coronavirus nsp5 ($3CL^{pro}$) coding sequences of HCoV-229E, FIPV and SARS-CoV, resepectively, with a C-terminal hexahistidine tag were cloned into the plasmid pMAL-c2X (pMAL-c2X-nsp5-His_6) to express an MBP-nsp5-H_6 fusion protein in *E. coli* (Hegyi et al., 2002; Ziebuhr et al., 1995; Ziebuhr et al., 1997). For HCoV-229E and FIPV nsp5, the according plasmid was transformed into *E. coli* TBI cells whereas *E. coli* Rosetta cells were used for SARS-CoV nsp5. To express HCoV-229E and FIPV nsp5, LB medium containing ampicillin (100 µg/ml) was inoculated with 5 ml of an overnight pre-culture. For SARS-CoV nsp5 expression, LB containing 100 µg/ml ampicillin and 34 µg/ml chloromaphenicol was inoculated with 5 ml of overnight culture. The cells were grown at 37°C until they reached an OD_{600} of 0.6. At this stage,

an aliquot of 1 ml was taken as a negative ("non-induced") control. The expression was induced by adding 0.3 mM (final concentration) IPTG for 16 h at 18°C and 1 ml was collected as positive control. The cells were harvested by centrifugation and Ni-IMAC purification was performed as described in Material and Methods. The eluted fractions containing the MBP-nsp5-H_6 fusion protein were analyzed by 12% Tris-tricine-PAGE along with non-induced and induced cell lysates (Figure 15). To release nsp5 with its authentic N-terminus from MBP, the fusion protein was incubated with factor Xa at 4°C for 2-3 days. Factor Xa recognizes the IEGR↓XXXX sequence located between MBP and nsp5-H_6. The cleaved MBP + nsp5-H_6 mixture was loaded onto a Ni-NTA matrix and nsp5-H_6 was purified as described in Materials and Methods. Due to efficient N-terminal autoprocessing by SARS-CoV nsp5-H_6 within its (slightly extended) N-terminal sequence including the authentic nsp4|nsp5 cleavage site, the factor Xa cleavage step was not required in this case.

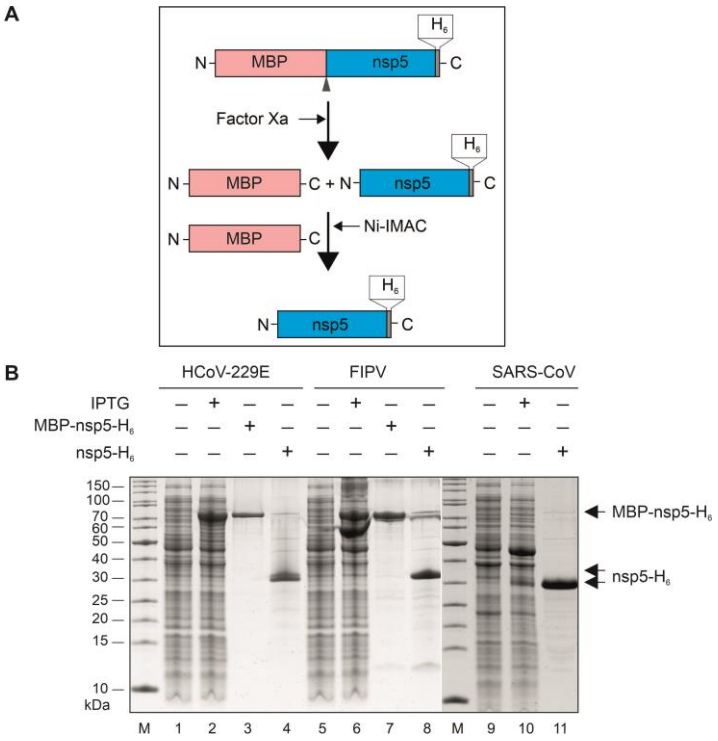


Figure 15: Tris-tricine-PAGE analysis of the expression and purification of HCoV-229E-, FIPV- and SARS-CoV- nsp5-H_6. (A) Schematic representation for the generation of nsp5-H_6. The arrowhead indicates the factor Xa cleavage site between MBP and nsp5-H_6. **(B)** Tris-tricine-

PAGE analysis of HCoV-229E, FIPV and SARS-CoV nsp5-H_6 purification by Ni-NTA affinity chromatography. HCoV-229E and FIPV MBP-nsp5-H_6 were expressed in TB1 transformed with pMAL-c2X-MBP-HCoV-229E-nsp5-H_6 and MAL-c2X-MBP-FIPV-nsp5-H_6, respectively. SARS-CoV MBP-nsp5-H_6 was expressend in *E. coli* Rosetta cells which had been transformed with pMAL-c2X-SARS-CoV-nsp5-H_6. HCoV-229E and FIPV MBP-nsp5-H_6 fusion proteins were purified by affinity chromatography using Ni-NTA. Thereafter, the fusion proteins were cleaved with factor Xa to separate MBP from nsp5-H_6. Subsequently, the mixture was loaded onto a Ni-NTA column. Bound nsp5-H_6 was eluted with imidazole whereas MBP was separated in the flow through. SARS-CoV-nsp5-H_6 can release itself from its N-terminal MBP tag upon induction due to having the C-terminal additional residues of nsp4 (P4, P3, P2, P1) at the N-terminus of nsp5. 5 µg of Ni-NTA affinity-purified nsp5-H_6 proteins were analyzed on a 12% Tris-tricine PAGE. Abbreviations: M – unstained protein marker; Lane 1, 5, 9 uninduced – the cells were not treated with IPTG; Lane 2 6, 10 induced – the cells were treated with IPTG; Lane 3, 7 MBP-nsp5-H_6 – maltose-binding protein-nsp5-H_6 fusion protein; Lane 4, 8, 11 nsp5-H_6 eluate fractions collected from the Ni-NTA column. The arrow indicates the purified MBP-nsp5-H_6 and nsp5-H_6 proteins with sizes that correspond to their calculated molecular weights of approximately 76 and 34 kDa, respectively.

4.5 Optimization of HCoV-229E nsp7-8-9-H_6 polyprotein maturation with HCoV-229E nsp5-H_6

The 43 kDa HCoV-229E nsp7-8-9-H_6 polyprotein was purified using the pASK-Ub-nsp-H_6 system as described in Materials and Methods. An *in vitro* proteolytic assay was conducted using 30 µg of nsp5-H_6 with 90 µg of nsp7-8-9-H_6 precursor substrate and incubated at 4°C in buffer containing 20 mM Tris-HCl/pH 8, 200 mM NaCl and 10 mM β-mercaptoethanol. The reaction samples were collected at the indicated time points and the percentage of cleavage was analyzed on a 12 % Tris-tricine-PAGE. The experiments showed that the substrate was cleaved rapidly by nsp5-H_6 into the expected nsp7, nsp8, and nsp9-H_6 processing products (Figure 16). All proteolytically processed nsps will have authentic N-termini, but while nsp7 and nsp8 also have authentic C-termini, nsp9 carries an additional hexahistidine tag at its C-terminus. Interestingly, a small cleavage product of approximately 12 kDa could as be observed besides the expected products. Most likely, this protein corresponds to a C-terminal nsp8 fragment as reported previously for another nsp8 protein construct expressed in *E. coli* (Li et al., 2010).

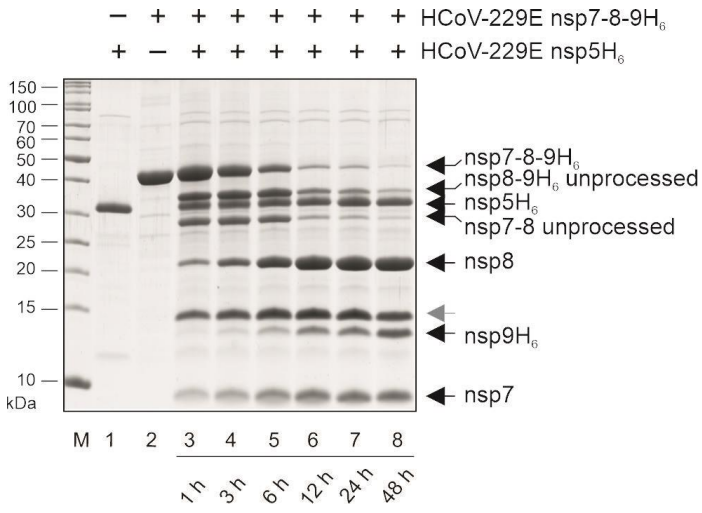


Figure 16: Tris-tricine-PAGE analysis of nsp5 protease-mediated proteolytic processing of HCoV-229E nsp7-8-9-H$_6$ polyprotein. To assess the release of mature nsp7, nsp8 and nsp9-H$_6$ from the nsp7-8-9-H$_6$ polyprotein precursor, 90 µg of precursor substrate was incubated with 30 µg of nsp5-H$_6$ at 4ºC in 20 mM Tris-Cl/pH 8, 200 mM NaCl, 10 mM β-mercaptoethanol. At the indicated time points, 12 µl of sample was collected and the percentage of cleavage determined on a Tris-tricine-PAGE with 5 µg of protease and 10 µg of uncleaved nsp7-8-9-H$_6$ loaded as controls. The protease-mediated cleavage products are marked on the right side with black arrows, the gray arrow indicates the (presumably aberrant) cleavage product of nsp8.

4.6 Generation of nsp7, nsp8 and pre-assembled nsp(7+8) complexes with authentic termini

The proteolytic assay was conducted in a large scale to generate tag-free HCoV-229E, FIPV and SARS-CoV nsp7, nsp8 and nsp(7+8) complexes. To produce nsp7 and nsp8 with authentic N- and C-termini and a preformed nsp(7+8) complex, nsp5-H$_6$, nsp7-L-H$_6$, nsp8-L-H$_6$ and nsp7-8-9-H$_6$ proteins were first purified by Ni-NTA affinity chromatography followed by anion exchange chromatography and dialyzed in dialysis buffer for further proteolytic processing. The purified nsp7-L-H$_6$, nsp8-L-H$_6$ and nsp7-8-9-H$_6$ precursor substrates were subjected to nsp5-H$_6$-mediated proteolysis for 48 hours at 4°C. Figure 17 shows the processing of HCoV-229E nsp7-L-H$_6$, nsp8-L-H$_6$ and nsp7-8-9-H$_6$ as an example. In the presence of the main protease, nsp7-L-H$_6$, nsp8-L-H$_6$ and nsp7-8-9-H$_6$ were cleaved into nsp7+L-H$_6$, nsp8+L-H$_6$ and nsp7+8+9-H$_6$, respectively. After completion of the cleavage, nsp5-H$_6$, L-H$_6$, nsp9-H$_6$ and

unprocessed His_6-containing protein was bound to Ni-NTA and tag free nsp7 and nsp8 and the nsp(7+8) complex were collected in the flow-through fraction. The same protocol was applied to generate tag-free FIPV and SARS-CoV nsp7, nsp8 and nsp(7+8) complexes. The tag-free proteins were further purified by anion exchange chromatography to remove nucleic acids and remaining bacterial proteins. Finally, nsp7, nsp8 and nsp(7+8) complexes were injected onto a prep grade size-exclusion column (Superdex 200), the eluted fractions combined and divided into two aliquots. One part was dialyzed into buffer containing 50 mM HEPES-KOH/pH 8, 150 mM KCl, 40 % [v/v] glycerol and 10 mM β-mercaptoethanl and stored at -20°C for further activity assays. The second part was dialyzed into 20 mM HEPES-KOH/pH 8, 150 mM KCl, 10 mM β-mercaptoethanol and used to analyze the multimerization state of the complexes by crosslinking and size-exclusion chromatography. Figure 17 shows the cleavage and purification of HCoV-229E nsp7, nsp8 and nsp(7+8). Lanes 1, 5, 9 are nsp5-H_6, lane 2, 6, 10 show the anion exchange chromatography-purified nsp7-L-H_6, nsp8-L-H_6 and nsp7-8-9-H_6, lane 3, 7 and 11 the protease-mediated processing of the respective protein; lanes 4, 8 and 12 show the final enriched nsp7 and nsp8 and nsp(7+8) complexes with authentic N- and C termini (Figure 17; lanes 4, 8 and 12).

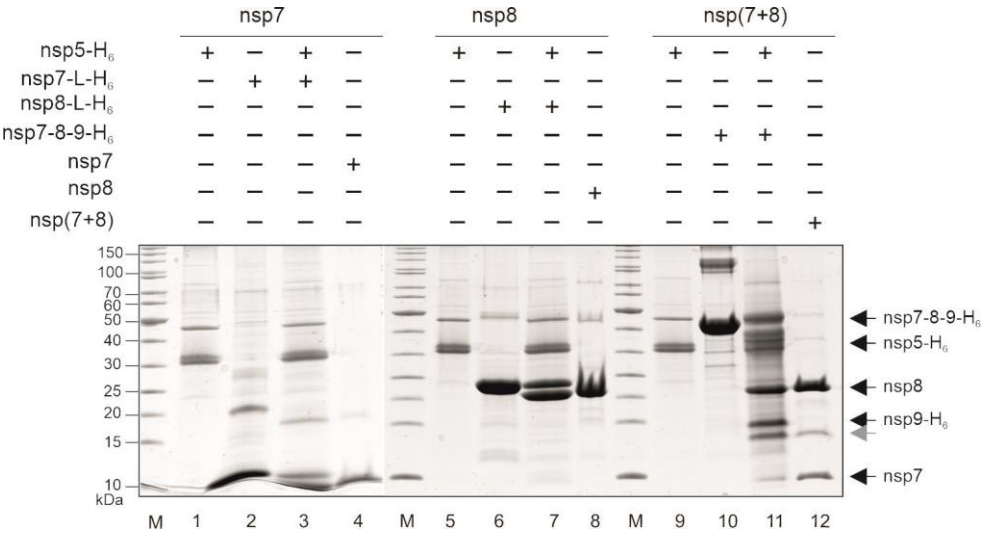


Figure 17: Tris-tricine-PAGE analysis illustrating the purification of HCoV-229E nsp7, nsp8 and pre-formed nsp(7+8) complexes with authentic N- and C-termini. Tris-tricine-PAGE showing the cleavage process used to generate tag-free versions of nsp7, nsp8 and nsp(7+8) complexes. Lanes 1, 2, 3, 4 show the generation of tag-free nsp7 (lane 1; 5 µg of purified nsp5-H_6, lane 2; 10 µg of purified unprocessed nsp7-L-H_6, Lane 3; 10 µg of nsp7-L-H_6 incubated with 5 µg of nsp5-H_6 to remove the L-H_6 tag from the C-terminus of nsp7, lane 4; 5 µg tag-free nsp7). Lanes 5, 6, 7, 8 show the generation of tag-free nsp8 (Lane 5; 5 µg of purified nsp5-H_6, Lane 6; 10 µg of purified unprocessed nsp8-L-H_6, lane 7; 10 µg of nsp8-L-H_6 incubated with 5 µg of nsp5-H_6 to remove the L-H_6 tag from nsp8, lane 8; 10 µg of tag-free nsp8). Lanes 9, 10, 11, 12 show the generation of a tag-free pre-assembled nsp(7+8) complex (lane 9; 5 µg of purified nsp5-H_6, lane 10; 10 µg of purified and unprocessed nsp7-8-9-H_6 polyprotein precursor substrate, lane 11; 10 µg of purified nsp7-8-9-H_6 polyprotein substrate treated with 5 µg of nsp5-His_6 protease to release the nsp7, nsp8 and nsp9-H_6 proteins, lane 12; 10 µg of pre-assembled nsp(7+8) tag-free complexes). Gray arrow heads indicate an aberrant cleavage product of nsp8. Abbreviations: M – unstained protein marker; +/- on top of the gel indicate the presence or absence of proteins; the positions of specific proteins are indicated to the right.

4.7 Multimeric states of HCoV-229E, FIPV and SARS-CoV nsp(7+8) complexes

4.7.1 BS^3 crosslinking

For BS^3 crosslinking, 5 µg of tag-free pre-assembled nsp(7+8) were used (Figure 18). Contradictory to the stable heterodimer protein complexes formed by nsp7-H_6 and nsp8-H_6, the pre-assembled complexes of HCoV-229E, FIPV and SARS-CoV also formed heterotrimeric and heterotetrametric complexes along with the previously abserved nsp(7+8) heterodimer, in addition to nsp7 and nsp8 monomers (Figure 18; lane 2 and 4), suggesting that the heterodimer is the initial "building block" of larger assemblies. Although the SARS-CoV nsp(7+8) complex showed very low amounts of the 1:2 ratio heterotrimeric complex. These multimeric species were not detected in the control reaction without BS^3. The stoichiometries of protein assemblies observed upon crosslinking were predicted based purely on their apparent molecular weight. Crosslinking results confirmed the formation of similar complexes for HCoV-229E, FIPV and SARS-CoV tag-free pre-assembled complexes despite these viruses belong to different coronavirus species or even genera.

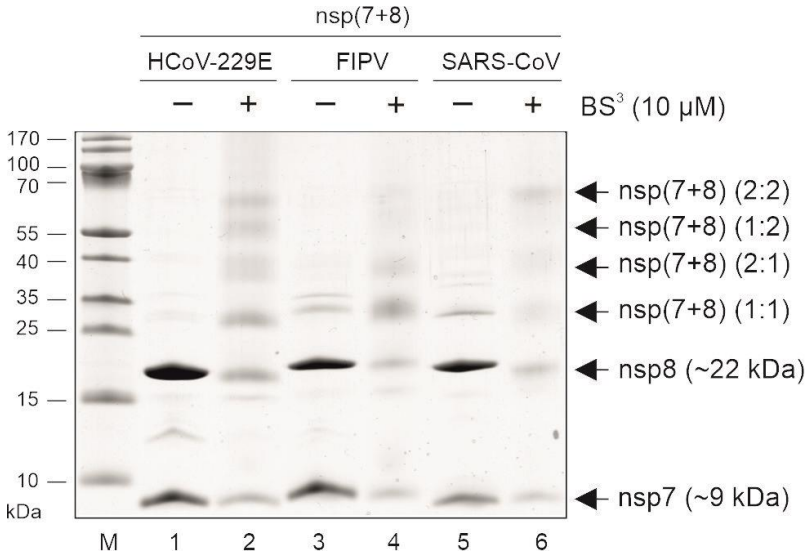


Figure 18: Tris-tricine-PAGE analysis of multimerization states of nsp(7+8) complexes of HCoV-229E, FIPV and SARS-CoV using BS^3 crosslinking. The nsp(7+8) complexes were cross-linked with BS^3 in reaction buffer (20 mM HEPES-KOH/pH 8, 30 mM KCl, and 2 mM β-mercaptoethanol) supplemented with 5 µg of protein and 10 µM BS^3. Cross-linking was carried out at 37°C for 30 minutes and quenched with 50 mM $(NH)_4HCO_3$ for another 30 minutes at 37°C. After terminating the reaction, the samples were mixed with 1x Laemmli 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 a 12 % Tris-tricine-PAGE. Lanes 1, 3, 5 – nsp(7+8) complexes without BS^3 treatment and lanes 2, 4, 6 – nsp(7+8) complexes treated with BS^3. Lane M, marker proteins (molecular masses in kilodaltons) indicated to the left. nsp – nonstructural protein. Black arrows on the right indicate various oligomeric states of the nsp(7+8) complexes upon cross-linking as predicted on the basis of their apparent molecular masses.

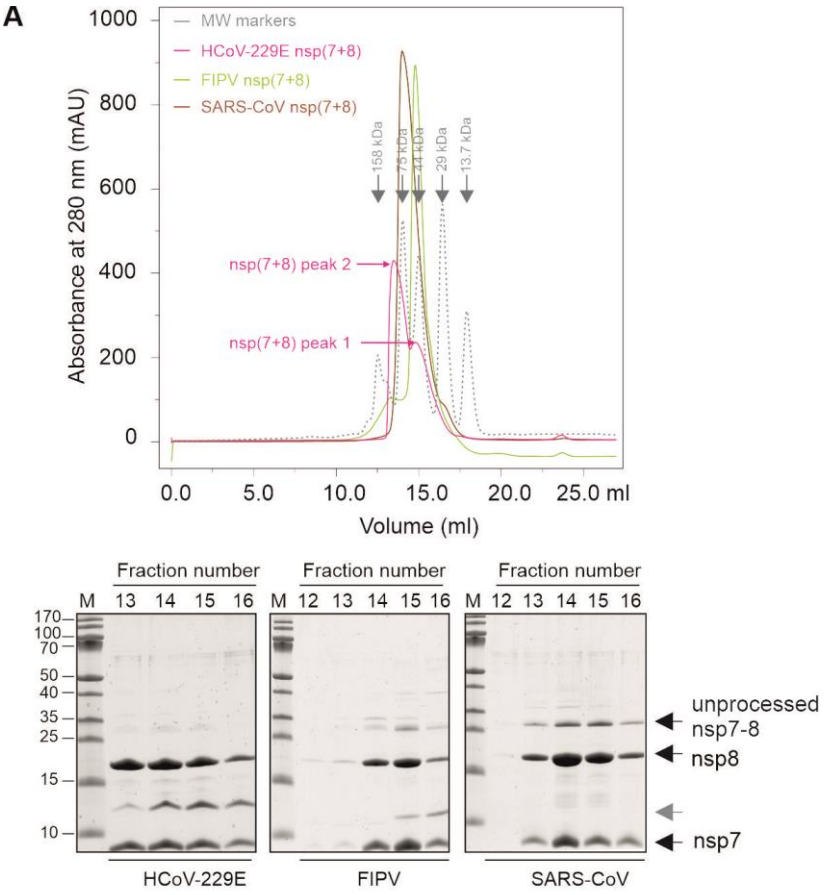
4.7.2 Analytical size-exclusion chromatography

The pre-assembled nsp(7+8) complexes (10 mg/ml each) were also subjected to analytical size-exclusion chromatography in buffer containing 20 mM HEPES-KOH/pH 8, 150 mM KCl and 10 mM β-mercaptoethanol. As shown in Figure 19, the elution profile of HCoV-229E nsp(7+8) showed two defined peaks with retention volumes (V_e) of 14.8 ml corresponding to an estimated molecular weight of 50.5 kDa (peak 1) and 13.49 ml, corresponding to an estimated molecular weight of 103 kDa (peak 2) (Figure 19). Peak 1 most probably represents a heterotrimer consisting of two copies of nsp7 and one copy of nsp8. Peak 2 may represent a hexameric complex consisting of three molecules nsp7 (3×9.3 kDa = 27.9 kDa) and three molecules nsp8

(3×22 kDa = 66 kDa). These results suggest that the hexameric complex has a limited stability and disassembles into smaller complexes under the conditions used.

In agreement with the reported FCoV heterotrimeric crystal structure, the FIPV pre-formed nsp(7+8) complex showed a major peak with a symmetrical profile and a V_e of 14.8 ml (corresponding to an estimated molecular weight of 51.5 kDa), indicating a heterotrimeric conformation (2:1) consisting of two molecules nsp7 and one molecule nsp8 (Figure 19). In addition to this major peak, some of the protein was found to elute earlier resulting in a relatively broad peak and indicating the formation of several larger (but probably unstable) protein complexes in some cases.

In contrast to previous studies for SARS-CoV (Zhai et al., 2005), the elution profile of the pre-assembled SARS-CoV nsp(7+8) complex revealed one major peak with a retention volume of 13.98 ml (Figure 19), corresponding to 75 kDa and suggesting the formation of stable heterotetramers consisting of two copies nsp7 and two copies nsp8. It should be noted that the estimated sizes of all nsp(7+8) complexes (as deduced from the elution profile of marker proteins) did not precisely match their calculated molecular masses (10-15 kDa above the expected sizes). The reason for this is not clear but could be due to the very extended (helical) conformation reported for the nsp8 N-terminal domain (Xiao et al., 2012; Zhai et al., 2005). To corroborate these size-exclusion chromatography data, native mass spectrometry analyses of nsp(7+8) complexes were performed in collaboration with Boris Krichel and Charlotte Uetrecht (Heinrich Pette Institute, Hamburg)



B

	HCoV-229E nsp(7+8)		FIPV nsp(7+8)	SARS-CoV nsp(7+8)
	peak 1	peak 2		
V_e	14.80	13.49	14.77	13.98
$K_{av}= V_e-V_o/V_c-V_o$	0.411	0.320	0.409	0.358
Estimated MW	50.5 kDa	103 kDa	51.5 kDa	75 kDa

Figure 19: Comparative size-exclusion chromatography analysis of pre-assembled nsp(7+8) complexes. (A) Size-exclusion chromatography elution profiles of the nsp(7+8) complexes. HCoV-229E nsp(7+8) complex: pink chromatogram, FIPV nsp(7+8) complex: green chromatogram, SARS-CoV nsp(7+8) complex: brown chromatogram. 10 mg/ml of each nsp(7+8) complex was loaded onto a Superdex 200 (10/300 GL) analytical size-exclusion column. HCoV-229E nsp(7+8) complex 1 eluted at 13.49 ml and complex 2 at 14.8 ml, while the FIPV nsp(7+8) complex eluted at 14.77 ml and SARS-CoV nsp(7+8) complex at 13.98 ml. The Y-axis represents the absorbance at 280 nm, the X-axis represents the elution volume in ml. Proteins of known molecular weights (RNase A: 13.7 kDa, carbonic anhydrase: 29 kDa, ovalbumin: 44 kDa, conalbumin: 75 kDa and aldolase: 158 kDa) were used as markers (gray chromatogram), and their elution peaks were used to calculate approximate molecular weights for fractions containing nsp(7+8) complexes. **(B)** The multimeric states of nsp(7+8) complexes were estimated based on their retention elution volume (V_e), void volume (V_o) and column volume (V_c) using the formula $K_{av}= V_e-V_o/V_c-V_o$. Based on this calculation, the nsp(7+8) complex 1 of HCoV-229E with its retention volume (13.49 ml) corresponded to a molecular mass of approximately 103 kDa, the nsp(7+8) complex 2 of HCoV-229E with its retention volume (14.80 ml) corresponded to a molecular mass of approximately 50.5 kDa, the nsp(7+8) complex of FIPV with its retention volume (15.27 ml) corresponded to an approximate molecular weight of 51.5 kDa while the SARS-CoV nsp(7+8) complex showed a retention volume of 13.98 ml, corresponding top a molecular mass of approximately 75 kDa.

4.7.3 Native mass spectrometry

Next, to further analyze the formation and stoichiometries of HCoV-229E, FIPV and SARS-CoV nsp(7+8) complexes, purified nsp7-8-9-H_6 precursor polyproteins were subjected to proteolytic processing with the respective nsp5-H_6 proteases and the resulting proteins were analyzed by native MS.

Native mass spectrometry results suggested that nsp5-H_6 and nsp9-H_6 did not interact with any of the other subunits (Figure 20). In line with previous structure analyses (Anand et al., 2002; Anand et al., 2003; Egloff et al., 2004; Sutton et al., 2004) both nsp5-H_6 and nsp9-H_6 were found to form homodimers (Figure 20). Nsp7 that was not part of complexes with nsp8 also formed a homodimer. The HCoV-229E nsp(7+8) complex was found to form heterodimers (1:1) and heterotetrametric (2:2) complexes of 61.89 kDa (Figure 20 A). Similar to the HCoV-229E nsp(7+8) complex, SARS-CoV nsp(7+8) complex formed both heterodimeric (1:1) and heterotetrametric (2:2) complexes. The approximate size of the heterotetrametric complex was 62.38 kDa (Figure 20 B). In contrast to an earlier study (Zhai et al., 2005), there was no evidence for the formation of hexadecameric complexes. However, the native mass spectrometry data are consistent with the recent description of SARS-CoV nsp(7+8) tetramer formation following proteolytic cleavage of a nsp7-8-9-10 precursor substrate (Krichel et al., 2020). In agreement with the FCoV heterotrimeric co-crystal structure reported earlier (Xiao et al., 2012), our results also showed the formation of a heterotrimeric complex for FIPV nsp(7+8) with a size of 40.37 kDa confirming the presence of two molecules nsp7 and one molecule nsp8 in this complex (Figure 20 C). These stoichiometries are predicted based on the molecular weights of the subunits (2×9.3 kDa + 2×21.8 kDa = 62.2 kDa). However, to provide further evidence, molecular ions of three different charge states of the nsp7+8 hetero-tetramer of HCoV-229E and SARS-

CoV and heterotrimers of FIPV nsp(7+8) complexes were selected as precursor and subjected to CID-MS/MS to successively induce subunit dissociation by increasing the collision voltage stepwise (data not shown).

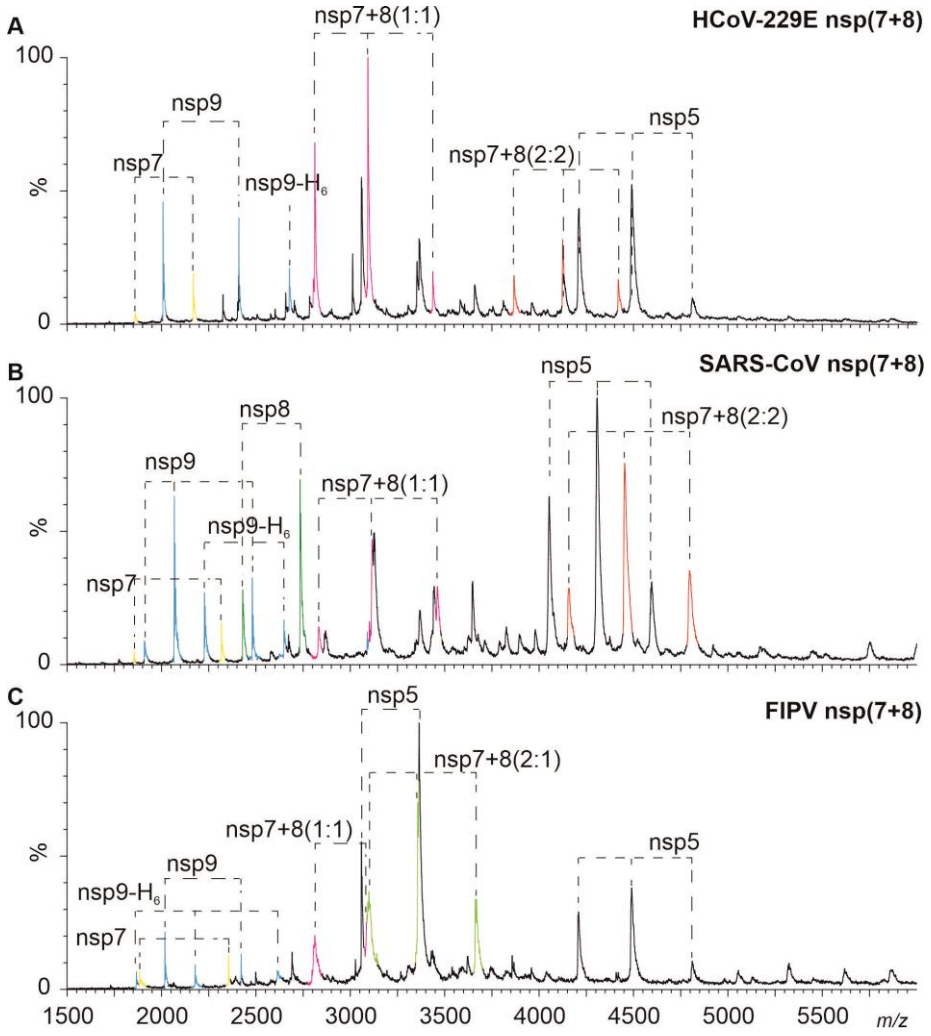


Figure 20: Identification of the stoichiometries of HCoV-229E, SARS-CoV and FIPV nsp(7+8) complexes using native mass spectroscopy. Exemplary native mass spectra illustrate the landscapes of covalent and non-covalent processing products. Coronavirus nsp7-8-9-H_6 proteins were cleaved with recombinant 3C-like protease (nsp5) at 4°C. Samples were injected into an electrospray capillary and native MS spectra were acquired at the indicated time points. **(A)** Mass spectrum of HCoV-229E nsp7-8-9-H_6 and HCoV-229E $3CL^{pro}$ (ratio ~1:6, 18 µM, after 20 h incubation time) in 500 mM AmAc, 1 mM DTT, pH 8) showing a dominant signal for non-covalent nsp(7+8) heterodimers and lower intense species of nsp(7+8) heterotetramers **(B)** Mass spectrum of SARS-CoV nsp7-8-9-H_6 and SARS-CoV $3CL^{pro}$ (ratio ~1:10, 17 µM, 40 h in 300 mM AmAc, 1 mM DTT, pH 8) showing a dominant signal for non-covalent nsp(7+8) hetero-tetramer and lower intense species of nsp(7+8) hetero-dimer. **(C)** Mass spectrum of FIPV nsp7-8-9-H_6 and FIPV $3CL^{pro}$ (ratio ~1:8.5, 10 µM, 40 h in 300 mM AmAc, 1 mM DTT, pH 8) showing a dominant signal for non-covalent nsp(7+8) heterotrimer and low intense species of nsp(7+8) heterodimer. The bimodal mass distribution of nsp8 does not influence complex formation: Mass-to-charge (m/z) are labeled, different mass species that were assigned are colored. HCoV-229E, FIPV and SARS-CoV 1:1 heterodimeric mass species were labeled in pink, HCoV-229E and SARS-CoV 2:2 heterotetrameric mass species were labeled red, FIPV

nsp(7+8) 2:1 heterotetrameric mass species were labeled chartreuse, nsp5-H_6 mass was labeled in gray, nsp7 mass was labeled in yellow, nsp8 mass was labeled in green, nsp9 mass labeled in sky blue and nsp9, nsp9-H_6 masses were labeled in cyan.

4.7.4 Expression and purification of HCoV-229E, FIPV and SARS-CoV nsp12

To determine the roles of pre-assembled nsp(7+8) complexes, nsp7 and nsp8 in coronavirus RNA synthesis in vitro, the 106-kDa nsp12 was also expressed and purified using the Ubiquitin-nsp-fusion system. The appropriate pASK-Ub-nsp12-H_6 plasmid was transformed into *E. coli* TB1 [pCGI] cells. Due to the low yield of nsp12, large culture volumes of up to 10 liters were used. The expression and purification were performed as described in Material and Methods. Along with nsp12, inactive (mutant) forms of the protein were expressed and purified as negative controls: nsp12_SAA of HCoV-229E (D4828A/D4829A) and FIPV (D4780A/D4781A) (Figure 21). For SARS-CoV, we mutated the codon for D618 (corresponding to D4987 in the polyprotein 1ab numbering) within motif A as a negative control (Figure 21 B). There was no major difference in the yield, stability and purity of the different coronavirus nsp12 proteins produced in *E. coli* in this study. Figure 21 shows the purified tag-free nsp7, nsp8 and nsp(7+8) and nsp12-H_6 proteins that were used in subsequent polymerase activity assays.

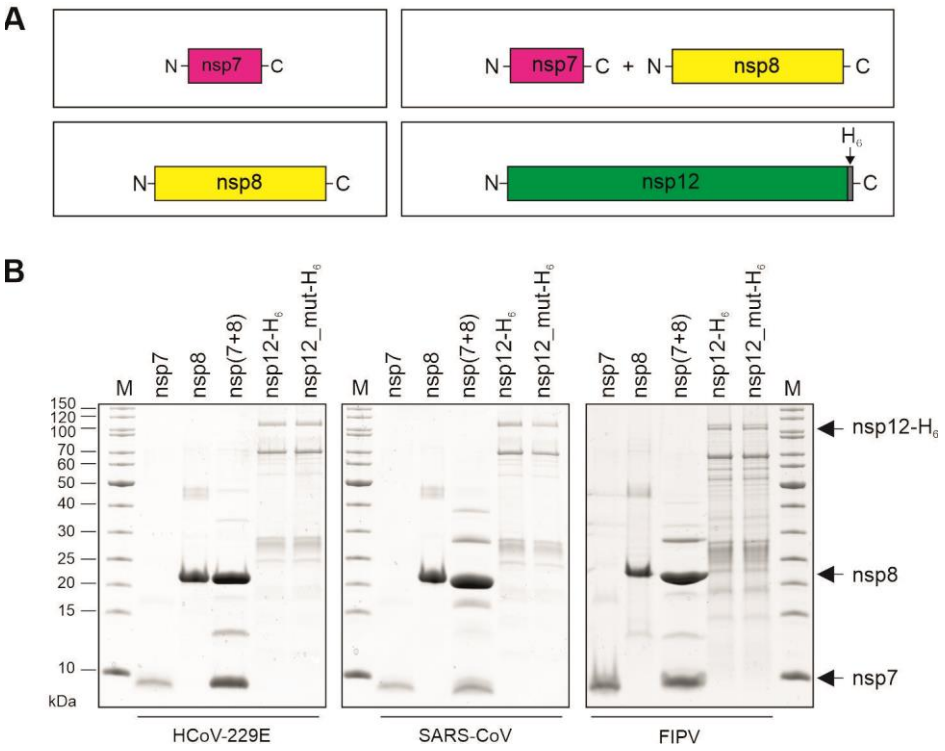


Figure 21: Nsp7, nsp8, nsp(7+8) complexes and nsp12-His$_6$ of HCoV-229E, FIPV and SARS-CoV produced in this study. (A) Schematic representation of the purified nsp7, nsp8, nsp(7+8) proteins carrying authentic N- and C-termini and nsp12-H$_6$, which has its authentic N-terminus but carries a C-terminal hexahistidine tag. **(B)** 12 % SDS-polyacrylamide gel showing the purified nsp7, nsp8, nsp(7+8) and nsp12-His$_6$ proteins of HCoV-229E, SARS-CoV and FIPV, respectively. Nsp12_mut - nsp12 variants containing alanine substitutions of two conserved aspartic acid residues. N - NH_2-terminus; His$_6$ - hexahistidine tag, C - carboxyl-terminus; Lane M - marker proteins (molecular masses [in kilodaltons] are indicated to the left).

4.8 Biochemical characterization of coronavirus nsp7, nsp8, nsp(7+8) complex and nsp12

4.8.1 The pre-assembled nsp(7+8) complex and nsp12 are essential for efficient RNA synthesis

To verify the role of nsp7, nsp8, the nsp(7+8) complex and nsp12-H$_6$ in RNA synthesis, an *in vitro* polymerase activity was carried out using double-stranded RNA substrates. We used a 28-oligonucleotide primer sequence mimicking the 3'-end of the HCoV-229E genome RNA which was annealed with the corresponding complementary RNA oligonucleotid containing a 10-nucleotide $(CU)_5$ overhang at the 5'-end (Figure 22 A). The 5'- heteropolymeric $(CU)_5$ overhang served as template for the respective enzymes.

Based on earlier reports describing the requirement of metal ions for nsp12 activity, we supplemented the polymerase assays with Mg^{2+} or Mn^{2+} ions to determine the divalent metal ion requirements for coronavirus polymerase activity (Gordon et al., 2020; Imbert et al., 2006; Subissi et al., 2014; te Velthuis et al., 2010a; te Velthuis et al., 2012; Tvarogova et al., 2019; Xiao et al., 2012). Newly synthesized RNA was detected by the incorporation of [α-^{32}P]-GTP in the presence of ATP and GTP. As negative control, the protein was omitted in the reactions (Figure 22; lane 1 and 11). As shown in Figure 22, the use of nsp7, nsp8, nsp7+nsp8 and the nsp(7+8) complex did not result in a radiolabeled product in the presence of either Mg^{2+} or Mn^{2+} (Figure 22; lanes 2, 3, 4, 5, 12, 13, 14, 15). A minimal (hardly detectable) polymerase activity was shown for nsp12 alone and nsp7+nsp12 in the presence of Mg^{2+} or Mn^{2+} (Figure 22; lanes 6, 7, 16 and 17). In contrast, a combination of nsp8+nsp12 and a mixture of nsp7+nsp8+nsp12 showed a slightly more efficient polymerase activity in the presence of Mg^{2+} (Figure 22; lanes 8 and 12). The nsp7+nsp8+nsp12 mixture had less activity in the presence of Mn^{2+} ions (Figure 22; lane 19). A mixture of nsp8+nsp12 displayed robust polymerase activity in the presence of Mn^{2+} (Figure 22; lane 18). The by far most efficient RNA synthesis was observed if pre-assembled nsp(7+8)+nsp12 complexes were used in the reaction (Figure 22 B; lanes 10 and 20). The polymerase activity was somewhat higher in the presence of Mn^{2+} compared to Mg^{2+} ions. Interestingly, the length of the products was slightly increased if (in addition to nsp8 and nsp12) nsp7 was included in the reaction, suggesting that the nsp(7+8) complex helps to keep the nsp12 polymerase bound to the template strand until the copy process is completed. Overall, the data reveal a critical role for nsp8 in coronavirus nsp12-mediated RNA synthesis and show that a nsp(7+8) complex that is (pre)assembled by interactions prior to or during M^{pro}-

mediated proteolytic processing of pp1a/pp1ab supports RNA synthesis much more efficiently than a mixture of (individually produced) nsp7 and nsp8 proteins.

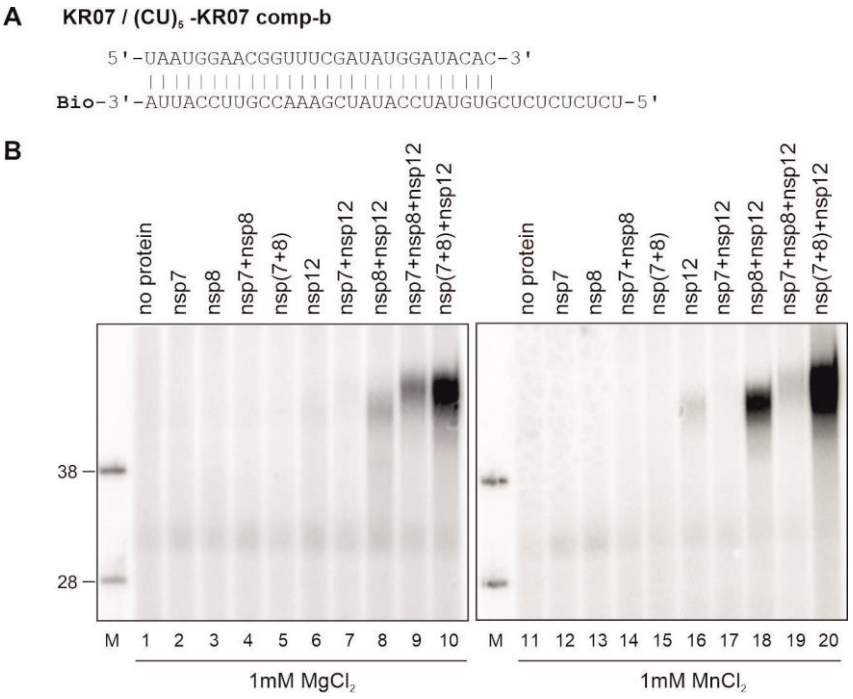


Figure 22: Pre-formed HCoV-229E nsp(7+8) complex efficiently supports its cognate nsp12 to elongate double-stranded RNA with a 10-nt 5' overhang: (A) RNA substrate used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** Template-dependent primer extension was performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM protein, 1 µM RNA substrate, 100 µM ATP, 100 µM GTP and 0.17 µM [α-^{32}P] GTP. Reactions in lanes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 were supplemented with 1 mM $MgCl_2$ while the reactions in lanes 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 were supplemented with 1 mM $MnCl_2$. In lanes 1 and 11, no proteins were added to the reactions. Nsp7 (lanes 2 and 12), nsp8 (lanes 3, 13), nsp 7 and nsp8 (lanes 4 and 14), nsp12 (lanes 6 and 16), nsp7 and nsp12 (lanes 7 and 17), nsp8 and nsp12 (lanes 8 and 18), nsp7, nsp8 and nsp12 (lanes 9 and 19), nsp(7+8) and nsp12 (lanes 10 and 20) were used in different reactions as indicated above the autoradiograms. In protein mixtures, proteins were used at a 1:1(:1) molar ratio. Products were resolved in a TBE-buffered 12 % polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, [5'-^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left.

4.8.2 The nsp(7+8)+nsp12 complex is capable to incorporate all NTPs during RNA elongation with both Mg^{2+} and Mn^{2+}

As already mentioned, a previous report suggested that SARS-CoV nsp8 acts as a noncanonical RdRp capable of synthesizing short RNA primers by a template-dependent *de novo* mechanism (Imbert et al., 2006). Te Velthuis *et al.* proposed that SARS-CoV nsp8 alone and as an nsp7+8 complex can extend an RNA substrate

containing a stretch of 20 C/U nucleotides in the template strand if the complementary nucleotides are provided (te Velthuis et al., 2012). Xiao *et al.* proposed a primer-independent polymerase activity for FCoV nsp8 (if part of a complex with nsp7) (Xiao et al., 2012). In a subsequent study by Subissi *et al.*, the Imbert laboratory reported that neither nsp7L8 (L = hexahistidine linker) nor nsp12 has any polymerase activity (Subissi et al., 2014), thereby partly contradicting conclusions made by this and other laboratories.

To validate and help resolve these controversial observations, possible functions of nsp7, nsp8, nsp(7+8) complexes and nsp12 were investigated in more detail in the present study. First, primer-dependent polymerase activity was investigated by providing various partially double-stranded RNA substrates in the presence of Mg^{2+} or Mn^{2+} (Figure 23 A). The RNA substrates consisted of the 28-nt KR07 RNA annealed to the corresponding complementary RNA containing a 5'-U_{10}, -A_{10}, -C_{10} or -$(CU)_5$ overhang. These 5'-overhangs served as template strands in primer-dependent polymerase activity assays. The experiments were performed with nsp7, nsp8, the nsp(7+8) complex, nsp12 and combinations of nsp7+nsp12, nsp8+nsp12, nsp(7+8)+nsp12, respectively. Reactions containing no enzyme were used as negative control (Figure 23 B, C, D, E: lanes 1, 9, 17, 25).

To assess AMP incorporation by these proteins, the KR07/U_{10}-KR07 comp-b partially double-stranded RNA substrate with a 5'-U_{10} overhang was used in the presence of ATP and [α-^{32}P]-ATP as nucleotide cosubstrate. The reactions contained nsp7, nsp8, nsp(7+8) or nsp12 or combinations of the nsp7 and nsp8 proteins with nsp12. In all reactions containing nsp8, radiolabeled products could be observed, albeit with different intensity (Figure 23 B; lanes 3, 4, 7, 8, 11, 12, 15, 16). In the presence of nsp12 or the nsp12_SAA mutant, very little radiolabeled product was observed if nsp8 was omitted

from the reaction (Figure 23 B, lanes 5, 6, 13, 14). The data suggest that the nsp8-mediated 3'-terminal adenylyltransferase activity reported previously for nsp8 (Tvarogova et al., 2019) was largely responsible for the production of radiolabeled products in these reactions. More importantly, a major (single) product of the expected size was only produced if the pre-assembled nsp(7+8) complex was used in combination with nsp12 but not in combination with the nsp12 active-site mutant protein (Figure 23 B; lanes 8 and 16). Similar results were obtained in reactions in which Mg^{2+} ions were replaced with Mn^{2+} ions (Figure 23 B; lanes17-32), with a few notable exceptions. Thus, in contrast to the results obtained in the presence of Mg^{2+}, nsp8 and nsp12 (in the absence of nsp7) generated a prominent product of (nearly) the expected length in the presence of Mn^{2+} (Figure 23 B, lane 23). This product was slightly shorter than the product produced by a mixture of the nsp(7+8) complex with nsp12, consistent with the data shown above (Figure 22 B). Taken together, the data show that nsp12 efficiently produces the complement of an oligouridylate template strand if a preformed nsp(7+8) complex is included in the reaction. This RdRp activity of nsp12 overlaps with the terminal adenylyltransferase activity mediated by nsp8 (Tvarogova et al., 2019). The data also suggest that Mn^{2+} ions stimulate the nsp12-mediated RdRp activity in reactions in which nsp7 was omitted. Interestingly, however, the absence of nsp7 leads to a slightly shorter product.

To test whether the enzymes or enzyme complexes can also incorporate UMP, the KR07/ A_{10} KR07 comp-b partially double-stranded RNA containing a 5'- A_{10} overhang was employed (Figure 23 A; substrate 2). Only the nsp(7+8)+nsp12 complex was found to efficiently incorporate UMP in the presence of Mg^{2+} (Figure 23 C; lane 8). All other proteins and protein combinations were found to be inactive (Figure 23 C; lanes 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15). Nsp12_SAA alone or in combination with other proteins

had no detectable activity. However, in the presence of Mn^{2+}, minor amounts of radiolabeled products were found in reactions containing nsp8, nsp(7+8), nsp12 and nsp7+nsp12 as well as by the nsp12_mut and nsp7+nsp12_mut (Figure 23 C; lanes 19, 20, 21, 22, 27, 28, 29, 30, 31). While Mn^{2+} supported the formation of a product of (nearly) the expected length by the combination of nsp8 and nsp12 (Figure 23 C; lane 23), a slightly larger product was generated in the reaction containing the nsp(7+8) complex and nsp12 (Figure 23 C; lane 24).

Next, to assess GMP incorporation by nsp7, nsp8, nsp(7+8) and nsp12, a synthetic partially double-stranded RNA substrate containing a homopolymeric (C_{10}) 5'-overhang was used (Figure 23 A; substrate 3). The assays were performed in the presence of Mg^{2+} or Mn^{2+}. The reactions with Mg^{2+} showed minimal amounts of GMP incorporation by nsp12, nsp7+nsp12 and nsp8+nsp12 (Figure 23 D, lanes 5, 6, 7) whereas nsp12 in combination with the nsp(7+8) complex incorporated GMP efficiently into products in the presence of Mg^{2+} (Fig. 23 C, lane 8). The product migrated more slowly than expected in the urea-polyacrylamide gel, possibly due to formation of a stable G-quadruplex structure preventing complete strand separation of the product RNA (regions) under the conditions used. In the presence of Mn^{2+}, a major labeled product was also observed if the reaction contained nsp8+nsp12 while slightly less product was produced by nsp(7+8)+nsp12 (Figure 23, lanes 23 and 24). The nsp12_SAA mutant was found to be inactive with both Mg^{2+} and Mn^{2+} (Figure 23 D; lanes 13-16 and 29-32).

Finally, the assays were performed with the KR07/ $(CU)_5$ KR07 comp-b partially double-stranded RNA containing a 5'-$(CU)_5$ overhang (Figure 23 A; substrate 4). Reactions were performed in the presence of Mg^{2+} or Mn^{2+} and ATP and GTP. Nsp12, nsp7+nsp12 and nsp8+nsp12 catalyzed only a minor amount of product formation (Figure 23 E; lanes 5-8), while a large amount of product was detected in the presence

of nsp(7+8)+nsp12 in the presence of Mg^{2+} ions (Figure 23 E; lane 8). In the presence of Mn^{2+} instead of $Mg^{2+,}$ small amounts of radiolabeled products were observed in the presence of nsp12 and nsp7+nsp12 (Figure 23 E; lanes 21 and 22). Consistent with the data shown in Figure 23 B, C, and D, the presence of Mn^{2+} stimulated the synthesis of a major product in reactions containing nsp8 and nsp12 (but no nsp7). This product was of slightly smaller size compared to the product synthesized in reactions containing nsp(7+8) and nsp12, suggesting that the nsp7 protein (as part of the nsp(7+8) complex) facilitated the 5'-to-3' extension of the RNA product to the very 5'-end of the template strand. (Figure 23 E; lanes 23 and 24). As expected, the nsp12_SAA mutant had no detectable activity in the presence of Mg^{2+} or Mn^{2+} (Figure 23 E; lanes 13-16 and 29-32).

Taken together, the data revealed that a pre-assembled nsp(7+8) complex profoundly supports nsp12 polymerase activity, resulting in efficient elongation of primed RNA substrates with different homo- or heteropolymeric template sequences and provision with matching nucleotides (Figure 23 B, C, D, E). In striking contrast to earlier observations made by Subissi *et al.* for SARS-CoV nsp7L8 and nsp12, we were able to detect terminal adenylyltransferase activity in the presence of the KR07/ U_{10} KR07 comp-b substrate using nsp8, nsp(7+8), and nsp12 in the presence of Mg^{2+} or Mn^{2+} ions (Figure 23 B) (Subissi et al., 2014). Furthermore, minor primer-dependent polymerase activity was observed for nsp12, nsp7+nsp12 and nsp8+nsp12 on cytidylate-containing templates in the presence of Mg^{2+}. Also, the data suggest that Mn^{2+} affects the nsp12 polymerase activity in a way that makes it less dependent of the presence of the preformed nsp(7+8) complex.

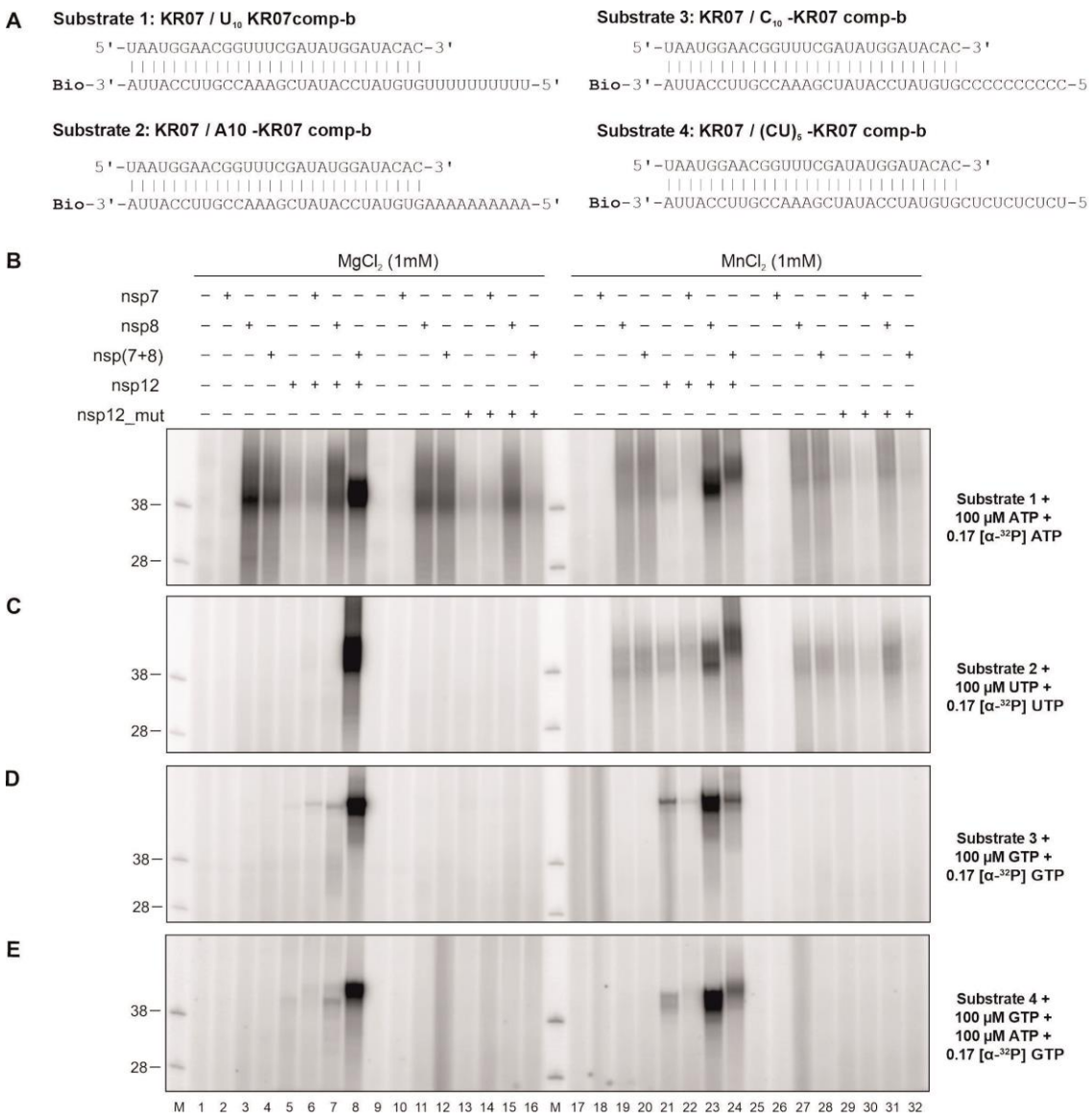


Figure 23: Polymerase activity of the HCoV-229E replicase complex (nsp(7+8)+nsp12) on partially double-stranded RNA substrates. (A) Partially double-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** Primer elongation assay was performed with RNA substrate 1 **(C)** Primer elongation assay was performed with RNA substrate 2 **(D)** Primer elongation assay was performed with RNA substrate 3 **(E)** Primer elongation assay was performed with RNA substrate 4. Primer-dependent polymerase assays were performed at 30°C for 1 h in standard reaction buffer supplemented with 1 µM of RNA substrate and 100 µM nucleotides as indicated to the right. The proteins used in the assays are indicated above the autoradiogram. Lanes 6, 14, 22 and 30: nsp7 and nsp12 or nsp12_mut as indicated; lanes 7, 15, 23 and 31: nsp8 and nsp12 or nsp12_mut as indicated; lanes 8, 16, 24 and 32: nsp(7+8)+nsp12 or nsp12_mut as indicated. Proteins were used at a 1:1(:1) molar ratio. nsp12_mut: nsp12 variant containing alanine substitutions of conserved active-site aspartic acid residues D4828 and D4829 (numbers indicate amino acid positions in the HCoV-229E polyprotein 1ab). + or -: reactions performed in the presence or absence of the respective protein. Products were resolved in a TBE-buffered 12 % polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of the autoradiograms.

4.8.3 Nsp(7+8)+nsp12 and Mn^{2+} are required for *de novo* initiation of RNA synthesis

An earlier report proposed *de novo* RNA synthesis for SARS-CoV nsp7L8+nsp12 using RNAs containing the 3'-UTR and fused 5'- and 3'-UTR in the presence of Mg^{2+} (Subissi et al., 2014) while Ahn *et al*. proposed that Mn^{2+} is essential for *de novo* RNA synthesis by nsp12 using both + and – stranded RNA (Ahn et al., 2012). Yet another report proposed primer-independent polymerase activity (*de novo*) for FCoV nsp8 and in complex with nsp7 in the presence of Mg^{2+} (Xiao et al., 2012).

To extend and evaluate these (partially contradictory) studies, polymerase assays were conducted using recombinant HCoV-229E nsp7, nsp8, nsp(7+8), nsp12 proteins and combinations of these proteins including nsp7+nsp12, nsp8+nsp12, and nsp(7+8)+nsp12. The 28-nt ssRNA KR07-b was used as a substrate (Figure 24 A; substrate 1). In this synthetic RNA, the 3'-OH group was modified with biotin to avoid any 3'-terminal nucleotidylation by the proteins used. To initiate RNA synthesis, we included Mg^{2+} or Mn^{2+} and all 4 NTPs (ATP, GTP, CTP, and UTP) in the reaction mixtures and products were detected by the incorporation of radiolabeled [α-^{32}P]-UTP (Figure 24 B). In the presence of Mg^{2+}, no products were synthesized by nsp7, nsp8, nsp(7+8), nsp12 or combinations of nsp7+nsp12, nsp8+nsp12, nsp(7+8)+nsp12, nsp7+nsp12_SAA, nsp8+nsp12_SAA or nsp(7+8)+nsp12_SAA (Figure 24 B; lanes 1-16). If Mg^{2+} was replaced with Mn^{2+} under otherwise identical reaction conditions, a major radiolabeled product was (only) detected in the presence of nsp(7+8) and nsp12 (Figure 24 B; lane 24). The product size was comparable to that of the template RNA used in this reaction, suggesting that (*de novo*) RNA synthesis had been initiated at (or very close to) the 3'-terminus of the template RNA and was terminated when the polymerase complex had reached the 5'-terminus of the template strand. Mn^{2+} seems to

be essential for primer-independent (i.e., *de novo* initiation of) RNA synthesis, since no product was detectable if Mg^{2+} was used in the reaction. To verify that the RNA was generated by nsp(7+8)+nsp12 rather than a potentially contaminating *E. coli* polymerase, the mutant nsp12_SAA protein was used in a control reaction. In this case, RNA synthesis was completely abolished, corroborating the conclusion that the synthesis of the product RNA was mediated by nsp12.

To obtain more insight into specific substrate preferences or requirements, polymerase assays were performed with the synthetic 38-nt KR07 A_{10} -b RNA substrate (Figure 24 A; substrate 2). The reason for using this particular RNA oligonucleotide was that, for several viruses, it has been shown that a poly(A) tail at the 3'-end of the genome plays a regulatory role in the initiation of minus-strand RNA synthesis (Hsue and Masters, 1997; Madhugiri et al., 2014; Svitkin et al., 2007; Züst et al., 2008). Additionally, we sought to determine whether the enzyme complex initiates RNA synthesis at (or within) the adenylylated 3'- terminus by incorporating UMP as the initial nucleotide or at the cytidylate immediately upstream of the poly(A) tail. Again, the assays with Mg^{2+} did not support RNA synthesis (Figure 24 C; 1-16) using KR07 A_{10}-b, whereas Mn^{2+} supported the initiation of *de novo* RNA synthesis if nsp(7+8) and nsp12 were included in the reaction (Figure 24 C; lane 24). Compared to the product(s) observed in the experiment shown in Figure 24 B, additional products of slightly larger size were detectable, suggesting that RNA synthesis (i) was initiated at different positions close to the 3’-end of the template strand and (ii) terminated when the enzyme complex reached the 5'-terminus of the template. Again, the use of a mutant form of nsp12 abolished RNA synthesis, confirming that the activity was mediated by nsp12. Collectively, these *in vitro* results suggest that the nsp(7+8)+nsp12 complex and Mn^{2+} are the essential components for initiation of RNA synthesis which stands in contrast to

earlier reports (Ahn et al., 2012; Subissi et al., 2014; te Velthuis et al., 2010a; te Velthuis et al., 2012; Xiao et al., 2012).

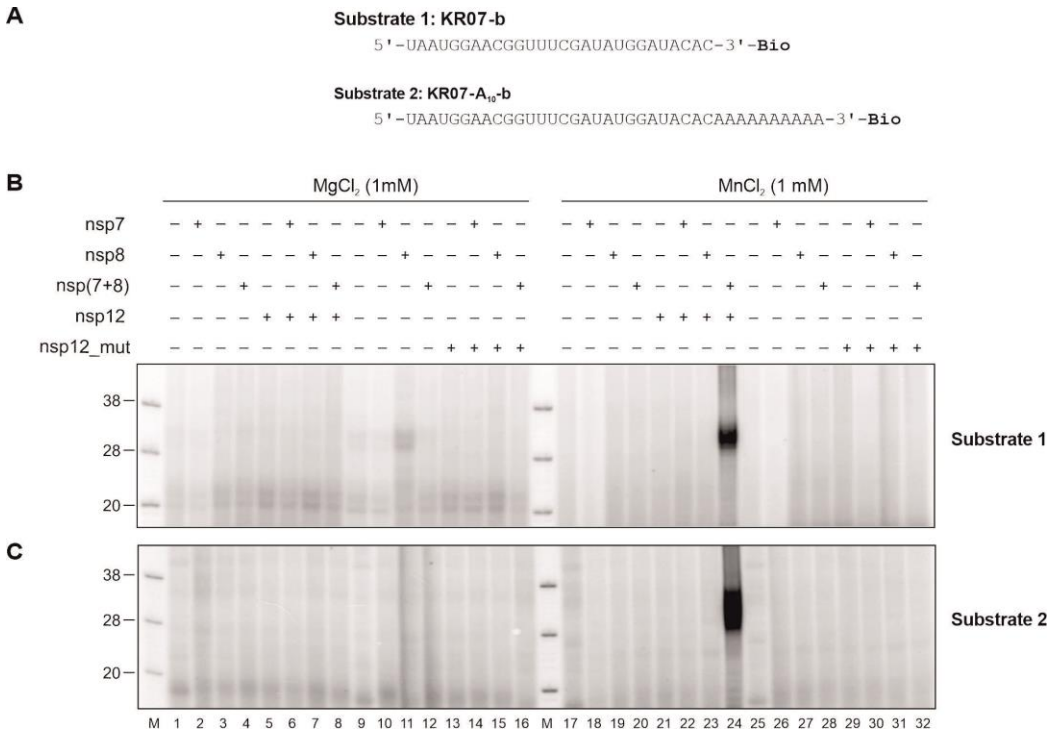


Figure 24: HCoV-229E replicase complex (nsp(7+8)+nsp12) initiates *in vitro* RNA synthesis by a *de novo* (primer-independent) mechanism. (A) Single-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** *De novo* initiation assay was performed with RNA substrate 1. **(C)** *De novo* initiation assay was performed with RNA substrate 2. *De novo* initiation assays were performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of protein, 1 µM of RNA substrate, NTP mix (100 µM of ATP, CTP, GTP and 75 µM UTP) and 0.17 µM [α-^{32}P]-UTP in the presence of $MgCl_2$ or $MnCl_2$ as indicated above the autoradiograms. The RNA substrate used in the respective reactions is indicated to the right. Lanes 6, 14, 22, and 30: nsp7 and nsp12 or nsp12_mut as indicated; lanes 7, 15, 23, and 31: nsp8 and nsp12 or nsp12_mut as indicated; lanes 8, 16, 24, and 32: nsp(7+8)+nsp12 or nsp12_mut as indicated. Proteins were used at a 1:1(:1) molar ratio. nsp12_mut: nsp12 variant containing alanine substitutions of conserved aspartic acid residues D4828 and D4829 (numbers indicate positions in the polyprotein 1ab of HCoV-229E). + or -: reactions were performed in the presence or absence of the indicated protein. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

4.8.4 Comparative analysis of HCoV-229E, SARS-CoV and FIPV nsp(7+8)+nsp12 replicase complexes *de novo* initiation and elongation of RNA synthesis

To assess whether the HCoV-229E, FIPV and SARS-CoV nsp(7+8)+nsp12 complexes share similarities in their activities or differ from each other due to non-identical stoichiometries of their nsp7+8 architectures, the *de novo* initiation and elongation activities were investigated for the respective proteins. The three *in vitro* constituted coronavirus replicase complexes were incubated with synthetic oligonucleotide ssRNAs whose sequences were derived from the 3'-end of the HCoV-229E genome (Figure 25 A: substrate 1-4). Comparative analysis of the activities of HCoV-229E, FIPV and SARS-CoV nsp(7+8)+nsp12 complexes (in the presence of Mn^{2+}) revealed a moderate amount of product RNA generated by *de novo* initiation and a large amount of 3'-elongated ("back-primed") products if KR07 RNA with its free 3'-OH group was used (Figure 25 B; lane 2, 7 and 12). If unblocked KR07 A_{10} substrate was used, the replicase complexes of HCoV-229E and FIPV also generated *de novo* and backprimed products (Figure 25 B; lanes 4 and 14). Unlike these two complexes, the SARS-CoV replicase complex showed almost exclusively *de novo* initiated products with unmodified KR07 A_{10} template RNA (Figure 25 B; lane 9). To distinguish *de novo* initiated product RNAs from backprimed products, assays with 3' biotinylated RNAs (KR07 -b and KR07 A_{10} -b) were performed (Figure 25 B; lanes 3, 5, 8, 10, 13 and 15). The size observed for the major product indicate that it was truly generated by the mechanism of *de novo* initiation by the replicase complexes of HCoV-229E, SARS-CoV and FIPV, even though trace amounts of longer products could still be detected if the 3'-teminally blocked RNA substrates were used in reactions containing HCoV-229E and FIPV

replicase complexes (Figure 25 B; lanes 3, 13, and 15). Possibly, some of the enzyme complexes utilized in this case the nascent RNA as template for multiple rounds of RNA synthesis.

Furthermore, elongation activity was investigated for these three replicase complexes with various partial-duplex RNA substrates containing different 5'-overhangs (Figure 25 C). In all cases, the different coronavirus nsp(7+8)+nsp12 replicase complexes used the primed RNA to synthesize products of the expected size in the presence of the appropriate NTP (Figure 25 D; lanes 2-18) and Mg^{2+}. As discussed above, incorporation of GMP residues resulted in a product that migrated more slowly, most likely due to the formation of a G-quadruplex structure that could not be denatured under the conditions used. Both *de novo* and elongation results collectively suggest that all nsp(7+8)+nsp12 replicase complexes have similar polymerase activities even though they operate in conjunction with nsp(7+8) complexes that have slightly different multimerization states.

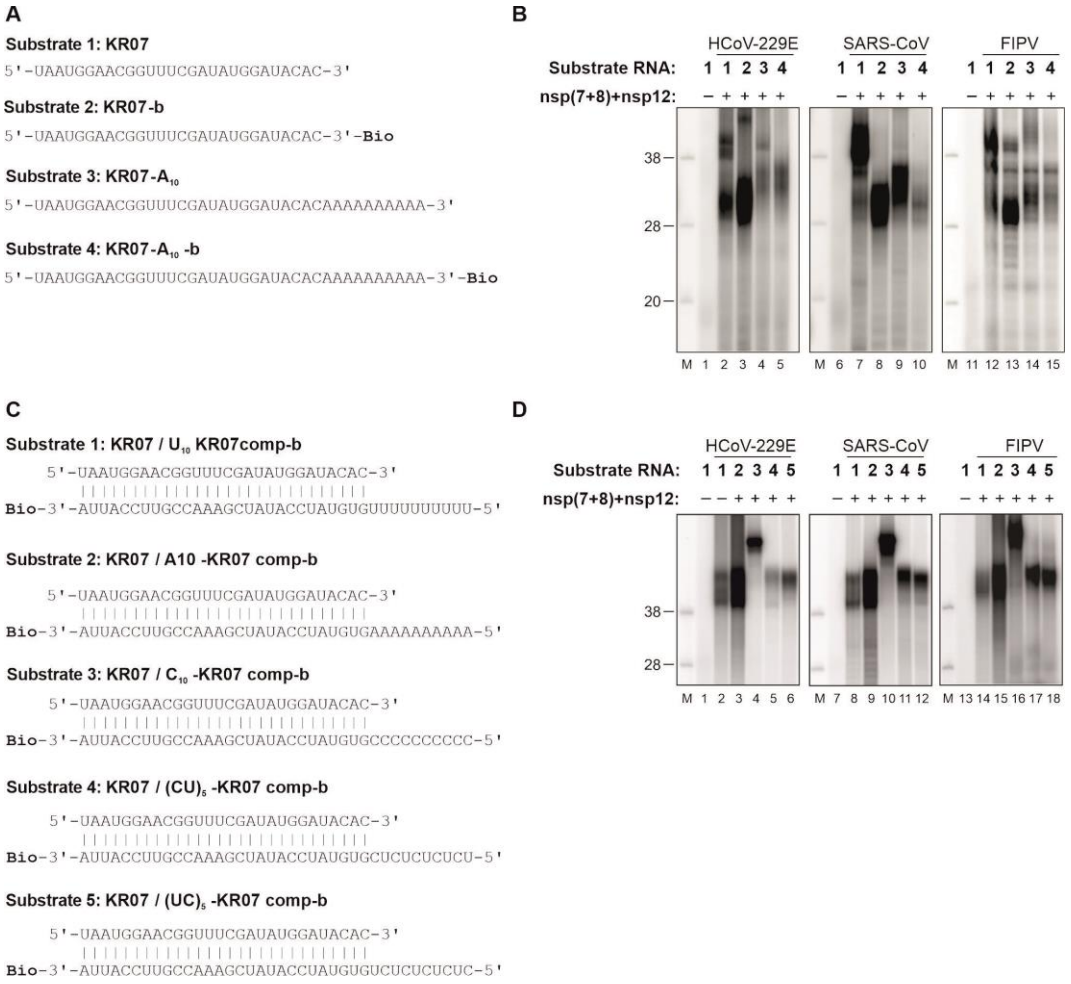


Figure 25: Replicase complexes (nsp(7+8)+nsp12) of alpha- (HCoV-229E and FIPV) and betacoronaviruses (SARS-CoV) display efficient *de novo* initiation and primer-dependent elongation activity *in vitro*. (A) Single-stranded RNA substrates used in the experiment B. 'Bio' indicates RNA 3'-biotinylation. **(B)** *The de novo* initiation assay was performed in standard reaction buffer containing 1 µM of protein, 1 µM of RNA substrate, NTP mix (100 µM of ATP, CTP, GTP and 75 µM of UTP) and 0.17 µM [α-^{32}P]-UTP in the presence of 1 mM $MnCl_2$. **(C)** Partially double-stranded RNA substrates used in the experiment D. 'Bio' indicates RNA 3'-biotinylation. **(D)** Primer-dependent elongation was performed in standard buffer supplemented with 1 µM of protein and 1 µM of RNA substrate in the presence of 1 mM $MgCl_2$. Nucleotides used in primer extension assays: substrate 1 - 100 µM ATP and 0.17 µM [α-^{32}P]-ATP; substrate 2 - 100 µM UTP and 0.17 µM [α-^{32}P]-UTP; substrate 3 - 100 µM GTP and 0.17 µM [α-^{32}P]-GTP; substrates 4 and 5 - 100 µM ATP, 100 µM GTP and 0.17 µM [α-^{32}P] GTP. *De novo* synthesis and primer extension assays contained a 1:1 molar ratio of nsp(7+8) and nsp12. Reactions were incubated at 30°C for 60 min. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[α-^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

4.8.5 The pre-formed nsp(7+8) complex modulates the RNA binding capacity of nsp12

The RNA-binding properties of nsp7, nsp8, nsp(7+8), nsp12 and combinations of

nsp7+nsp12, nsp8+nsp12, and nsp(7+8)+nsp12 were examined to better understand the possible role of the nsp(7+8) complex in activating the polymerase activity of nsp12. To this end, the KR07/ $(CU)_5$ KR07 comp-b duplex RNA substrate containing a heteropolymeric 5' $(CU)_5$ -overhang was used in a shift assay (Figure 26 A). The assay was essentially done under the conditions used for the elongation assay reactions, except for the nucleotides. As shown in Figure 26 B, RNA-binding assays revealed that there was no detectable RNA binding by nsp7 (Figure 26 B; lane 2), while nsp8 alone with bound the test RNA less tightly than the nsp(7+8) complex (Figure 26 B; lane 3) for which a stable RNA-protein complex could be detected (Figure 26 B; lane 4). In contrast, nsp12 alone had only weak RNA-binding capability. The combined use of nsp8 and nsp12 resulted in the formation of a stable complex, while two forms of stable RNA-protein complexes were detected if nsp(7+8) was used in combination with nsp12. Similar observations were made with combinations of (i) nsp12_SAA and nsp8 or (ii) nsp12_SAA+ and the nsp(7+8) complex. These results suggest that nsp8 has a major role in mediating interactions with RNA. The most efficient binding was observed for the nsp(7+8) complex, suggesting that the association of nsp7 and nsp8 results in a structure that promotes tight binding to RNA and potentially also helps to recruit nsp12 to its RNA substrate to initiate RNA synthesis.

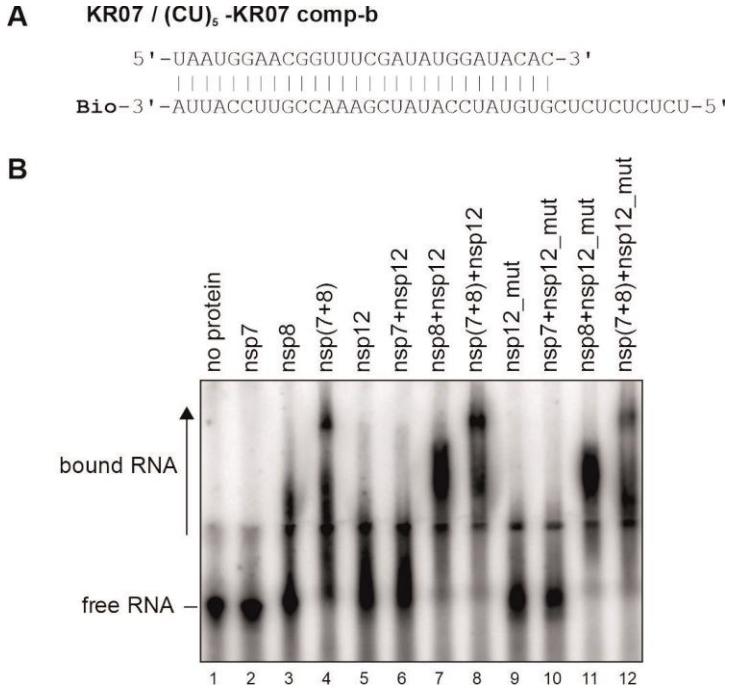


Figure 26: HCoV-229E nsp(7+8) complex promotes RNA-binding activity of its cognate nsp12: (A) RNA substrate used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** EMSA was performed in standard reaction buffer supplemented with 1 µM of protein, 0.06 µM of 5'-[^{32}P]-labeled partially dsRNA in the presence of 1 mM $MgCl_2$. Lanes 6 and 10: nsp7 and nsp12 or nsp12_mut as indicated; lanes 7 and 11: nsp8 and nsp12 or nsp12_mut as indicated; lanes 8 and 12: nsp(7+8) and nsp12 or nsp12_mut as indicated. Proteins were used at a 1:1(:1) molar ratio. nsp12_mut: nsp12 variant containing alanine substitution of conserved aspartic acid residues D4828 and D4829 (numbers indicate the positions of residues in the polyprotein 1ab of HCoV-229E). RNA-protein complexes were separated in a 0.5x TBE-buffered non-denaturing polyacrylamide (4% to 12%) gel and visualized by phosphorimaging. The positions of free and protein-bound [^{32}P]-labeled RNAs are indicated to the left.

4.9 Nucleotide selection and fidelity of the HCoV-229E nsp(7+8)+nsp12 complex

4.9.1 Metal ion titration assay

Like many viral polymerases, the coronavirus polymerase also possesses two conserved aspartic acid residues in motif C (within the conserved SDD sequence) that coordinate two metal ions in the active site of the protein (Gorbalenya et al., 2002; Kirchdoerfer and Ward, 2019; Xu et al., 2003). Studies of the poliovirus RNA-dependent RNA polymerase ($3D^{pol}$) have shown that high concentrations of Mg^{2+} or Mn^{2+} (5 mM) alter the biochemical properties of the protein and may result in an increased use of DNA

templates or incorporation of nucleotides with incorrect sugar moieties and bases (Arnold et al., 1999; Arnold et al., 2004).

We therefore investigated if (and to what extent) high concentrations of catalytic ions affect the elongation activity and the incorporation of non-complementary nucleotides during RNA elongation by the coronavirus polymerase. To determine the optimal metal ion concentration required for appropriate nucleotide selection during RNA synthesis, varying levels of Mg^{2+} and Mn^{2+} were used with the KR07/ $(CU)_5$ KR07 comp-b duplex RNA substrate (Figure 27 A) providing ATP and GTP as complementary nucleotides (Figure 27 B). As expected, the nsp(7+8)+nsp12 enzyme complex had no detectable polymerase activity in the absence of bivalent cations (Figure 27 B; lanes 1 and 13). In reactions supplemented with Mg^{2+}, products of expected length were observed at concentrations from 50 to 1000 μM (Figure 27 B; lanes 3 to 13) while for Mn^{2+} only concentrations ranging from 100 to 1000 μM produced detectable amounts of product (Figure 27 B: lanes 16 to 26). These results demonstrate that Mg^{2+} and Mn^{2+} ions are essential for primer-dependent elongation activity and slightly different minimal concentrations of the two ions are required to attain activity. Also, the elongation activity seems to be inhibited at higher Mn^{2+} (but not Mg^{2+}) concentrations.

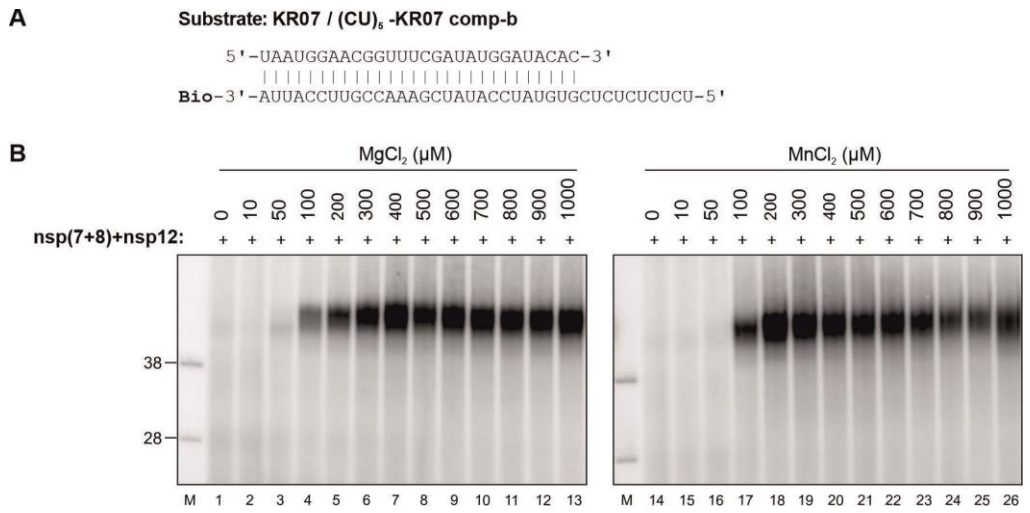


Figure 27: Determination of the effect of metal ion identities and concentrations on primer-dependent polymerase activity of the nsp(7+8)/nsp12 polymerase complex: (A) Heteropolymeric duplex RNA substrate with 5'-$(CU)_5$ overhang used. **(B)** Reaction contained 1 µM of nsp(7+8)+nsp12, 1µM RNA, 100 µM ATP + 100 µM GTP, 0.17µM [α-^{32}p]-GTP (1µCi/µl) and varying concentrations of Mg^{2+} or Mn^{2+} as indicated above the autoradiogram. RNA products were precipitated by adding 0.3 mM sodium acetate and 10 volumes ethanol and separated on 1x TBE-buffered urea-12% polyacrylamide gels. M, 5'-radiolabeled ssRNA (28nt and 38nt) were used as marker.

4.9.2 HCoV-229E nsp(7+8)+nsp12 nucleotide selection and fidelity at nonphysiological metal ion concentrations

Polymerases play a crucial role in maintaining genetic information and selecting the appropriate nucleotide during genome replication and transcription. RdRps have been shown to display a higher mutation rate than DNA polymerases (Castro et al., 2005; Drake, 1999). Except for nidoviruses with large genomes of more than 20 kb, such as coronaviruses, most RNA viruses lack dedicated proofreading enzymes during genome replication and hence depend entirely on the RdRp to control nucleotide selectivity and replication fidelity. Subissi *et al*. demonstrated that SARS-CoV nsp14 interacts with nsp12 and they proposed that nsp14/10 is able to correct errors (i.e., remove non-matching nucleotides) introduced by the nsp7-L-8+nsp12 complex while other viral factors potentially involved in nucleotide selection and fidelity remain to be investigated (Subissi et al., 2014).

To assess the possible role of metal ions in modulating replication fidelity of the nsp(7+8)+nsp12 complex, an RNA elongation assay was employed in which either complementary or non-complementary nucleotides for a partially duplex RNA substrate were provided in reactions containing either 1 mM Mg^{2+} or Mn^{2+} (Figure 28 A; substrate 1). As shown in Figure 28 B, products of expected length were catalyzed when ATP and GTP were provided (Figure 28 B and C; lanes 2 and 3). Also, radiolabeled products of (mostly) smaller sizes were found to be synthesized in almost all reactions if nonmatching nucleotides were provided, indicating nucleotide misincorporation under these conditions (Figure 28 B and C; lanes 4, 5, 6, 7). For example, significant misincorporation of ATP and CTP was detected in the presence of Mg^{2+}(Figure 28 B; lanes 4, 5, 6). Similarly, misincorporation of ATP, CTP and UTP was detected in the presence of 1 mM Mn^{2+} (Figure 28 C; lane 4, 6, 7).

Next, assays were performed with the KR07/ $(CU)_5$ KR07 comp-b duplex RNA substrate containing a 5'- $(CU)_5$ overhang (Figure 28 A; substrate 2). As shown in Figure 28 B and C; lanes 9 and 10, in the presence of Mg^{2+} or Mn^{2+} ions, products of expected length were generated, with ATP and GTP being incorporated as complementary nucleotides. Again, for both Mg^{2+} and Mn^{2+}, nucleotide misincorporation was detected when individual non-complementary nucleotides like GTP, CTP and UTP were provided (Figure 28 B and C; lanes 12, 13, 14). Of note, with substrate 2, the provision of (only) GTP (Figure 28 B; lane 12) resulted in products of the expected length if Mn^{2+} was used as metal ion cofactor, while the synthesis of full-length products was less efficient if the reaction was performed in the presence of Mg^{2+} ions (Figure 28 C; lane 13), suggesting slightly higher copy fidelity in the presence of Mg^{2+} (compared to Mn^{2+}). In other words, the presence of Mn^{2+} (rather than Mg^{2+}) ions seemed to promote the incorporation of GMP at uridylates present in the template strand.

Furthermore, we analyzed nucleotide selection with three different homopolymeric partially duplex RNAs containing either U_{10} or A_{10} or C_{10} 5'-overhangs in the presence of Mg^{2+} or Mn^{2+}. As shown in Figure 28 B and C, lanes 16, 21, 26, the nsp(7+8)+nsp12 complex generated products of expected length in the presence of the preferential complementary nucleotide. If non-complementary nucleotides were provided, products resulting from nucleotide misincorporation were observed with both Mg^{2+} and Mn^{2+} (Figure 28 B and C; lanes 17, 18, 19, 22, 23, 24, 27, 28, 29).

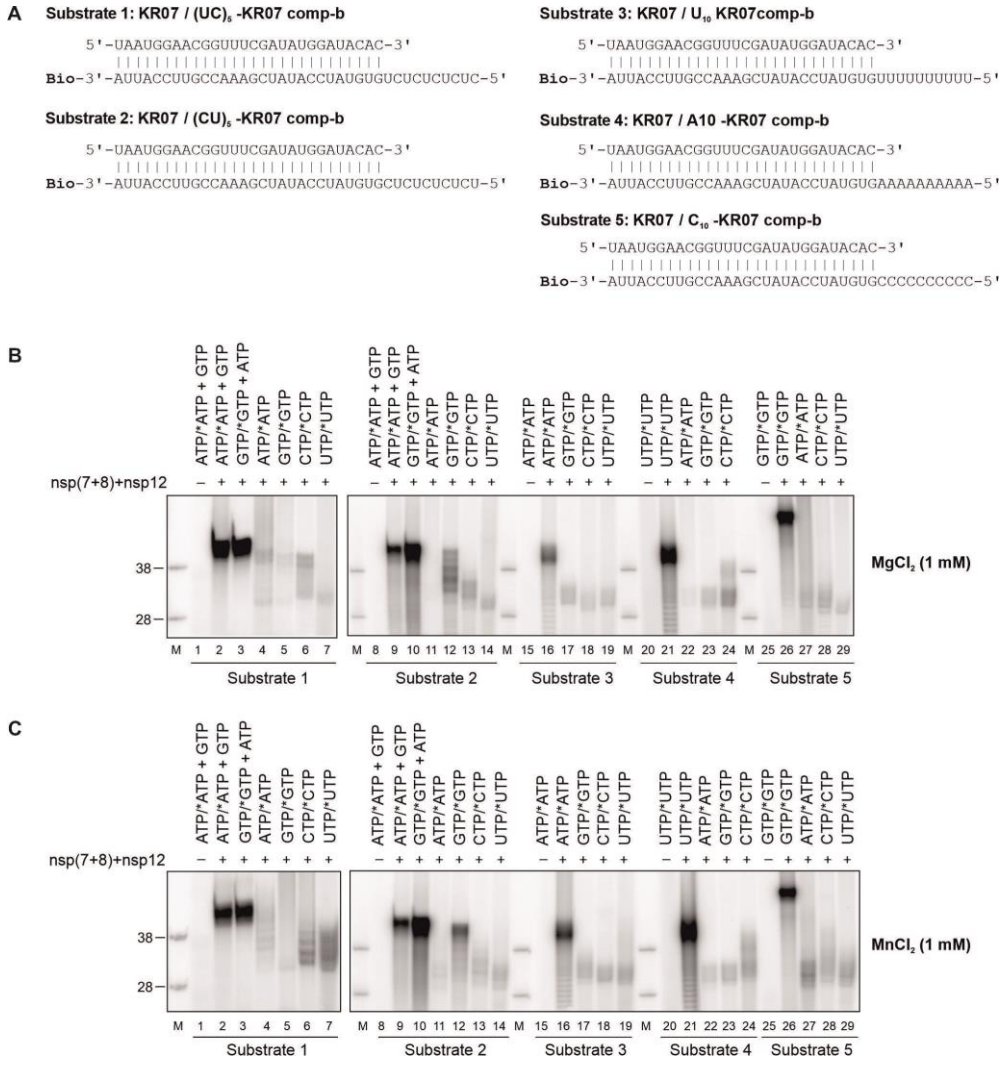


Figure 28: Nucleotide selectivity of the HCoV-229E nsp(7+8)+nsp12 complex in the presence of 1 mM Mg^{2+} or Mn^{2+} ions. (A) Partially double-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. The primer extension assay was performed in standard reaction buffer supplemented with 1 µM of polymerase complex (comprised of a 1:1 molar ratio of nsp(7+8) and nsp12), 1 µM of partially double-stranded RNA substrate and 100 µM ATP and GTP, 0.17 µM [α-^{32}P] ATP (*ATP) or 100 µM ATP and GTP, 0.17 µM [α-^{32}P] GTP

(*GTP) or 100 µM ATP, 0.17 µM [α-^{32}P] ATP *(ATP) or 100 µM GTP, 0.17 µM [α-^{32}P] GTP (*GTP) or 100 µM CTP, 0.17 µM [α-^{32}P] CTP (*CTP) as indicated above the autoradiograms. Reactions were performed at 30°C for 60 min either in the presence of 1 mM $MgCl_2$ **(B)** or $MnCl_2$ **(C)** as indicated on the right side of the autoradiograms. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[α-^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of the autoradiograms.

4.9.3 HCoV-229E nsp(7+8) replicase complex nucleotide selection and fidelity at reduced concentrations of Mg^{2+} and Mn^{2+}

As shown above, the nsp(7+8)+nsp12 enzyme complex requires Mg^{2+} or Mn^{2+} ions for activity, with maximum activities in the range of 400 µM for Mg^{2+} or 200 µM for Mn^{2+}, that is, below the concentrations used in the standard reaction buffer. To investigate possible effects of reduced metal ion concentrations on the enzyme complex's ability of selecting the correct nucleotide, five partial-duplex RNAs (Figure 29 A) were used in polymerase assays that were supplemented with matching and nonmatching nucleotides, respectively. When the appropriate (complementary) nucleotides were provided, the enzyme complex incorporated these nucleotides efficiently in the presence of 240 µM Mg^{2+} or 100 µM of Mn^{2+}, resulting in the expected products (Figure 29 B; lanes 2, 3, 9, 10, 15, 20, 25; Figure 29 C; lanes 2, 3, 8, 9, 15, 20, 26). Also, under reaction conditions with reduced metal ion concentrations, the abundance of products resulting from incorporation of nonmatching nucleotides was found to be reduced and this effect was particularly evident for Mn^{2+} (Figure 29 B, compare with Figure 28 B and C). Collectively, these *in vitro* results indicate that high metal ion concentrations reduce the fidelity of the coronavirus (nsp7+8)+nsp12 polymerase complex *in vitro*.

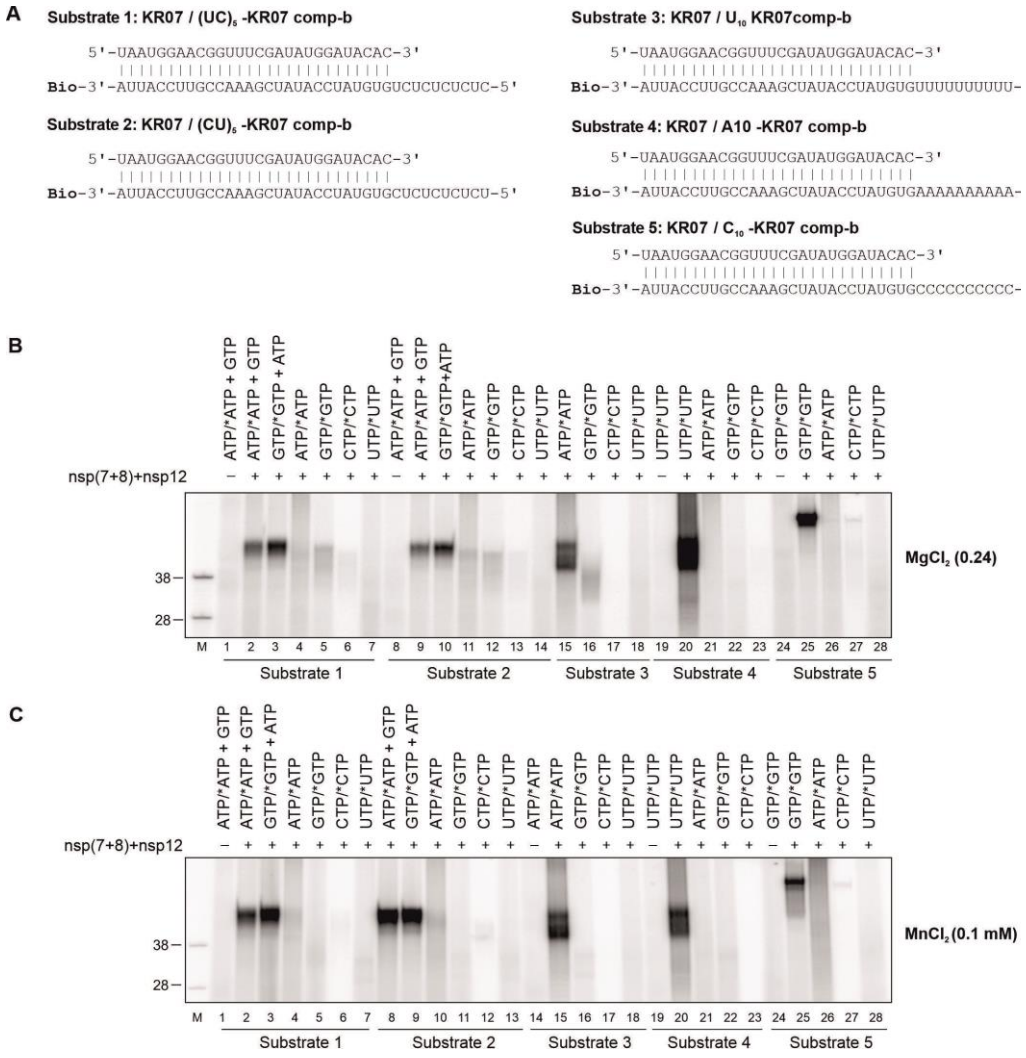


Figure 29: Nucleotide selectivity of the HCoV-229E nsp(7+8)+nsp12 complex in the presence of reduced Mg^{2+} or Mn^{2+} ion concentrations. (A) Partially double-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. The primer extension assay was performed in standard reaction buffer supplemented with 1 µM polymerase complex (comprised of a 1:1 molar ratio of nsp(7+8) and nsp12), 1 µM of partially double-stranded RNA substrate and 100 µM ATP and GTP, 0.17 µM [α-^{32}P] ATP (*ATP) or 100 µM ATP and GTP, 0.17 µM [α-^{32}P] GTP (*GTP) or 100 µM ATP, 0.17 µM [α-^{32}P] ATP *(ATP) or 100 µM GTP, 0.17 µM [α-^{32}P] GTP (*GTP) or 100 µM CTP, 0.17 µM [α-^{32}P] CTP (*CTP) as indicated above the autoradiogram. Reactions were performed at 30°C for 60 min in the presence of either 0.24 mM $MgCl_2$ **(B)** or 0.1 mM $MnCl_2$ **(C)** as indicated on the right side of the autoradiograms. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of the autoradiograms.

4.9.4 HCoV-229E nsp(7+8)+nsp12 complex can differentiate between rNTPs and dNTPs

To test the enzyme complex's ability to discriminate between NTPs and dNTPs during RNA synthesis, we performed primer-dependent polymerase activity assays with two

different duplex RNA substrates in the presence of 240 µM Mg^{2+} (Figure 30 A) or 100 µM Mn^{2+} (Figure 30 B) and the complementary NTPs or dNTPs. As shown in Figure 30 B, the enzyme complex only synthesized a product of expected length when NTPs were provided (Figure 30 B; lanes 2, 4, 7, 9, 12, 14, 17, 19). However, with both RNA substrates, trace amounts of smaller radiolabeled products were also detected in reactions containing dATP and dGTP in the presence of Mg^{2+} and Mn^{2+} ions (Figure 30 B; lanes 3, 5, 8, 10, 13, 15, 18, 20). These results suggest that, consistent with its presumed role as viral RNA-dependent RNA polymerase, the enzyme complex differentiates between NTPs and dNTPs. There was no evidence for efficient RNA-dependent DNA polymerase (reverse transcriptase) activity.

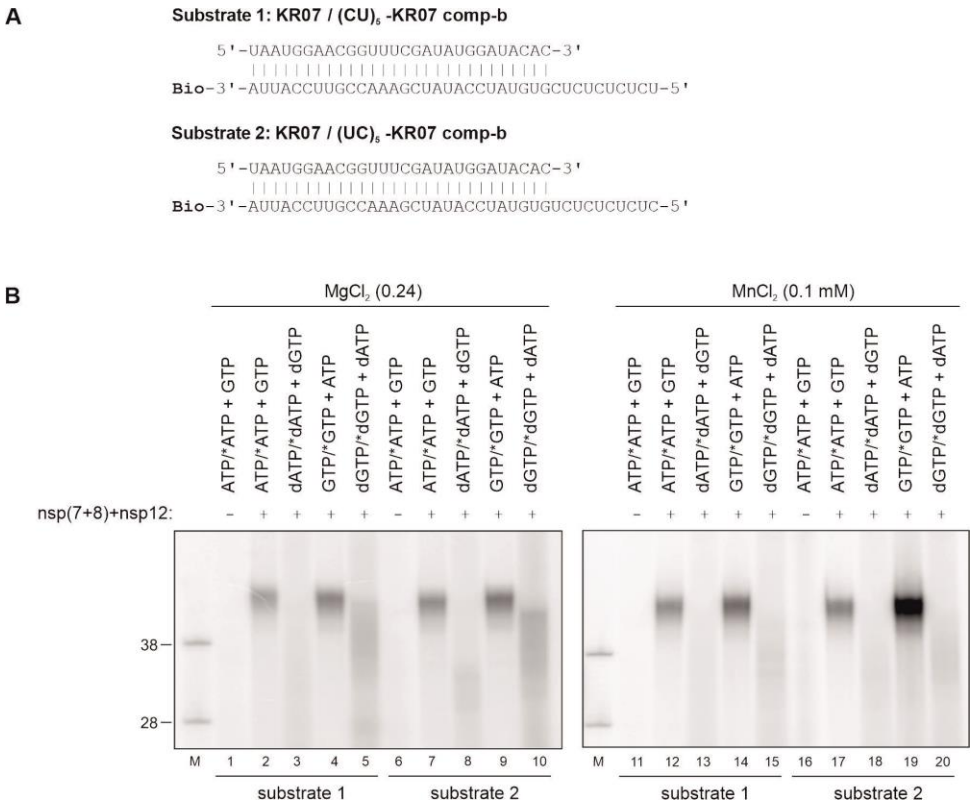


Figure 30: Selectivity of the HCoV-229E nsp(7+8)+nsp12 complex for NTPs rather than dNTPs. (A) Partially double-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** The primer extension assay was performed in standard reaction buffer supplemented with 1 µM polymerase complex (comprised of a 1:1 molar ratio of nsp(7+8) and nsp12), 1 µM RNA substrate and 100 µM ATP and GTP, 0.17 µM [α-^{32}P] ATP (*ATP) or 100 µM ATP and GTP, 0.17 µM [α-^{32}P] GTP (*GTP) or 100 µM dATP and dGTP, 0.17 µM [α-^{32}P] dATP (*dATP) or 100 µM dGTP and dATP, 0.17 µM [α-^{32}P] dGTP (*GTP) as indicated above the autoradiograms. Reactions were performed in the presence of either $MgCl_2$ (240 µM) or $MnCl_2$

(100 µM) as indicated on the right side of the autoradiograms and incubated for 1 h at 30°C. Products were resolved in a TBE-buffered 12 % polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[α-^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

5 Characterization of the equine arteritis virus replicase complex

5.1 Expression and purification of EAV replicase proteins

Earlier reports from Beerens *et al.* and te Velthuis *et al.* suggested a *de novo* polymerase activity for the 78-kDa nsp9 protein of EAV (the homolog of coronavirus nsp12) in the absence of other viral proteins (Beerens et al., 2007; te Velthuis et al., 2010b). To validate these data and obtain insight into possible cofactors of the nsp9-mediated RdRp activity, the coding sequences of nsp6-7-8, nsp7, nsp7α, nsp7β, nsp9, nsp10 and nsp12, were cloned as pASK-Ub-nsp-His_6 plasmid constructs. *E. coli* TB1 [pCGI] were transformed with the appropriate plasmid and 1 liter of Luria Broth containing ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) was inoculated with 5 ml of an overnight pre-culture. Cells were grown at 37°C until OD_{600} of 0.4. At this point, 1 ml of the cell suspension was collected as a control ("uninduced"). Then, the expression was induced by adding anhydrotetracycline (AHT) (200 ng/ml final concentration) and the cells were incubated at 18°C for 16 h. 1 ml of the cell suspension was collected and used to produce a total lysate of "induced" cells. Cells from the induced 1-liter culture were harvested by centrifugation and the recombinant proteins were purified by metal affinity chromatography using Ni-NTA and anion exchange chromatography. To remove degraded proteins and large protein aggregates, size-exclusion chromatography was performed, and the purified proteins were stored at –20°C in 50 mM HEPES-KOH/pH 8, 150 mM KCl, 40 % [v/v] glycerol and 10 mM β-mercaptoethanol until further use.

Figure 31 A provides an example for the expression and purification of EAV nsp9 by Ni-IMAC chromatography. The analysis shows overexpression of a ~78-kDa protein upon induction (Figure 31 A; compare lanes 1 and 2), which corresponds to the expected size of this protein, and the purified protein (Figure 31 A; lane 3). Also, a set of mutant derivatives of the pASK-Ub-nsp9-H_6 expression construct carrying a C-terminal His_6 tag was generated via site-directed mutagenesis. The mutations led to substitutions of predicted active-site residues in the NiRAN and/or RdRp domains of this protein (Lehmann et al., 2015). Nsp9 wild-type and mutant proteins were purified by affinity and anion exchange chromatography (Figure 31 B; lanes 1-14).

The coding sequence for the EAV main protease nsp4 was cloned in the pMAL-c2X vector. The protein was expressed as a MBP fusion protein with a C-terminal His_6 tag and purified as described in Materials and Methods (see 3.12.1).

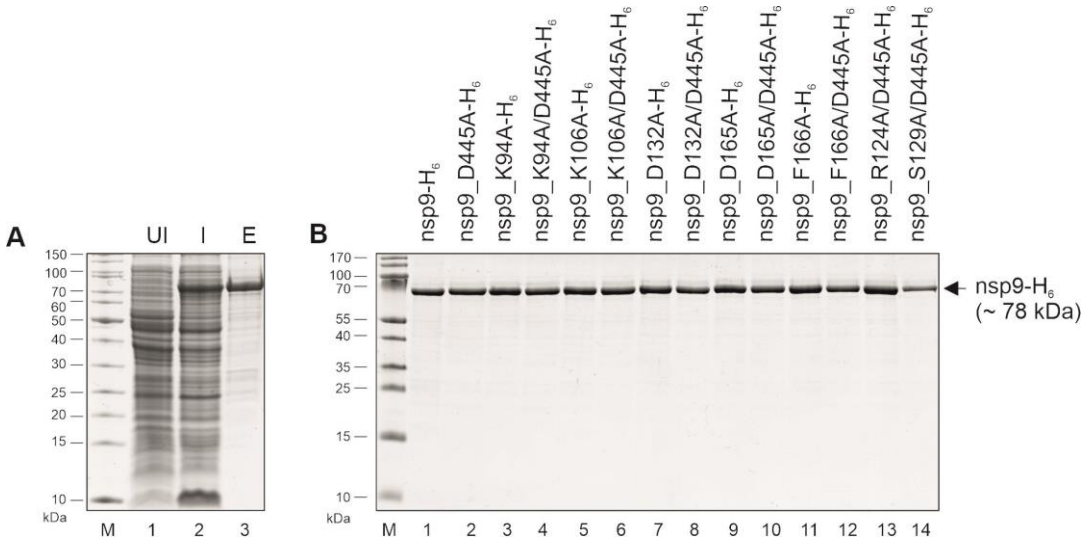


Figure 31: Tris-tricine-PAGE analysis documenting the expression and purification of EAV nsp9-H_6 and mutant derivatives of this protein: (A) Expression of the recombinant nsp9 (wildtype) was monitored by SDS-PAGE analysis (12% polyacrylamide gel) using cell lysates obtained from induced (lane 2) and uninduced (lane 1) cells. Nsp9 was purified by Ni-IMAC affinity chromatography (lane 3). **(B)** The Ni-NTA purified proteins were further purified by anion exchange chromatography. 3 µg of the anion-exchange purified wild-type EAV nsp9 and its mutant derivatives carrying the indicated single or combined alanine replacements of essential residues of the EAV nsp9 NiRAN and RdRp domain (D445 and K94, K106, D132, D165, F166 in combination with D445) were analyzed by Tris-tricine PAGE on a 12 % polyacrylamide gel. Size markers are depicted on the left side in kDa.

5.2 Biochemical characterization of EAV replicase proteins

5.2.1 EAV nsp9 possesses terminal adenylyltransferase activity

To explore the activity of nsp9 in more detail, we used four sets of substrate RNAs corresponding to those described by Lehmann et. al., 2016 (Lehmann et al., 2016). These substrates had a longer 5'-overhang compared to the substrates used for the coronavirus polymerase. The reaction mixtures were incubated at 30°C for 60 minutes with varying nucleotide conditions to assess if the recombinant nsp9 (i) acts as terminal adenylyltransferase activity on ssRNA, (ii) is able to synthesize the complement of a $(CU)_{10}$ overhang of a partial-duplex RNA substrate if ATP or GTP is provided and (iii) is able to synthesize the A_{20} complement of an U_{20} overhang of a partial-duplex RNA substrate in a reaction containing ATP. The assays were performed in buffer containing 20 mM HEPES-KOH/pH 8, 30 mM KCl, 1 mM $MgCl_2$, 0.2 % [v/v] Triton X-100, 1.5 mM β-mercaptoethanol, 1 mM DTT, 5 % [v/v] glycerol, 1 µM substrate RNA, 100 µM NTP, 0.17 µM [α-^{32}P]-NTP and 1 µM nsp9. To address the question of whether EAV nsp9 has terminal transferase activity, we used KR01 ssRNA along with a modified form of this RNA that carried a 3'-biotin group (KR01-b) and thus lacked a free 3'-hydroxyl group (Figure 32 A; substrate I and II). The reaction mixtures also contained 100 µM ATP and 0.17 µM [α-^{32}P]-ATP and 1 mM Mg^{2+}. As shown in Figure 32, lane 3, ATP was incorporated by EAV nsp9 if the reaction was performed with an RNA substrate carrying an unmodified 3' end, while no radiolabeled products were generated if a 3'-biotinylated RNA substrate was used (Figure 32, lane 2). To investigate if (and to what extent) the presence and sequence of an opposite ("template") strand affects this activity, partially double-stranded RNA substrates with different 5'-overhangs were used. Using the KR01/ JTR1-b RNA substrate, which contained a 5' $(CU)_{10}$-overhang

(Figure 32 A: substrate III), in a reaction supplemented with100 µM ATP and 0.17 µM [α-^{32}P]-ATP and Mg^{2+}, radiolabeled products were found to be synthesized whose sizes exceeded that of the "template" strand, suggesting that EAV nsp9 incorporated AMP in a template-independent manner in this case. The role of the template was further characterized using the KR01/ U_{20} JTR1-b substrate in which the bottom strand contained a U_{20} overhang at its 5'-end (Figure 32 A; Substrate IV). The reaction mixture also contained 100 µM ATP and 0.17 µM [α-^{32}P]-ATP as complementary nucleotides and 1 mM Mg^{2+}. Reaction mixtures were incubated at 30°C for 60 minutes under the buffer conditions used in previous assays. In this case, nsp9 efficiently incorporated ATP to generate a product whose size corresponded to that of the bottom strand (40 nt), suggesting that the U_{20} sequence of the bottom strand greatly facilitated the incorporation of ~20 AMPs into the top strand, thereby possibly acting as a template (Figure 32 B; lane 5). In contrast to AMP, there was no evidence for nsp9-mediated RNA 3'-guanylyltransferase activity in reactions containing RNA II as a substrate and 100 µM GTP/0.17 µM [α-^{32}P]-GTP (Figure 32 B; lane 6), indicating that nsp9 is unable to incorporate GMP under the conditions used.

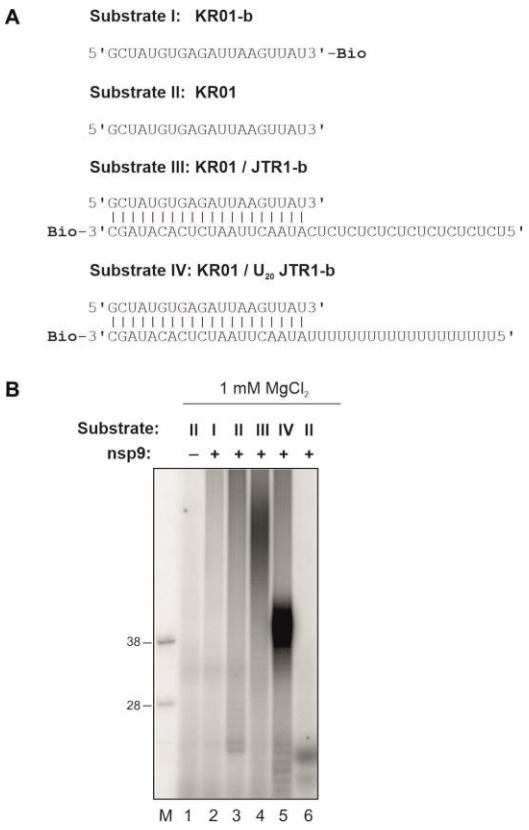


Figure 32: EAV nsp9-mediated polyadenylyltransferase activities on ss- and partially ds-RNA substrates: (A) RNA substrates I to IV used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** Reactions were performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of protein and 1 µM of RNA substrate as indicated above the autoradiograms and in the presence of 1 mM $MgCl_2$. RNA substrates used in the respective reactions are indicated on the top of autoradiogram. Lane 1 – reaction performed in the absence of protein, but with 100 µM ATP and 0.17 µM [α-^{32}P]-ATP, lane 2 – as a negative control KR01-b RNA substrate was used with its 3'-OH blocked with biotin and the reaction was performed in the presence of 100 µM ATP and 0.17 µM [α-^{32}P]-ATP, lane 3 – the reaction contained 100 µM ATP and 0.17 µM [α-^{32}P]-ATP, lane 4 – the reaction was performed with substrate containing a free 3'-OH on the top strand and the 3'-OH of bottom strand was blocked with biotin, in the presence of 100 µM ATP and 0.17 µM [α-^{32}P]-ATP, lane 5 – reaction with 100 µM ATP and 0.17 µM [α-^{32}P]-ATP, lane 6 – 100 µM GTP and 0.17 µM [α-^{32}P]-GTP.Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

5.3 The NiRAN domain of EAV nsp9 transfers UMP and GMP to the nsp6-7-8 polyprotein

The data presented above suggest that EAV nsp9 has terminal adenylyltransferase activity and is able to transfer AMP to the (unmodified) 3'-end of single-stranded or double-stranded RNA substrates. Prompted by information on the coronavirus

polymerase complex, which is known to require the co-factor nsp(7+8) complex to attain primer-dependent and *de novo* polymerase activity (see chapter 4.8), we decided to express and purify a set of small nonstructural proteins and cleavage intermediates encoded by the 3'-terminal region of the EAV ORF1a. This set included nsp4, nsp6-7-8, nsp7, nsp7α and nsp7β, nsp10 and nsp12 (Figure 33). All these proteins contained a C-terminal hexahistidine tag.

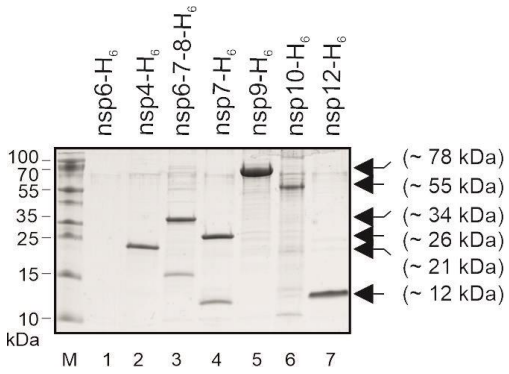


Figure 33: Tris-tricine-PAGE analysis of bacterially expressed and purified EAV nonstructural proteins: EAV nsp6-H_6, nsp6-7-8-H_6, nsp7-H_6, nsp9-H_6, nsp10-H_6 and nsp12-H_6 were expressed in TB1-pCGI transformed with the according pASK-Ub-nsp-H_6 construct. The proteins were purified by Ni-IMAC affinity chromatography. MBP-nsp4-H_6 was produced in TB1 cells transformed with pMAL-c2X-MBP-nsp4-H_6. The protein was purified by Ni-IMAC affinity chromatography and the eluted (MBP-nsp4-H_6) fusion protein was cleaved with factor Xa. The nsp4-H_6 protein was subsequently seperated from MBP using Ni-NTA affinity chromatography. The Ni-IMAC-purified proteins were further purified by anion exchange chromatography. 2 µg of nsp4-H_6, nsp6-H_6, nsp6-7-8-H_6, nsp7-H_6, nsp9-H_6, nsp10-H_6, and nsp12-H_6 were analyzed on a 12 % Tris-tricine-PAGE. M – prestained marker proteins (PageRuler) with molecular masses (in kDa) indicated to the left and recombinant EAV proteins with approximate molecular masses indicated to the right. Note that, due its small size (2.4 kDa) nsp6-H_6 is not resolved in this gel.

In an attempt to demonstrate EAV nsp9-mediated *de novo* polymerase activity, nsp9 was used in combination with other EAV ORF1a- and ORF1b-encoded nonstructural proteins using the 32-nt oligonucleotide KR07 A_4-b RNA as a template (Figure 34 A). The reactions were performed in buffer containing 20 mM HEPES-KOH/pH 8, 30 mM KCl, 1 $MgCl_2$, 0.2 % [v/v] Triton X-100, 1.5 mM β-mercaptoethanol, 0.5 mM DTT, 5 % [v/v] glycerol, 1 µM RNA substrate, 100 µM of NTPs, 0.17 µM [α-^{32}P]-UTP and 1 µM protein(s). As shown in Figure 34 B (lanes 1 to 11, 13,14), reactions containing

nsp4, nsp6, nsp7, nsp6-7-8, nsp9, nsp10 and nsp12, respectively, or combinations of nsp4+nsp6-7-8, nsp4+nsp9, nsp7+nsp9, nsp9+nsp10 and nsp9+nsp12 did not give rise to radioactively labeled products. Interestingly, the reaction performed in the presence of nsp6-7-8 and nsp9 (Figure 34 B, lane 12) resulted in a radiolabeled product, albeit with a size that was much smaller than expected. Upon mixing the serine protease nsp4 with nsp6-7-8+nsp9, the intensity of the radiolabeled product was reduced (Figure 34 B; lane 15), suggesting that nsp4-mediated cleavage of the nsp6-7-8 polyprotein (between nsp6↓7 and nsp7↓nsp8) into three individual proteins (nsp6+nsp7+nsp8) diminished the activity responsible for this radiolabeled product. In another reaction (Figure 34 B, lane 16), the nsp9_D445A RdRp active-site mutant protein was used to determine if the formation of this radiolabeled product involved the catalytically active RdRp domain. Surprisingly, the use of this RdRp active-site mutant protein instead of the wild-type nsp9 had no effect on the formation and abundance of this product. To further investigate the activity responsible for the formation of this radiolabeled product, we targeted the NiRAN domain by replacing putative active-site (and control) residues (Lehmann et al., 2015). Specifically, we replaced the conserved K94 and the non-conserved residues K106 of the NiRAN domain with alanine (Lehmann et al., 2015) and assays were carried out in the presence of nsp4+nsp6-7-8 (Figure 34 B; lanes 16 and 17). Interestingly, with nsp9_K94A (having an inactive NiRAN), the labeled product was no longer detectable (Figure 34 B; lanes 17) while nsp9_K106A (control) retained activity (Figure 34 B; lanes 18). Taken together, these data suggest that the reaction resulting in the radiolabeled product was catalyzed by the NiRAN rather than the RdRp domain of nsp9. To explore potential roles of yet other replicase gene-encoded proteins for *de novo* initiation of nsp9-mediated RNA synthesis, a combination of nsp4, nsp6-7-8, nsp10, nsp12 and nsp9 was tested using KR07 A_4 -b as a template RNA. The

reaction was performed in the presence of all 4 NTPs, 0.17µM [α-^{32}P]-UTP and 1mM $MnCl_2$ (Figure 34 B; lane 19). Again, no (additional) product of the expected size was detected.

To investigate if these radiolabeled products appear only in the presence of radiolabeled UTP, the assays were also performed with two other radiolabeled [α-^{32}P]-NTPs, each at 0.17 µM, while the concentrations of the 4 nonlabeled NTPs remained unchanged (Figure 34 B; lanes 20 and 21). The combination of nsp4, nsp6-7-8, and nsp9 did not give rise to a radiolabeled product if 0.17 µM [α-^{32}P]-ATP was used instead of 0.17 µM [α-^{32}P]-UTP (Figure 34 B, lane 20), suggesting that ATP is not incorporated into the radiolabeled product observed in several other reactions (Figure 34 B; lanes 12, 15, 16, 18, 19, 22). In striking contrast, a major radiolabeled product was obtained when 0.17 µM [α-^{32}P]-UTP was replaced with [α-^{32}P]-GTP (Figure 34 B; lane 21). In this case, the nsp6-7-8+nsp9 complex gave rise to a product that migrated a bit more slowly compared to the product observed in reactions containing [α-^{32}P]-UTP (lanes 15 and 21). To compare the products observed for EAV replicase gene-encoded proteins, an additional control experiment was performed using the previously characterized HCoV-229E nsp(7+8)+nsp12 complex in the presence of NTPs (each at 100 µM), 0.17 µM [α-^{32}P]-UTP and Mn^{2+}. As expected, the HCoV-229E nsp(7+8)+nsp12 complex produced a radiolabeled RNA with a size similar to that of the 3'-biotinylated template RNA (Figure 34 B; lane 23). Thus, in contrast to the coronavirus data, none of the EAV protein combinations used in this set of experiments gave rise to a full-length copy of the template RNA provided. Instead, faster migrating [α-^{32}P]-UMP- or [α-^{32}P]-GMP-labeled products were detected in the presence of nsp6-7-8 and nsp9. These products remained detectable with nsp9_D445A (RdRp mutant) but not with nsp9_K94A (NiRAN mutant), possibly implicating the NiRAN (rather than RdRp) domain in this

activity (see below). With regard to the arterivirus RdRp activity, the protein and substrate requirements that are needed to attain activity *in vitro* remain elusive at this stage.

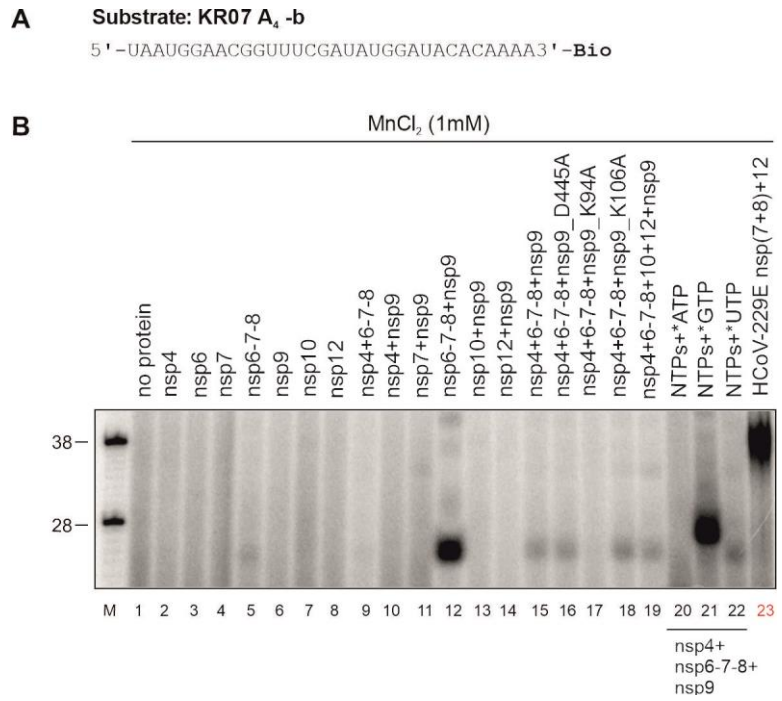


Figure 34: Analysis of *de novo* RNA synthesis by EAV ORF1a- and ORF1b-encoded nonstructural proteins and the HCoV-229E nsp(7+8)+nsp12 polymerase complex. (A) RNA substrate used in *de novo* RNA synthesis assay experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** *De novo* RNA synthesis assays were performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of each protein (alone or in the indicated combination), 1 µM of RNA substrate in the presence of 1 mM $MnCl_2$. The reactions were performed in the presence of 100 µM NTPs and 0.17 µM [α-^{32}P]-UTP in lanes 1 to 19, 22, and 23. In lane 20: with 100 µM NTPs and 0.17 µM [α-^{32}P]-ATP and and in lane 21: with 100 µM NTPs and 0.17 µM [α-^{32}P]-GTP. Lane 1: no protein was included. In lanes 16, 17, and 18, respectively, the nsp9 wild-type protein was replaced with nsp9_D445A (substitution in motif A of the nsp9 RdRp domain), nsp9_K94A (substitution of the active-site K94 residue of the nsp9 NiRAN domain), and nsp9_K106A (substitution of the non-conserved K106 residue of the nsp9 NiRAN domain). Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as marker. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

5.4 Nucleotidylation activity

5.4.1 The NiRAN domain of EAV nsp9 transfer UMP and GMP nucleotides to nsp6-7-8

The experiments presented above using EAV ORF1a/1b encoded nsps failed to provide evidence for RNA synthesis, even though the combination of nsp6-7-8 and nsp9 gave rise to a radiolabeled product of unexpected size if [α-^{32}P]-UTP or [α-^{32}P]-GTP and Mn^{2+} were included in the reaction. The identity of this product was not clear but could be related to protein nucleotidylation or 5'-nucleotidylation of the RNA substrate. This idea was mainly supported by the observation that an alanine replacement of the NiRAN active-site residue K94 (Lehmann et al., 2015) abolished the activity resulting in this particular product, while the RdRp active-site mutation (D445A) retained this activity (Figure 34 B, lane 16). These data prompted us to explore a potential protein nucleotidylation activity of the NiRAN domain of EAV nsp9.

Protein uridylylation was tested in reactions containing 1 µM of individual proteins (nsp6-7-8, nsp9, nsp9_K94A) or a 1:1 mixture of proteins (each at 1 µM) in which nsp6-7-8 was mixed with nsp9 or nsp9_K94A. The reaction was performed in the presence of 25 µM UTP and 0.17 µM [α-^{32}P]-UTP in buffer containing 20 mM HEPES-KOH/pH 8, 30 mM KCl, 1 mM $MnCl_2$, 0.2 % [v/v] Triton X-100, 5 % [v/v] glycerol, 1.5 mM β-mecaptoethanol and 1 mM DTT. Following incubation at 30°C for 60 min, the reaction products were heated at 100°C for 10 minutes and seperated by Tris-tricine-PAGE to identify proteins with covalently bound nucleotides. As shown in Figure 35 (lanes 2 and 3), weak signals indicating uridylylation could be observed for nsp6-7-8 and nsp9 , but not for nsp9_K94A (Figure 35; lanes 2, 3, and 5). In contrast, if nsp6-7-8 and nsp9 were used in combination, uridylylated proteins could be readily detected. The mobilities of

the radiolabeled proteins corresponded to those of nsp6-7-8 and nsp9 in the Coomassie-stained gel (Figure 35; lane 4, top panel). No labeled products were detected in the reaction containing nsp6-7-8 and nsp9_K94A (Figure 35; lane 6), suggesting that the NiRAN domain has protein uridylylation activity in *cis* and *trans* and that these activities critically depend on the K94 residue. The data also suggest that the presence of nsp6-7-8 stimulates the (self)uridylylation activity of nsp9, possibly by a conformational change induced upon binding of nsp9 to nsp6-7-8. The basis and biological relevance for the minor uridylylation observed for nsp6-7-8 remain to be corroborated in future studies.

The assay was repeated under identical conditions, except that UTP was replaced with 25 μM GTP, 0.17 μM [α-^{32}P]-GTP. As shown in Figure 35 (lanes 8, 9 and 12), week signals indicating inefficient guanylylation were obtained for nsp6-7-8 and nsp9, while efficient guanylylation of nsp6-7-8 and slightly less efficient (self)guanylylation of nsp9 was observed in the reaction containing nsp6-7-8 and the wild-type nsp9 (Figure 35; lane 10). Only trace amounts of guanylylated proteins were observed if nsp9 was replaced with nsp9_K94A carrying a NiRAN active-site mutation (Figure 35; lane 12). Based on these data, it is tempting to suggest (but needs to be confirmed) that the labeled products observed in the polymerase assays shown in Figure 34 represent peptides that are (i) derived from nsp6-7-8 (and, possibly, nsp9) and (ii) guanylylated/uridylylated by the nsp9 NiRAN domain.

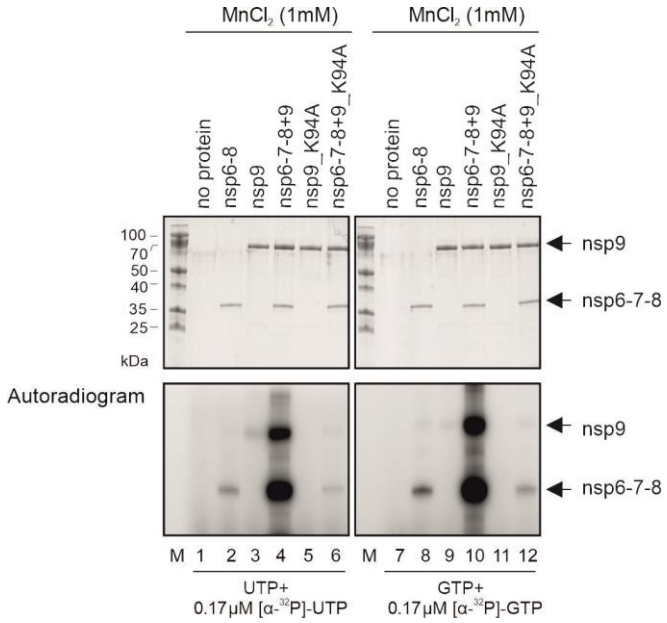


Figure 35: Identification of nsp6-7-8 as an acceptor protein for nsp9-mediated uridylylation and guanylylation. Individual proteins or combinations of proteins (each at 1 µM) were incubated at 30°C for 60 min in standard reaction buffer containing 1 mM MnCl2 and 25 µM UTP/0.17 µM [α-32P]-UTP or 25 µM GTP/0.17 µM [α-32P] GTP. Proteins used in different reactions are indicated above (lanes 2-6, 8-12). In control reactions, no proteins were included (lanes 1 and 7). The reaction mixtures were separated in a 12% Tris-tricine polyacrylamide gel. Proteins were stained with Coomassie brilliant blue (top) and radiolabeled proteins were detected by phosphorimaging of the same gel (bottom). Arrows on the right indicate the positions of nsp9 and nsp6-7-8. M – prestained protein marker (PageRuler).

5.4.2 Further characterization of guanylylated and uridylylated EAV nsp6-7-8-H_6 complexes

To obtain more information on the nucleotidylated forms of nsp6-7-8, we embarked on mass spectrometry analyses in a collaboration with Dr. Uwe Linne (Department of Chemistry, University of Marburg). Uridylylation and guanylylation assays, along with control reactions, were performed with nsp6-7-8 and nsp9 (each at 10 µM) in standard reaction buffer including 1 mM Mn^{2+} and 500 µM UTP or GTP. Following incubation at 30°C for 60 min, the reaction mixtures containing (i) nsp6-7-8, (ii) nsp6-7-8+nsp9, (iii) nsp6-7-8, nsp9, and UTP or (iv) nsp6-7-8, nsp9, and GTP, were desalted and subjected to mass spectroscopy using the ESI source of a Synapt G2Si mass spectrometer. As shown in Figure 36 A and B, a single population of mass species was

observed: as an apo protein, containing the intact molecular mass 33789 Daltons for nsp6-7-8 and the same mass was detected for the complex of nsp6-7-8+nsp9. This indicates that the detected mass is related to the nsp6-7-8 protein and the mass spectrometry failed to produce spectra that allowed for the detection of nsp9. The mass spectrum for the uridylylation reaction with nsp6-7-8+nsp9+UMP yielded two different molecular mass populations: nsp6-7-8 as apo form and the stable and covalent complex nsp6-7-8-UMP (Figure 36 C). The distribution of these mass species is different, with the apo form being less abundant. The highest mass species of 34095 Daltons corresponds to the nsp6-7-8-UMP complex, while the lower intensity species with 33789 Daltons represents the nucleotide-free nsp6-7-8 protein. The size difference between nucleotide-free nsp6-7-8 and nsp6-7-8-UMP complex is 306 Daltons, which corresponds to the mass of UMP. When the guanylyltransferase reaction (nsp6-7-8+nsp9+GMP) was subjected to mass spectrometry, two mass species were detected (Figure 36 D). Of these, the lower intensity mass species with 33789 Daltons represents the nucleotide-free nsp6-7-8 and the higher mass species (34134 Daltons) the nsp6-7-8-GMP complex. The identified mass difference between these two species is 345 Daltons, corresponding to the mass of GMP, confirming that exactly one molecule of GMP was bound covalently to nsp6-7-8. If the binding had been non-covalent, the complexes would have been separated since the samples were kept in a low-pH buffer compatible with mass spectrometry analyses (pH 2.0).

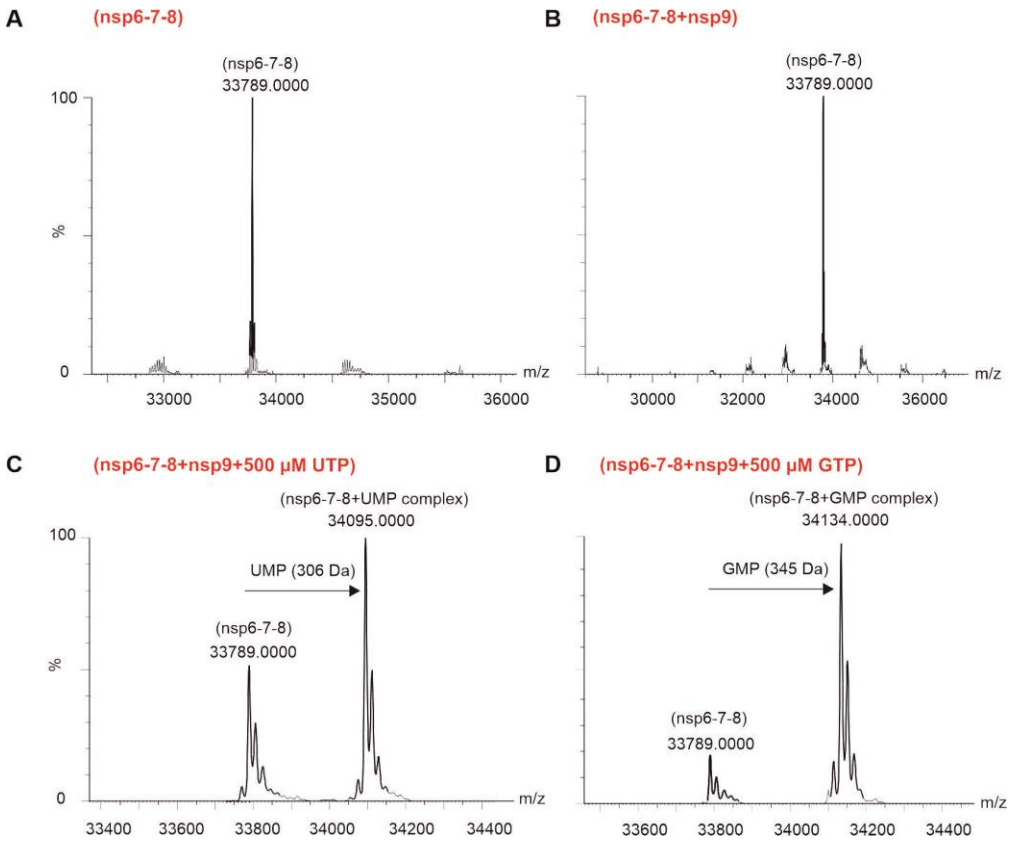


Figure 36: Mass spectroscopy reveals nsp9-mediated covalent transfer of UMP and GMP to nsp6-7-8. Shown are deconvoluted intact protein mass spectra obtained for EAV nsp6-7-8 (in its unmodified and nucleotidylated form) in reactions containing nsp9 and UTP or GTP. **(A)** nsp6-7-8, **(B)** nsp6-7-8, nsp9, **(C)** nsp6-7-8, nsp9, and UTP, **(D)** nsp6-7-8, nsp9, and GTP. The nucleotidylation assay and control reactions were performed with the indicated proteins (each at 10 µM) and nucleotides (500 µM) in buffer containing 20 mM HEPES/pH 8, 30 mM KCl, 5 % [v/v] glycerol, 0.2 % [v/v] Triton X-100, 1.5 mM β-mercaptoethanol, 1 mM DTT, and 1 mM $MnCl_2$ at 30ºC for 1 hour. Samples were desalted and analyzed using a Synapt G2Si mass spectrometer. Intact protein masses relating to proteins and protein-NMP complexes are indicated. Also indicated are the mass differences between proteins and protein-NMP complexes. The X-axis represents mass-to-charge (m/z) values, and the Y-axis represents the relative abundance.

5.4.3 Identification of the nsp7 N-proximal domain as a target of nsp9-mediated nucleotidylation

To delineate possible targets of the nsp9 nucleotidylation activity within nsp6-7-8, we decided to produce recombinant forms of nsp7 and two proteins, nsp7α and nsp7β, that are released from nsp7 by nsp4-mediated proteolytic cleavage (van Aken et al., 2006). The cleavage site is conserved among arteriviruses and mutagenesis of this site in a molecular clone of EAV was shown to abolish viral replication. Structural analyses of

EAV and PRRSV of nsp7α did not reveal specific functional motifs although nsp7/nsp7α was shown previously to be essential for viral RNA replication (Chen et al., 2017; Li et al., 2015; Manolaridis et al., 2011; Zhang et al., 2013). Chen *et al.* reported that nsp7α interacts with nsp9 (Chen et al., 2017).The EAV nsp7, nsp7α and nsp7β proteins were expressed in *E. coli* and purified as described in Material and Methods. To assess the overall purity and molecular size of the proteins, equal amounts of protein were analyzed by Tris-tricine-PAGE. The apparent molecular masses of nsp7 (26 kDa) nsp7α (13 kDa), and nsp7β (13 kDa) in SDS-polyacrylamide gels corresponded well to the calculated masses of these proteins. In Figure 37, the purified nsp7, nsp7α and nsp7β proteins used for subsequent analyses are shown.

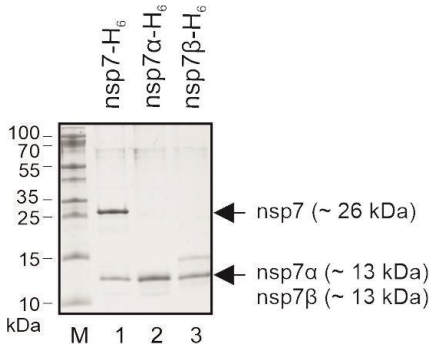


Figure 37: Tris-tricine PAGE analysis of bacterially expressed and purified EAV nsp7, nsp7α and nsp7β proteins: C-terminally His-tagged nsp7, nsp7α, and nsp7β were expressed in TB1[pCGI] cells transformed with the appropriate pASK-Ub-EAV-nsp-H_6 construct. The proteins were purified by Ni-IMAC affinity chromatography, anion exchange chromatography and size-exclusion chromatography. 2 µg of nsp7, nsp7α and nsp7β were analyzed in a 12% Tris-tricine-PAGE. M – prestained marker proteins (PageRuler). Positions of the respective proteins are indicated on the left with calculated molecular masses.

Nsp7, nsp7α and nsp7β were used in nucleotidylation assays with nsp9 in buffer containing 20 mM HEPES-KOH/pH 8, 30 mM KCl, 0.2 % [v/v] Triton X-100, 5 % [v/v] glycerol, β-mercaptoethanol, 1mM DTT, and 1 mM $MnCl_2$ or $MgCl_2$.

As shown in Figure 38 A and B, there was no evidence for nsp9-mediated adenylylation or cytidylation of nsp7, nsp7α or nsp7β. By contrast, nsp9 was found to produce guanylylated forms of nsp7 and nsp7α (Figure 38 B, lanes 15, 16, 18). Also, the presence

of nsp7 appeared to stimulate the self-guanylylation activity of nsp9 (Figure 38 B, lane 15). Protein guanylylation activity was dependent on Mn^{2+}, while Mg^{2+} did not support this activity, confirming and extending an earlier study in which EAV nsp9 self-nucleotidylation activity was shown to require the presence of Mn^{2+} ions (Lehmann et al., 2015). In the reaction with nsp7β, no labeled product was detected, suggesting that the nsp7α region of nsp7 is the target for nsp9-mediated guanylylation. Very similar results were obtained in the uridylylation assay (Figure 38 D). Thus, for example, nsp9 was found to uridylate nsp7 and nsp7α, while there was no evidence for uridylylation of nsp7β. Also, self-uridylylation of nsp9 was increased in the presence of nsp7 and protein uridylylation activity required the presence of Mn^{2+}, while Mg^{2+} did not support this activity. (Figure 38 D; lanes 19-27). The presence of nsp7β in the reaction containing nsp7α and nsp9 did not affect guanylylation and uridylylation activity (Figure 38 B, lane 18, and D, lane 36). Taken together, the data lead me to conclude that the nsp7α region of nsp7 can act as an acceptor for nsp9-mediated nucleotidylation. The structural basis for the preference for GTP and UTP over other nucleotides remains to be determined. Guanylylation and uridylylation activities of nsp9 are equally efficient and both reactions share the requirement of Mn^{2+} as catalytic metal ion.

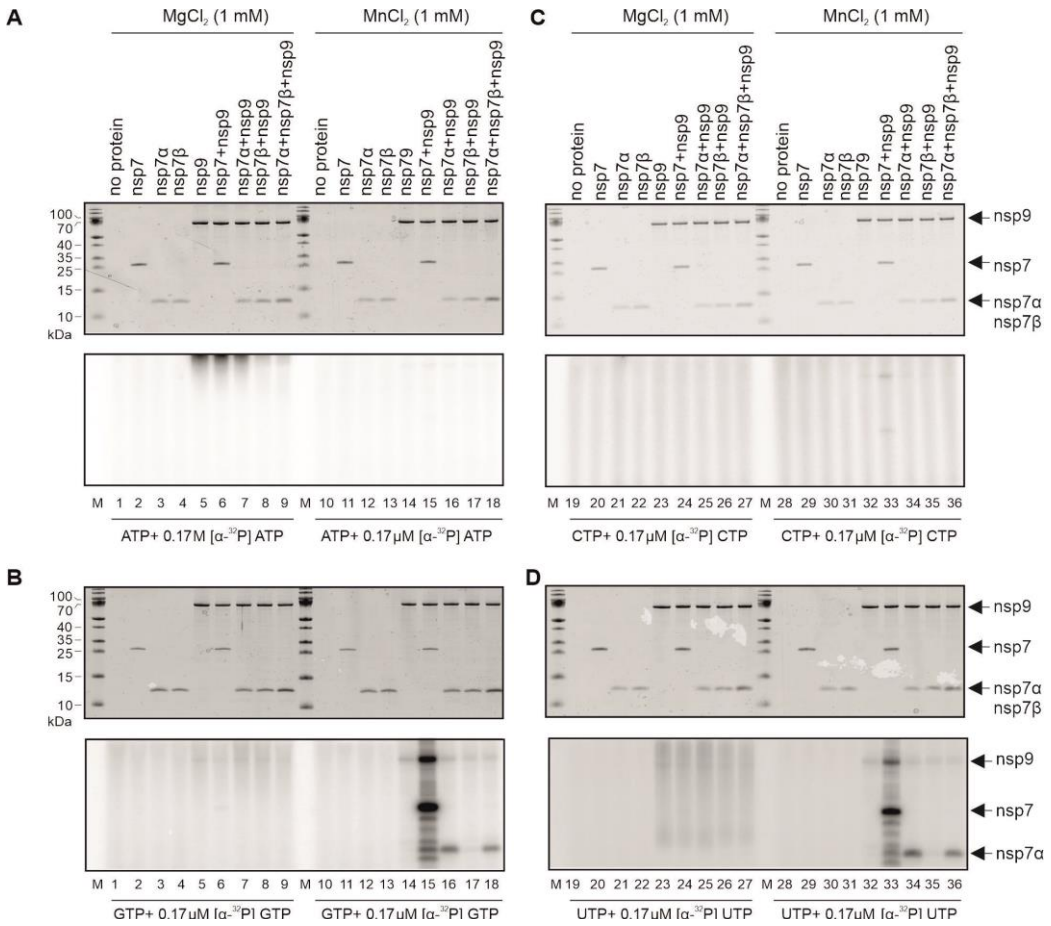


Figure 38: Protein substrate, nucleotide co-substrate and metal ion requirements of EAV nsp9-mediated nucleotidylation activites. Nsp9-mediated protein nucleotidylation was assessed using nsp7, nsp7α and nsp7β as potential substrates. Reaction products were separated in 12 % Tris-tricine polyacrylamide gels and stained with Coomassie brilliant blue (top panels). Radiolabeled products were analyzed by phosphorimaging of the SDS-polyacrylamide gels (bottom panels). **(A)** The adenylylation assay was performed in the presence of 25 µM ATP, 0.17 µM [α-^{32}P]-ATP and 1 mM $MgCl_2$ (lanes 1-9) or $MnCl_2$ (lanes 10-18). **(B)** The guanylylation assay was performed in the presence of 25 µM GTP, 0.17 µM [α-^{32}P]-GTP and 1 mM $MgCl_2$ (lanes 1-9) or $MnCl_2$ (lanes 10-18). **(C)** The cytidylation assay was performed in the presence of 25 µM CTP, 1.7µM of [α-^{32}P]-CTP and 1 mM $MgCl_2$ (lanes 19-27) or $MnCl_2$ (lanes 28-36). **(D)** The uridylylation assay was performed in the presence of UTP, 1.7 µM [α-^{32}P]-UTP and 1 mM $MgCl_2$ (lanes 19-27) or $MnCl_2$ (lanes 28-36). Assays were performed using 1 µM of nsp7, nsp7α, nsp7β, nsp9 or a 1:1 ratio of nsp7+nsp9, nsp7α+nsp9, nsp7β+nsp9 and nsp7α+nsp7β+nsp9 (each at 1 µM) as indicated on the top. Lanes 1, 10, 19, 28 - no protein was included in the reaction. The position of individual proteins in the Coomassie-stained gels (top) and radiolabeld proteins in the autoradiograms (bottom) are indicated to the right. Protein size markers (in kDa) are indicated to the left.

To verify the identities of the [α-^{32}P]-UMP/GMP-labeled proteins observed by SDS-PAGE and autoradiography, we performed mass spectrometry analysis of the nsp7-UMP and nsp7-GMP complexes generated by nsp9-mediated nucleotidylation. To this end, uridylylation and guanylylation assays were performed using EAV nsp7 and nsp9

(each at 10 µM) in the presence of 1 mM $MnCl_2$ and 500 µM UTP or GTP, along with control reactions in which specific compounds were omitted. Following incubation at 30°C for 60 min, the reaction mixtures were were desalted and subjected to mass spectrometry as described above for the nsp6-7-8 and nsp9 complexes and, in more detail, in Materials and Methods. To obtain the mass spectra of non-nucleotidylated proteins, nsp7 and nsp7+nsp9 were used as controls. As shown in Figure 39 A and B, a single mass species was observed for the apo form of nsp7 (26463.0020 Da) and and the same mass was detected for nsp7 if mixed with nsp9. The mass spectrometry failed to produce spectra for nsp9.

As shown in Figure 39 C, molecular masses for the nsp7 apo form and the nsp7-UMP adduct were obtained for the sample containing the nsp9-mediated uridylylation reaction (nsp7, nsp9, UMP). The apo form of nsp7 was less abundant. The difference between the two major mass species of 26769.0020 and 26463.0020 was 306 Daltons, which is precisely the mass of UMP, suggesting that nsp9 transfers a single UMP to nsp7. Similar results were obtained for the guanylylation reaction. Here, a slightly less complete conversion of nsp7 (26463.0020 Da) to nsp7-GMP (26808.0020 Da) was confirmed. The mass difference between the two major mass species was 345 Da, which corresponds to the mass of GMP (Figure 39 D) and confirms that nsp7 was modified by guanylylation at a single position, suggesting that nsp9 nucleotidylation targets a specific site in nsp7.

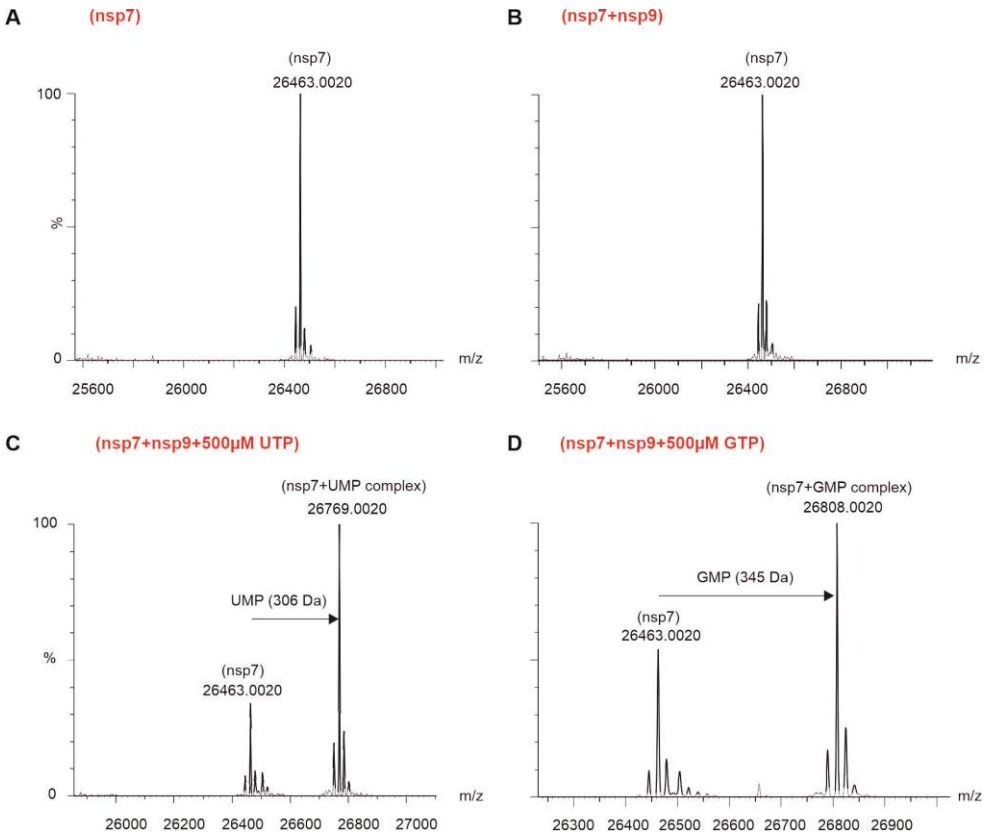


Figure 39: Mass spectroscopy reveals nsp9-mediated (mono)uridylylation and (mono)guanylylation of nsp7. Shown are deconvoluted intact protein mass spectra obtained for EAV nsp7 (in its unmodified and nucleotidylated form) in reactions containing nsp9 and UTP or GTP. **(A)** nsp7, **(B)** nsp7 and nsp9, **(C)** nsp7, nsp9, and UTP, **(D)** nsp7, nsp9 and GTP. Nucleotidylation assays and control reactions were performed with the indicated proteins (each at 10 µM) and nucleotides (500 µM) in buffer containing 20 mM HEPES-KOH/pH 8.0, 30 mM KCl, 5 % [v/v] glycerol, 1.5 mM β-mercaptoethanol, 1mM DTT, and 1 mM $MnCl_2$ at 30ºC for 1 hour. Samples were desalted and analyzed using a Synapt G2Si mass spectrometer. Intact protein masses relating to nsp7 and nsp7-NMP complexes are indicated. Also shown are the mass differences between proteins and protein-NMP complexes. The X-axis represents mass-to-charge (m/z) values and the Y-asis represents the relative abundance.

5.5 Conserved residues of the NiRAN domain are essential for nucleotidylation activity, but not RdRp domain of EAV nsp9.

The study by Lehmann *et al.* identified a number of residues conserved in NiRAN domains of viruses representing different families of the order *Nidovirales* (Lehmann et al., 2015). To assess the roles of these residues in protein nucleotidylation, we perfomed a mutational analysis. The expression plasmid was modified by PCR-based mutagenesis to allow the expression of mutant forms of EAV nsp9 in which the conserved NiRAN

residues K94, R124, S129A, D132, D165, F166 residues and the non-conserved NiRAN residue K106 and the conserved RdRp D445 residue, respectively, were replaced with alanine (Figure 40). The mutant proteins were expressed in *E. coli* and purified to near homogeneity using affinity- and anion-exchange chromatography. There were no obvious differences in expression, purification or stability between the wild-type and mutant proteins.

In control reactions containing no protein (Figure 40; lanes 1, 13) or nsp7 alone, no radiolabeled products were observed (Figure 40; lanes 2, 14). Reactions containing nsp9 alone revealed minimal self-uridylylation and self-guanylylation as judged by the presence of trace amounts of a radiolabeled protein with a size similar to that of nsp9 (Figure 40; lanes 3, 15). Radiolabeled forms of nsp7 were readily detected when nsp7 was mixed with nsp9 in the presence of [α-^{32}P]-GTP or [α-^{32}P]-UTP. As shown above (Figure 38), nsp9 self-nucleotidylation activity was increased in the presence of nsp7 (Figure 40, lanes 4 and 16), possibly indicating structural changes in nsp9 when interacting with its substrate protein nsp7. The assays with mutant proteins were performed using nsp7 as a substrate and GTP and UTP as nucleotide cosubstrates. As shown in Figure 40, no nsp9 self-nucleotidylation or nsp7 nucleotidylation activities were detected for nsp9_K94A and nsp9_D165A, while nsp9_K106A, nsp9_R124A and the RdRp mutant nsp9_D445A retained their activities with both UTP and GTP. Nsp9_S129A, nsp9_D132A, and nsp9_F166A had moderately or greatly reduced activities. Together, these data confirm the functional importance of most nidovirus-wide conserved residues in this *in vitro* nucleotidylation assay.

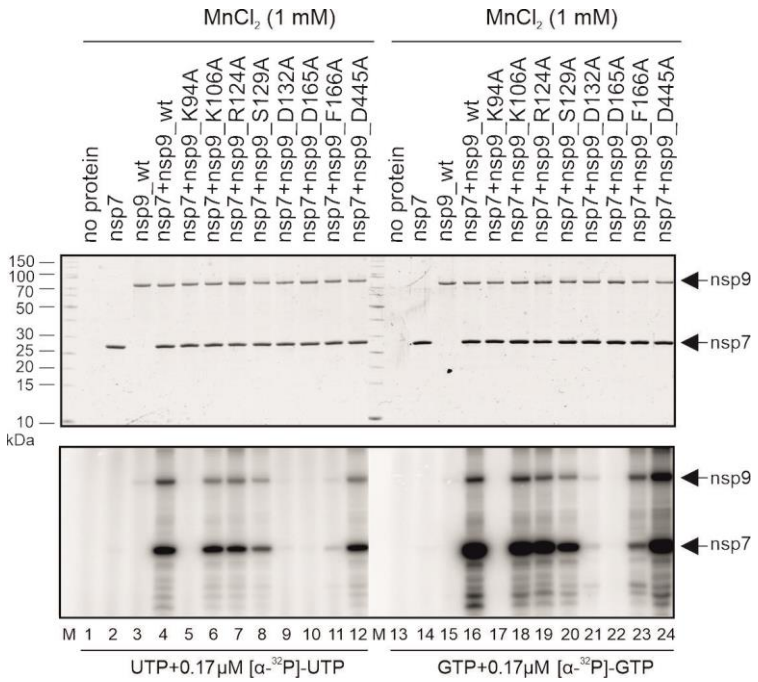


Figure 40: Nucleotide transferase assay using EAV nsp9 wild-type and mutant proteins in the presence of the acceptor protein nsp7. Conserved residues identified in the NiRAN domains of arteriviruses and other nidovirus families (Lehmann et al., 2015) were replaced with alanine. As controls, nonconserved or less conserved residues were replaced. Also, a mutant nsp9 protein, D445A, in which an RdRp active-site residue in motif A was substituted was included in the analysis. Reactions products were separated on a 12 % Tris-tricine polyacrylamide gel and stained with Coomassie brilliant blue (top panel). Radiolabeled products were analyzed by phosphorimaging of the gel (lower panel). **(A)** Lanes 1-12, reactions in the presence of 25 µM UTP, 0.17µM [α-^{32}P]-UTP. **(B)** Lanes 13-24, reactions in the presence of 25 µM GTP, 0.17 µM [α-^{32}P]-GTP. Nucleotidylation reactions were performed with the indicated proteins (each at 1 µM) or without protein (lanes 1 and13). The positions of nsp7 and nsp9 are indicated to the right of both the Commassie-stained Tris-tricine polyacrylamide gel (top) and the autoradiogram (bottom). Protein size markers (in kDa) are indicated to the left.

6 Discussion

6.1 Goal I

Like other plus-strand RNA viruses, nidoviruses carry an RdRp enzyme which directs viral RNA synthesis in combination with other viral and cellular proteins (van Hemert et al., 2008b) and is the key enzyme of the viral replication/transcription complex (RTC). The coronaviral RdRp was identified by comparative sequence analysis of the first complete coronavirus genome sequence and found to contain an SDD sequence at a position that, in many other viral RdRps, is occupied by GDD (den Boon et al., 1991; Gorbalenya et al., 1989b; Gorbalenya et al., 2002; van Dinten et al., 1999; van Dinten et al., 1996; Xu et al., 2003). Even though the common motifs of the nidoviral RdRp have been identified for some time, the mechanism or strategy of initiation of RNA synthesis is still poorly understood.

So far, two well defined replication initiation strategies have been discovered for viral RdRps, *de novo* initiation and primer-dependent RNA synthesis. During *de novo* initiation, the flavivirus RdRp for example initiates RNA synthesis by forming a phosphodiester bond between the first initiation NTP and the second incoming NTP (Chen and Patton, 2000; Kao et al., 1999; Kao et al., 2001). The RdRp then utilizes this dinucleotide for the elongation of the RNA. The second mechanism (primer-dependent initiation) requires a nucleic acid or protein to initiate RNA synthesis. The *Picornaviridae* family employs a protein-primed initiation mechanism. For example, in poliovirus, a tyrosine of the VPg protein provides the hydroxyl group for the formation of a phosphodiester bond with the first nucleotide to elongate the RNA (Paul et al., 2000; Paul et al., 1998). Interestingly, influenza A virus uses 5'-capped oligonucleotides that are released from cellular mRNAs by endonucleolytic cleavage as primers for the initiation of viral RNA synthesis (Plotch et al., 1981).

In previous studies, a range of functions have been proposed to be associated with coronavirus nsp8, nsp7+8 and nsp12: (i) primase activity was proposed for nsp8, generating primers for the main RdRp (Imbert et al., 2006; Zhai et al., 2005), (ii) primer-independent and primer-dependent polymerase activity was proposed for SARS-CoV and FCoV nsp7+8 (te Velthuis et al., 2012; Xiao et al., 2012), (iii) primer-dependent polymerase activity was proposed for SARS-CoV nsp12 in the presence of Mg^{2+} (te Velthuis et al., 2010a), and (iv) *de novo* initiation activity was reported for a nsp7-L-nsp8+nsp12 complex in the presence Mg^{2+}, which however failed to reproduce the previously proposed primase activity for nsp8 (Subissi et al., 2014). Overall, there has been little reproducibility concerning the role of the central RdRp enzyme and/or the co-factors nsp8 and nsp(7+8) in RNA polymerase activity *in vitro*.

Therefore, this study aimed to identify and characterize the role of nsp7, nsp8, nsp(7+8) complex and nsp12 from both alpha- and betacoronaviruses. The rationale behind this study was: (i) to identify similarities and differences in the architecture of the nsp(7+8) complex of viruses representing different coronavirus species and genera, (ii) to address the precise role of nsp7, nsp8, nsp(7+8), nsp12, nsp(7+8)+nsp12 proteins in RNA elongation and *de novo* initiation, and (iii) to analyze the effect of metal ions on the replicase complex nucleotide selectivity and fidelity.

6.1.1 Generation of an active nsp(7+8) complex and multimeric states of alpha- and betacoronaviral nsp(7+8) complexes

Previously, protein purification strategies used to determine structures and characterize biochemical properties were generally based on recombinant nsp7, nsp8 and/or nsp12 proteins containing exogenous tags at either the N- or C-terminus of the proteins. In this study, we aimed at obtaining tag-free, biologically active complexes with authentic N- and C-termini by exploiting the proteolytic activity of nsp5. This approach allowed us

to further characterize the multimeric states of nsp(7+8) complexes and their biochemical properties and eliminated potential effects of artificial sequences on the structure and/or function(s) of the proteins under study.

Previous structural studies on SARS-CoV and SARS-CoV-2 nsp7 and nsp8 reported interactions resulting in the formation of hexadecamers (8:8) and heterodimers (1:1) (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; te Velthuis et al., 2012; Wang et al., 2020; Zhai et al., 2005), while FIPV nsp7 and nsp8 was reported to form a heterotrimer (2:1) (Xiao et al., 2012). To validate these observations, the multimerization states of several pre-assembled coronavirus nsp(7+8) complexes were analyzed by chemical crosslinking, analytical size-exclusion chromatography and native mass spectrometry (Figures 18, 19, 20 and supplementary Figure S2). Unlike the previously observed hexadecameric structure of SARS-CoV nsp7 and nsp8, the pre-assembled nsp(7+8) complexes from HCoV-229E and SARS-CoV produced in the current study formed various multimeric states including heterodimeric (1:1), heterotrimeric (1:2 and 2:1) and heterotetrametric (2:2) complexes (Figure 18). In agreement with the heterotrimeric structure of FCoV nsp7+8 and in contrast to the complexes produced by the HCoV-229E and SARS-CoV proteins, there was no evidence for the formation of a tetrameric nsp(7+8) complex in the case of FIPV (Figure 18, 19, 20 and supplementary Figure S2) (Xiao et al., 2012). Instead, the FIPV nsp7 and nsp8 were shown to form heterotrimeric (2:1) rather than heterotetrameric (2:2) complexes. The formation of such heterotrimeric (and heterodimeric) nsp(7+8) complexes could also be demonstrated for a very closely related virus, TGEV, using native mass spectrometry (Supplementary Figure S2). In contrast to the hexadecameric nsp7+8 supercomplex reported previously for SARS-CoV (Zhai et al., 2005), very strong evidence for the (predominant) formation of a stable heterotetrameric complex

was obtained by different approaches, including size-exclusion chromatography, chemical crosslinking experiments and native mass spectrometry (Figures 18, 20 and supplementary Figure S2). These results are in line with a recently published study of complexes produced by nsp5-mediated processing of the SARS-CoV nsp7-8-9-10 polyprotein (Krichel et al., 2020) and the crystal structure reported for a SARS-CoV-2 nsp7+8 heterotetramer (Konkolova et al., 2020). For MERS-CoV, which, like SARS-CoV, is a member of the genus *Betacoronavirus*, the nsp(7+8) complex was shown to form heterodimeric, heterotrimeric and heterotetrameric species using native mass spectrometric analysis (supplementary Figure S2). Analysis of the multimeric state of the HCoV-229E nsp(7+8) complex in solution by analytical size-exclusion chromatography suggested the formation of tetrameric and larger, potentially hexameric, complexes (Figure 19), while native mass spectrometry suggested the formation of (some) heterodimeric (1:1) and heterotrimeric (2:1) complexes and a clearly predominant heterotetrametric (2:2) species for HCoV-229E and PEDV (Supplementary Figure S2). Based on the combined crosslinking, size-exclusion chromatography and native mass spectrometry results it seems reasonable to conclude that, upon processing of the polyprotein, a stable heterodimeric nsp(7+8) subcomplex is formed that is subsequently used as building block for the formation of larger assemblies, such as heterotrimers (primarily in FIPV and TGEV, and less frequently in HCoV-229E, PEDV and MERS-CoV) or heterotetramers, which have been identified for HCoV-229E, PEDV, SARS-CoV and MERS-CoV complexes. The (differentially assembled) nsp(7+8)/nsp12 trimeric and nsp(7+8)/8/12 tetrameric complexes in recently published polymerase structures suggest specific functions for nsp8 and nsp(7+8) that bind to different surfaces of the nsp12 protein (Gao et al., 2020; Hillen et al., 2020), while no evidence for a tetrameric nsp(7+8) complex was obtained in these

studies. However, given that those structural studies did not employ preformed nsp(7+8) complexes produced by proteolytic processing from polyprotein precursors, the formation of heterotetrameric nsp(7+8) complexes that interact with nsp12 remains a likely scenario. The structures determined recently for heterodimeric nsp7+8 complexes of SARS-CoV-2 and SARS-CoV in conjunction with nsp12 were most likely based on a head-to-head interaction of nsp7 and nsp8, which appears to be the most conserved form of interaction (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Wang et al., 2020).

6.1.2 Nsp(7+8) and nsp12 are essential for RNA elongation

Another important part of the study focused on the roles of nsp7 and nsp8 in supporting the elongation activity and *de novo* polymerase activity of nsp12. Initially, HCoV-229E nsp7, nsp8, the pre-assembled nsp(7+8) complexes with authentic N- and C-termini and nsp12 were purified (Figure 21 and Supplementary Figure S3) and characterized. To assess the functional roles of the purified proteins in RNA elongation (i.e., primer-dependent polymerase) activity, a partially double-stranded RNA substrate was used (Figure 22). Neither nsp8 nor the nsp(7+8) complex displayed primer-dependent polymerase activity in the presence of either Mg^{2+} or Mn^{2+}, which is in agreement with the observations made by Tvaragová *et al*. (2019) and in contrast to reports from te Velthuis *et al*. and Xiao *et al*. (te Velthuis et al., 2012; Tvarogova et al., 2019; Xiao et al., 2012) (Figure 22). While previous reports suggested that SARS-CoV nsp12 alone is inactive (Subissi et al., 2014), we observed a low (but clearly detectable) polymerase activity when HCoV-229E nsp12 was used in primer-dependent polymerase assays using a template with a stretch of CU nucleotides and Mg^{2+} or Mn^{2+} being present in the reaction buffer. However, the strongest primer-dependent polymerase activity was observed if the pre-assembled nsp(7+8) complex was used in combination with nsp12.

RNA product formation was significantly reduced if separately purified nsp7+nsp8 was used in combination with nsp12, suggesting that the pre-assembled nsp(7+8) complex adopts a structure that supports the appropriate interaction of nsp12 with the primed RNA substrate, thereby facilitating primer elongation and processive RNA synthesis (Figure 22). Strikingly, in the presence of Mn^{2+}, nsp8 (alone) in combination with nsp12 resulted in relatively efficient RNA synthesis giving rise to a slightly shorter product compared to the that obtained with the nsp(7+8) complex. The reason for the superior activity of nsp8+nsp12 in the presence of Mn^{2+} (compared to Mg^{2+}) is not clear but may be due to subtle conformational changes in nsp12 that make it less dependent on interactions with the nsp(7+8) complex. Stimulatory effects of Mn^{2+} on polymerase activity have also been reported for other viral polymerases and have been suggested to be linked to relaxed nucleotide specificity and different binding strengths of the two ions to the phosphodiester backbone of nucleic acids (Arnold et al., 1999; Brooks and Andersen, 1978; Cirino et al., 1995; Huang et al., 1997; Polyanichko et al., 2004; Tabor and Richardson, 1989). Conformational changes upon interaction with Mn^{2+} that affect enzymatic activity have also been reported for the hepatitis C virus NS5B polymerase (Bougie et al., 2003; Ranjith-Kumar et al., 2002b). Based on a recent cryo-EM structure analysis of a tetrameric SARS-CoV-2 nsp7+8+8+12 complex with RNA, it has been proposed that, during RNA synthesis, the long N-terminal helical extensions of nsp8 slide along the exiting RNA, thereby preventing premature dissociation of nsp12 during catalysis (Hillen et al., 2020). It remains to be studied if Mn^{2+} ions help retain the processivity factor function of nsp8 under experimental *in vitro* conditions in which no functional nsp(7+8) complex is being formed.

To determine whether the different stoichiometries observed for nsp(7+8) complexes of different coronaviruses correlate with different elongation activities, primer-dependent

polymerase activity assays were performed with HCoV-229E, FIPV and SARS-CoV nsp(7+8) complexes and the cognate nsp12 using partially double-stranded RNA substrates with homo- or heteropolymeric single-stranded template regions in the presence of Mg^{2+} (Figure 25 C). The elongation activities were found to be equally efficient for all three coronavirus replicase complexes irrespective of their nsp(7+8) stoichiometry. Taken together, the data consistently show that nsp12-mediated RNA synthesis is greatly simulated by a pre-formed nsp(7+8) complex. Under the *in vitro* conditions used in this study, no major differences were observed for tetrameric versus trimeric nsp(7+8) complexes regarding their potential to support nsp12 polymerase activity.

This conclusion is supported by data shown in Supplementary Figure S4 A where primer-dependent polymerase activities of alpha- and betacoronavirus nsp12 proteins along with nsp(7+8) complexes were analyzed using a primed RNA with a heteropolymeric template strand. Irrespective of the complexation state of the pre-assembled nsp(7+8) complexes (predominantly 2:1 or 2:2), similar activities were obtained for all coronavirus nsp12s tested (Supplementary Figure S4 B). In another experiment, we tested if a heterologous coronavirus nsp(7+8) complex can substitute the function of the cognate nsp(7+8) complex. To do this, we used the nsp(7+8) complexes of TGEV, HCoV-229E, PEDV, FIPV, SARS-CoV or MERS-CoV in combination with TGEV nsp12 in a primer-dependent polymerase activity assay (Supplementary Figure S5). As expected, TGEV nsp12 was active in the presence of its cognate nsp(7+8) complex. Interestingly, TGEV nsp12 retained its activity if the FIPV nsp(7+8) complex was provided instead of the cognate TGEV nsp(7+8) complex. By contrast, no activity was detectable if the TGEV nsp12 was used in combination with any other coronavirus nsp(7+8) complex (Supplementary Figure S5 B). Once again, the

data demonstrate the critical role of the nsp(7+8) complex for nsp12 polymerase activity. More imporatantly, the data show that nsp12 interacts with the nsp(7+8) complex and that these interactions are (only) conserved among very closely related viruses, such as TGEV and FIPV, which both belong to the same virus species called *Alphacoronavirus 1*. At this stage, however, it cannot be excluded that TGEV and FIPV nsp12 polymerases share a preference for nsp(7+8) complexes with a heterotrimeric (rather than heterotetrameric) architecture.

6.1.3 A nsp(7+8) complex, nsp12 and Mn^{2+} are essential for the *de novo* initiation of RNA synthesis

Apart from elongation activity, the roles of nsp7, nsp8, nsp(7+8), nsp12 and combinations of nsp7+nsp12, nsp8+nsp12, nsp(7+8)+nsp12 were investigated for the initiation of RNA synthesis in the absence of an RNA primer. Using single-stranded 28- or 38-nucleotide-long RNAs with sequences corresponding to the 3'-end of the HCoV-229E genome (Figure 24 A) as substrates, RNA synthesis resulting in products of the expected length were observed for the nsp(7+8)+nsp12 complex if Mn^{2+} ions were included in the reaction mixture (Figure 24 B and C). None of the individual nsp7, nsp8, nsp12 proteins or the nsp(7+8) complex promoted RNA synthesis, and no primase activity was detected for nsp8. In the available cryo-EM structures for SARS-CoV-2 nsp7+8+12 complexes, the nsp7+8 complex was found to be located far from the nsp12 active site in protein-RNA complexes including a primed (partially double-stranded) RNA, suggesting that RNA synthesis is catalyzed by nsp12 with no direct involvement of nsp8 in this case (Gao et al., 2020; Hillen et al., 2020). It remains to be investigated by additional structural studies if nsp(7+8) has a distinct role in the *de novo* initiation process of RNA synthesis or, following initiation, contributes to elongation by its presumed role as a processivity factor. Our observed *de novo* initiation activity stands

in contrast to the results obtained for the SARS-CoV nsp7-L-8+nsp12 complex showing *de novo* initiation in the presence of Mg^{2+} as the sole catalytic ion (Subissi et al., 2014) and the proposed *de novo* activity for the isolated SARS-CoV nsp7+8(te Velthuis et al., 2012) and nsp12 proteins (Ahn et al., 2012; te Velthuis et al., 2012). The data presented here clearly show that the nsp(7+8)+nsp12 complex initiated RNA synthesis at the very 3'-terminus of the KR07 -b substrate (terminal cytidylate of template RNA T1) (Figure 24 B) while the enzyme complex initiated RNA synthesis from the same position but also from within the oligo(A) stretch present in the KR07 A_{10} -b template RNA (Figure 24 C). A possible explanation for why the nsp(7+8)+nsp12 complex initiated RNA synthesis at different sites could be that, under these *in vitro* conditions, the start site selection is less contrained. Also, it should be noted that, in infected cells, minus-strand RNA synthesis starts with the synthesis of a stretch of uridylates probably templated by the poly(A) tail (Hofmann and Brian, 1991). The data show that Mn^{2+} and nsp12 are essential but not sufficient for the *de novo* RNA synthesis. Also required is a "naturally obtained" pre-assembled nsp(7+8) complex. Although we were able to provide convincing biochemical evidence for *de novo* initiation of RNA synthesis by the coronavirus nsp(7+8)+nsp12 replicase complex, similar to activities reported for flavivirus and cystovirus polymerases, there is no structural information yet on how this initiation occurs (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Wang et al., 2020). The coronaviral polymerase structure shows a resemblance to the picornaviral polymerase (3Dpol), which uses a protein primer (VPg-UpUp) for the initiation of RNA synthesis rather than *de novo* initiation. In this context, it is worth noting that the coronavirus nsp9 has recently been shown to be NMPylated by the NiRAN domain of nsp12 and speculated to potentially act as a protein primer to initiate coronavirus RNA synthesis (Slanina et al., 2021). Clearly, further studies are required

to firmly establish the mechanisms and factors involved in the initiation of RNA synthesis in the context of coronavirus infection in cell culture.

An *in vitro* optimization of the role of the catalytic ions Mg^{2+} or Mn^{2+} in *de novo* initiation of RNA synthesis revealed that Mn^{2+} is the key player for *de novo* RNA synthesis by the HCoV-229E nsp(7+8)+nsp12 complex. Mg^{2+} did not stimulate RNA synthesis if short synthetic RNAs were used as templates. Assays using HCoV-229E, FIPV and SARS-CoV nsp(7+8)+nsp12 complexes with a 3'-biotinylated RNA or an unmodified RNA substrates containing a 3'-OH group (Figure 25 A) in the presence of Mn^{2+} showed initiation of RNA synthesis by a *de novo* mechanism, generating predominantly a template-sized product with trace amounts of backprimed products if the 3'-biotinylated KR07-b and KR07 A_{10}-b substrates were used (Figure 25 B). Using the 3'-unmodified KR07 and KR07 A_{10} substrates, the replicase complexes were able to initiate RNA synthesis *de novo*, thereby generating a small amount of template-sized products, while the majority of products were longer, indicating that RNA synthesis involved backpriming due to basepairing of the template's 3'-end with internal sequences (hairpin conformation) (Ackermann and Padmanabhan, 2001; Laurila et al., 2005; Luo et al., 2000). Apparently, the 3'-region hybridized internally to an upstream sequence and was then elongated. The backpriming or copy-back mechanism is generally considered common to *in vitro* experimental conditions while *de novo* initiation of replication is the true mechanism occuring in flavi- and cystovirus-infected cells (Ackermann and Padmanabhan, 2001; Butcher et al., 2001; Laurila et al., 2002; Noton et al., 2014). Similar to what we have shown for the nsp(7+8)+nsp12 complexes of HCoV-229E, FIPV and SARS-CoV using 3'-biotinylated and unmodified RNA templates, both template-size products and hairpin products were previously identified in studies using HCV RNA (Ackermann and Padmanabhan, 2001; Luo et al., 2000;

Nomaguchi et al., 2003; Nomaguchi et al., 2004). If the 3'-OH of the substrate RNA ws modified by a chain terminator, cordycepin, the formation of back-primed product was prevented without affecting the formation of the template-sized product. Based on these results, the authors concluded that *de novo* and copy-back mechanisms were responsible for generating these two products (Luo et al., 2000). In previous studies, Mn^{2+} (as opposed to Mg^{2+}) ions have been proposed to affect polymerase activities specific ways, for example, by stimulating or modulating the catalytic and non-catalytic site(s) of the polymerase by inducing structural rearrangements or supporting additional activities like terminal transferase activity and also by activating *de novo* initiation more efficiently than 3'-elongation (Arnold et al., 1999; Butcher et al., 2001; Ng et al., 2004; Pinto et al., 1979; Poranen et al., 2008b; Ranjith-Kumar et al., 2002a; Ranjith-Kumar et al., 2002b; Shim et al., 2002; Tabor and Richardson, 1989; Wright et al., 2012). As demonstrated for the coronavirus nsp12-mediated Mn^{2+}-dependent initiation of RNA synthesis, the polymerases of positive-stranded RNA viruses from the *Flaviviridae* family (HCV, BVDV, EV, CSFV, WNV) and negative-stranded RNA viruses such as VSV and RSV have been shown to synthesize RNA (*in vitro*) by *de novo* and backpriming mechanisms (Noton et al., 2014; Ranjith-Kumar et al., 2002a; Ranjith-Kumar et al., 2002b; Yi et al., 2003; Yu et al., 2007). However, flaviviruses do not seem to require any additional virally encoded or host factors to initiate RNA synthesis while VSV and RSV require phosphoprotein P as a cofactor for the polymerase protein L (Morin et al., 2012; Noton et al., 2014). It has been proposed that, within the family *Flaviviridae*, HCV, BVDV and GBV RdRps can employ both Mg^{2+} and Mn^{2+} metal ions for the initiation of RNA synthesis (Luo et al., 2000; Ranjith-Kumar et al., 2002a). In this regard, we compared alpha- and betacoronavirus replicase complexes using both catalytic ions, Mg^{2+} and Mn^{2+}, and the ssRNA KR07 A_4-b as a substrate (Supplementary

Figure 4 C). Despite differences in their efficiency of activities and multiple initiation sites both alpha- and betacoronavirus replicase complexes similarly catalyzed *de novo*-initiated products by incorporating UMP as initial nucleotide templated by (one of) the 3'-terminal adenylate(s) of the substrate in the presence of Mn^{2+} (Supplementary Figure S4 D). Similar to what we observed for the HCoV-229E nsp(7+8)+nsp12 complex, the replicase complexes of PEDV, TGEV, FIPV, SARS-CoV and MERS-CoV did not exhibit activity on this template in the presence of Mg^{2+} (Supplementary Figure S4 D). Our comparative analysis also indicates that, irrespective of (i) the quarternary assembly of the nsp(7+8) complex (tetramer/trimer) and (ii) the respective coronavirus (species/genus) used, all these replicase complexes were able to initiate RNA synthesis *de novo* (that is, without a primer) in the presence of Mn^{2+}.

With regard to data reported previously by Subissi *et al.* (Subissi et al., 2014), one or more possibilities may account for the partly discrepant observations made. First, the previous study used significantly longer RNA substrates compared to those used in this work. Secondary structures presumably present in these longer RNAs might have either induced elongation by the RdRp or inhibited *de novo* initiation.

Second, we observed similar patterns of *de novo* initiation activity with a large number of short RNA substrates (only a few representative substrates were presented in this thesis).

Third, in the previous work, a weak initiation of RNA synthesis was shown with Mg^{2+}. In our assays, the enzyme complexes of both alpha- and betacoronaviruses consistently initiated RNA synthesis (*do novo*) only with Mn^{2+} but not with Mg^{2+}.

Fourth, in the previous work showing weak *de novo* RNA synthesis in the presence of nsp12, a physical linker had been inserted between nsp7 and nsp8. It is not known how this artificial protein construct affects *de novo* initiation. In our study, we show efficient

RNA synthesis with the most naturally obtained, pre-assembled complex in the presence of nsp12 using short synthetic RNA substrates.

GST-pulldown, strep-tag interaction experiments and structural studied provided convincing evidence that nsp8, nsp7L8 and nsp7+8 physically interact with nsp12 (Gao et al., 2020; Hillen et al., 2020; Imbert et al., 2008; Kirchdoerfer and Ward, 2019; Subissi et al., 2014; Tan et al., 2018; Wang et al., 2020). The assembly of nsp7 with nsp8 is critical and, based on crystal structures of nsp7+8 of SARS-CoV and FIPV, nsp7 was suggested to serve as mortar to stabilize the nsp7+8 complex and to enhance the RNA-binding affinity of the nsp8 protein (te Velthuis et al., 2012). This is in line with nsp8 being the critical molecule for the interaction with RNA (Figure 26). Tvarogová *et al*. reported for HCoV-229E that an alanine replacement of one conserved lysine in nsp8 (K3771A) abolished both RNA binding and TATase activity of the nsp8 protein and also abolished viral replication (Tvarogova et al., 2019). The N-terminal extensions of nsp8 with their alternative conformations in crystal structures of nsp7+nsp8 complexes (Xiao et al., 2012; Zhai et al., 2005) are mobile, and also in a larger complex with RdRp (Gao et al., 2020; Kirchdoerfer and Ward, 2019) the nsp8 extensions appear to remain flexible. Hillen *et al.* suggested that the nsp8 extension slides along the exiting RNA to prevent premature dissociation of the RdRp during replication (Hillen et al., 2020). K58 located at the N-terminus of nsp8 is conserved amongst members of the family *Coronaviridae* and is thought to interact with the exiting RNA (Hillen et al., 2020; Subissi et al., 2014; te Velthuis et al., 2012; Tvarogova et al., 2019). Collectively, the results obtained in this (and previous) work support the idea that alpha- and betacoronavirus nsp(7+8) complexes help to stabilize the interaction of nsp12 with the RNA substrate and product, thereby possibly acting as a major processivity factor of the RTC (Figure 26).

6.1.4 Model for the coronaviral replication

Finally, we propose a model for the initiation of coronaviral RNA synthesis and elongation in which proteolytic processing is an essential requirement for the generation of mature replicase enzymes and co-factors. Based on our *in vitro* results, the proteolytically released nsp7 and nsp8 initially assemble to form a stable heterodimer. Subsequently, these heterodimers likely interact with another heterodimer resulting in heterotrimers (possibly, by expelling one molecule of nsp7) as shown for FIPV and TGEV and, less frequently, for HCoV-229E, PEDV and MERS-CoV. Importantly however, in HCoV-229E, PEDV and MERS-CoV, a large proportion of nsp(7+8) complexes was shown to form heterotetrameric assemblies (Figure 41), just like in SARS-CoV in which heterotetramers (and no heterotrimers) are being formed. These pre-assembled nsp(7+8) complexes are thought to present the RNA substrate to nsp12, activated by Mn^{2+}, which then initiates processive RNA synthesis. To also allow for efficient elongation of the nascent RNA strand, Mn^{2+} is subsequently replaced by Mg^{2+} which is present at much higher concentrations in the cell.

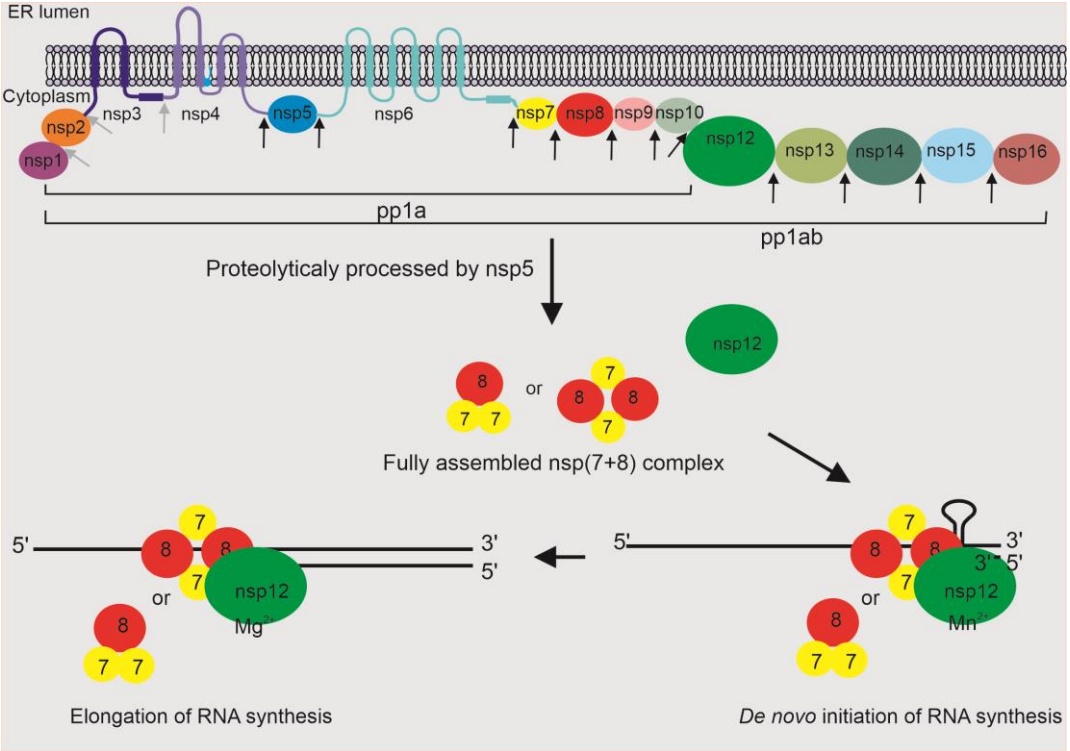


Figure 41: Schematic representation of the proteolytic processing of coronavirus polyproteins and assembly of nonstructural protein complexes that drive RNA synthesis. The coronavirus polyproteins contain proteinases residing in nsp3 (PL^{pro}; grey arrowheads indicate cleavage sites) and nsp5 (M^{pro}; black arrowheads indicate cleavage sites), nsp1, nsp2, hydrophobic domains (TM1, TM2 and TM3) in nsp3, nsp4 and nsp6 are indicated. Nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16 are thought to be part the RTC and are released from pp1a and/or pp1ab by nsp5-mediated cleavage at conserved sites. Nsp7 and nsp8 assemble as heterotrimeric (2:1) (TGEV and FIPV) or heterotetrameric (2:2) (HCoV-229E, PEDV, SARS-CoV and MERS-CoV) complexes. The fully assembled nsp(7+8) complex binds RNA and promotes or stabilizes interactions of nsp12 with the RNA. To initiate *de novo* RNA synthesis, the enzyme complex selectively requires Mn^{2+} as catalytic ion while, in the elongation reaction, Mg^{2+} is probably used as catalytic ion, which is present in much larger concentrations in the cell.

6.1.5 Nucleotide selection and fidelity of the HCoV-229E nsp(7+8)+nsp12 complex

Another goal of this work was to get insight into factors affecting nucleotide selection and fidelity of the nsp(7+8)+nsp12 complex. A previous study reported the degradation of RNAs synthesized by SARS-CoV nsp7L8/nsp12 in the presence of 1 mM Mg^{2+} when the nsp10/14 complex was added (Subissi et al., 2014). Another study suggested that the SARS-CoV nsp12 V553 and M611 residues participate in the regulation of fidelity

and nucleotide selection (Sexton et al., 2016). The data obtained in that study also suggested that nsp12 and the core component of the proofreading complex, nsp10/14, independently contribute to the fidelity during viral RNA synthesis.

Here, we demonstrate that the nsp(7+8)+nsp12 complex (itself) plays a role in nucleotide selection and fidelity independently of nsp10/14. Our investigations of nucleotide selection at high concentrations of Mg^{2+} and, particularly, Mn^{2+} revealed that the nsp(7+8)+nsp12 complex incorporated non-complementary nucleotides during RNA synthesis (Figure 28) on partial-duplex RNA substrates with template sequences containing U_{10}, A_{10}, C_{10}, $(CU)_5$ or $(UC)_5$. Thus, for example, GMP was efficiently incorporated if a template with a $(CU)_5$ stretch was provided (KR07/$(CU)_5$ KR07 comp-b; Figure 28 B and C). These findings led us to analyze a possible correlation between polymerase activity and metal ion concentration. Mg^{2+} is the most abundant divalent catalytic ion and its intracellular concentration is around 210 to 240 μM (Achuthan et al., 2014; Goldschmidt et al., 2006; Markesbery et al., 1984; Zhang and Ellis, 1989). Based on its abundance, Mg^{2+} is thus considered to be the physiologically most relevant catalytic metal ion for RdRp and DNA polymerases. The intracellular concentration of Mn^{2+} ranges from 0.1 to 40 μM (Ash and Schramm, 1982; Zhang and Ellis, 1989) and is much lower than that of Mg^{2+}. Taking this into consideration and consistent with the metal ion titration data shown in Figure 27, the nsp(7+8)+nsp12 complex exhibited robust activity at the biologically relevant concentration of Mg^{2+} and a two-fold increased concentration (compared to cellular levels) of Mn^{2+} (Figure 29). Importantly, the use of these lower concentrations of Mg^{2+} or Mn^{2+} ions reduced nucleotide misincorporations during RNA synthesis in reactions supplemented with nonmatching nucleotides (Figure 29). This effect was particularly evident if lower concentrations of Mn^{2+} were used (Figure 29 C). Similar observations were made for the DNA human

polymerase ι where physiological metal ion concentrations were found to increase the enzyme's catalytic efficiency and fidelity (Frank and Woodgate, 2007). When tested at these reduced metal ion concentrations, the coronavirus nsp(7+8)+nsp12 complex clearly discriminated between ribonucleotides (NTPs) and deoxyribonucleotides (dNTPs) (Figure 30 B and C), with an overall very low incorporation of dNTPs. Using a partial-duplex RNA substrate with a UC-containing single-stranded (template) sequence, dGTP was found to be incorporated slightly more efficiently than dATP, especially in the presence of Mg^{2+} (Figure 30 B).

Taken together, these data lead me to conclude that the nsp(7+8)+nsp12 complex is highly active at near-physiological concentrations of Mg^{2+} and Mn^{2+} and has superior nucleotide selectivity and fidelity compared to those determined at increased concentrations (1 mM) of these metal ions. The recent structural studies of coronavirus nsp12 and nsp12-containing complexes (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Wang et al., 2020) will greatly facilitate mutational analyses of residues suggested to be involved in metal ion interactions and nucleotide selectivity. Such studies can be expected to provide interesting new insight into the molecular mechanisms that control the fidelity and nucleotide selectivity of coronavirus RNA polymerases (alone and in conjunction with the nsp14 exonuclease).

6.2 Goal II

6.2.1 EAV nsp9 displays terminal adenylyl transferase activity

EAV nsp9 is a two-domain proteins carrying an RdRp domain with the characteristic SDD sequence signature in motif C (den Boon et al., 1991) and a NiRAN domain in the C- and N-terminal parts, respectively, of this central component of the arterivirus RTC (Lehmann et al., 2015). To date, there is no convincing evidence that nsp9 (on its own) has RdRp activity (Lehmann et al., 2016). While the minimal coronavirus replicase

complex was found to consist of nsp(7+8) and nsp12, with nsp(7+8) acting as an essential cofactor for the nsp12-associated RdRp activity, there is currently no information on whether the arterivirus RdRp (residing in nsp9) requires additional cofactors for activity.

The present study aimed at characterizing activities of nsp9 together with other small nsp subunits encoded in EAV ORFs 1a and 1b. In initial experiments, we were able to establish that EAV nsp9 has efficient 3'-terminal adenylyltransferase activity and that this activity was greatly stimulated if a partially double-stranded substrate with a U_{20} single-stranded template region was provided (Figure 32), which is consistent with the observations made by Lehmann *et al*. (Lehmann et al., 2016). Previous studies of viral RNA polymerases of corona-, polio-, noro-, sapo-, flavi-, noda-, toga- and cystovirus reported a similar kind of primer extension activity (Arnold et al., 1999; Fullerton et al., 2007; Neufeld et al., 1994; Poranen et al., 2008a; Ranjith-Kumar et al., 2001; Tomar et al., 2006). In additional experiments, we assessed the incorporation of GMP by nsp9 into RNA products using partial-duplex RNA substrates that contained stretches of CU in their single-stranded regions (Figure 32). We failed to obtain evidence for the incorporation of GMP. Collectively, these data contradict earlier reports describing *de novo* initiation of RNA synthesis by EAV nsp9 (Beerens et al., 2007; te Velthuis et al., 2010b) and they are consistent with data reported by Lehmann *et al.* (Lehmann et al., 2016).

In an additional set of experiments, the potential role of other small arterivirus nsps in promoting nsp9-mediated RNA synthesis was investigated. By analogy with the coronavirus nsp(7+8)+nsp12 complex, it seemed reasonable to speculate that nsp9 requires co-factors to attain its polymerase activity. To test this hypothesis, we purified several EAV-encoded nsps, such as nsp4, nsp6, nsp6-7-8, nsp7, nsp10 and nsp12

(Figure 33). In subsequent polymerase assays, a 3'-biotinylated 32-nt RNA substrate with four 3'-terminal adenylates was used as a substrate. Incubation of the various EAV nsps with nsp9 (wildtype or D445A RdRp active-site mutant) provided no evidence for a labeled RNA product of the expected length corresponding to that of the template RNA used (Figure 34). Instead, faster migrating radiolabeled products were detected in reactions supplemented with [α-^{32}P]-GTP or [α-^{32}P]-GTP uridylylation. Importantly, these products were also detected if the RdRp D445A active-site protein variant was used, while they were not detected if the NiRAN active-site variant K94 was used in the reaction. These results confirmed that the NiRAN (rather than the RdRp) domain catalyzed these products. Based on the (self)guanylylation and (self)uridylylation activities established previously for the EAV nsp9-associated NiRAN domain and the recently reported protein-directed nucleotidyltransferase activities of coronavirus NiRAN domains (Slanina et al., 2021), it seems reasonable to speculate that the prominent radiolabeled products observed in the presence of GTP or UTP and nsp6-7-8 represent guanylylated or uridylylated peptides produced by the proteinase K treatment of the reaction mixtures that was done with all samples prior to denaturing polyacrylamide gel electrophoresis (Figure 34). Uridylylation and guanylylation was dependent on the presence of nsp6-7-8 and nsp9 in the reaction mixture. Incubation of other replicase nsps with nsp9 did not lead to the detection of any other labeled products, suggesting a defined substrate specificity for this nucleotidyltransferase activity. In contrast to EAV, the control experiment with coronavirus nsp(7+8)+nsp12 replicase complex resulted in radiolabeled products of the expected (near-template) length.

6.2.2 Uridylylation and Guanylylation of EAV polyprotein nsp6-7-8

As a second enzymatic activity associated with arterivirus nsp9, a nucleotidyltransferase was identified and proposed to be mediated by the N-terminal NiRAN domain (Lehmann et al., 2015). A previous study suggested that nsp9 has (self)nucleotidylation activity with a strong preference for UTP and GTP and the chemical nature between of protein-nucleotide adduct was proposed to be a phosphoamide bond. Prompted by studies of coronavirus nsp12-mediated nucleotidyltransferase activities (Slanina et al., 2021) and observations made in our polymerase assays (Figure 34), nucleotidylation assays were performed for EAV nsp9 and nsp6-7-8 as a potential substrate of the NiRAN domain. Compared to an earlier study (Lehmann et al., 2015), we used a modified protocol in which a mixture of "cold" NTPs and [α-^{32}P]-radiolabeled NTPs (instead of only radiolabeled NTPs) was included in the reaction mixture. Also, we used a lower concentration of Mn^{2+} to minimize nonspecific activity. We observed that EAV nsp6-7-8 was radiolabeled by nsp9 in the presence of UTP or GTP and Mn^{2+} ions (Figure 35). Previously, an autonucleotidylation activity of the NiRAN domain was reported but no (other) target protein could be identified (Lehmann et al., 2015). Interestingly, only trace amounts of radiolabeled nucleotides were bound to nsp9 in the absence of other proteins while the addition of nsp6-7-8 precursor stimulated the autonucleotidylation of nsp9. More importantly, the data revealed that nsp9 efficiently nucleotidylated the nsp6-7-8 protein (*in trans*), indicating multiple-turnover reactions in which one molecule of nsp9 modifies multiple acceptor protein molecules. Substitution of a proposed active-site residue, Lys-94, with Ala (nsp9_ K94A) reduced the activity profoundly to near-background levels, suggesting that the nucleotidylation activity is mediated by the EAV nsp9-associated NiRAN domain (Figure 35).

Furthermore, uridylylated or guanylylated nsp6-7-8 complexes proved to stable at low pH of 2.0 and high temperature (100 ºC), suggesting formation of a covalent bond in the protein-GMP/UMP adducts. The deconvoluted intact protein spectra confirmed that one molecule of UMP or GMP was bound to one molecule of nsp6-7-8 protein (Figure 36), suggesting a specific nucleotidylation of the target protein at a single site.

6.2.3 Identification of the nucleotidylation acceptor protein

To delineate the protein (domain) that is nucleotidylated in the nsp6-7-8 polyprotein precursor, we produced the proteins nsp7, nsp7α and 7β (Figure 37) in *E. coli* and incubated them with nsp9 in the presence of radiolabled nucleotides. In addition to finding the nucleotide acceptor, we also performed adenylylation, guanylylation, uridylylation and cytidylylation ractions to determine which nucleotides were transferred to the acceptor protein and we optimized the metal ion concentrations to be used in nucleotidylation assays.

Uridylylation or guanylylation reactions using nsp7 and nsp7α proteins revealed efficient transfer of radiolabel in the presence of nsp9 and Mn^{2+}, whereas the nsp7β protein was not labeled (Figure 38). These results suggested that the uridylylation or guanylylation activity targets a residue in the N-terminal part of nsp7. The use of [α-^{32}P]-ATP or [α-^{32}P]-CTP did not lead to labeling of either of nsp7 or nsp7α, confirming that the EAV NiRAN domain uses selectively UTP and GTP as nucleotide cosubstrates and requires Mn^{2+} for its nucleotidyltranferase activity *in trans*, which is consistent with the self-nucleotidylation activity data obtained in a previous study (Lehmann et al., 2015). The structural basis and potential biological implications of the selective use of UTP and GTP remain to be investigated.

In a recent study, the coronavirus nsp12 NiRAN domain was shown to specifically nucleotidylate nsp9, a small ORF1a-encoded RNA-binding protein. Largely based on

genetic data, it was speculated that the nucleotidylated form of the coronavirus nsp9 might be involved in a protein-primed initiation of RNA synthesis, similar to what was demonstrated previously for the VPg and VPg-like proteins of several other plus-strand RNA viruses, such as picorna-, calici- and potyviruses (Machin et al., 2001; Paul et al., 2000; Paul et al., 1998; Puustinen and Makinen, 2004; Wedekind et al., 1996). In these viruses, the viral RdRp di-uridylylates the side chain of a conserved residue (e.g. tyrosine) of the small virus-encoded VPg protein. The resulting VPg-UpUp protein is subsequently used to prime viral RNA synthesis. To date, there is no direct biochemical evidence to confirm that nucleotidylated forms of coronavirus nsp9 (Slanina et al., 2021) act to prime RNA synthesis and also our arterivirus *in vitro* polymerase assay with nsp6-7-8 or nsp7 in the presence of EAV nsp9 failed to detect any activity. It remains to be investigated whether additional protein cofactors and/or RNA sequences or RNA structural elements are required to attain EAV nsp9-mediated RdRp activity *in vitro*.

Several conserved residues located in the NiRAN domain were demonstrated to be essential for EAV and SARS-CoV replication in cell culture (Lehmann et al., 2015). This previous study also suggested that K94 is the catalytic residue for nucleotidylation, while R124 and F166 residues were postulated to be involved in interactions with NTP (Lehmann et al., 2015). In this regard, we performed a mutational study of the NiRAN domain to investigate the role of specific residues in transferring nucleotides to nsp7 (Figure 40). Alanine substitutions of conserved lysine (K94A) and aspartic acid (D165A) residues abolished uridylylation and guanylylation of nsp7, whereas the replacement of a nonconserved lysine in the NiRAN domain (K106A) with alanine or a conserved aspartic acid residue in the RdRp domain (D445A) retained nsp7 uridylylation and guanylylation activity. Residual nucleotidylation activity was detected for the proteins with alanine replacements of S129, D132 and F166. Taken together, the

results correlate well with a previous mutagenesis analysis of the NiRAN domain (Lehmann et al., 2015), which suggested that the lack of nucleotidyltransferase activity in the nsp9 D132A and D165A mutant proteins was due to a defect in binding of Mn^{2+}, while the loss of activity for the F166A mutant was due to the lack of NTP binding or loss of critical interactions between nsp7 and nsp9. In contrast to data reported by Lehmann *et al.* (Lehmann et al., 2015) on nsp9 self-nucleotidylation activity, we were able to show for EAV nsp9_R124A that this protein variant retains significant uridylylation and guanylylation activity using nsp7 as a substrate. This residue was speculated to be involved in NTP binding. One possible reason for these conflicting results could be that interactions between nsp7 and nsp9 induce conformational changes that stimulate NiRAN's nucleotidylation activity in *trans* (and *cis*) and thereby help compensate some of the functional defects induced by the substitution of the conserved arginine residue 124 with alanine. Interactions between nsp7 and nsp9 have been suggested earlier (Chen et al., 2017) and the significant stimulation of self-nucleotidylation activity observed in the present study (Figure 35) suggests that such interactions induce conformational changes in the NiRAN active site that stimulate activity.

7 Summary

In this work, biochemical studies of nidovirus replicase proteins were performed. The experiments were based on tag-less coronavirus nsp7, nsp8 and pre-assembled nsp(7+8) complexes that were generated by nsp5-mediated proteolytic processing *in vitro* of precursor proteins produced in *E. coli*. Our investigations suggest that the primer-dependent coronavirus nsp12 polymerase is most active in the presence of a pre-formed nsp(7+8) complex, resulting in efficient elongation of primed (partially double-stranded) RNAs. Both Mg^{2+} and Mn^{2+} were able to support elongation activity, albeit with a reduced fidelity if high concentrations of Mn^{2+} were included in the reaction. Nsp8 (alone) was able to support (a low level of) primer-dependent nsp12 polymerase activity if the reaction was performed in the presence of Mn^{2+}. Furthermore, nsp8 and the pre-assembled nsp(7+8) complex were confirmed to have 3'-terminal adenylyltransferase activity, which was stimulated by using a partially double-stranded RNA substrate with a 5' oligo(U) overhang.

Pre-assembled nsp(7+8) complexes derived from different alpha- and betacoronaviruses and used in combination with the cognate nsp12 were shown to support *de novo* RNA synthesis in the presence of Mn^{2+} (but not Mg^{2+}) on 3'-biotinylated single-stranded RNA templates. Electrophoretic mobility shift assays indicated that the pre-assembled nsp(7+8) complex binds tightly to RNA and stabilizes the RNA binding of nsp12. Experiments in which the porcine coronavirus TGEV nsp12 was combined with (noncognate) nsp(7+8) complexes of other coronaviruses revealed that the TGEV nsp12 only retained activity when it was used in combination with the nsp(7+8) complex of FIPV, the most closely related coronavirus among the coronaviruses included in this experiment, suggesting that protein-protein interactions between the nsp(7+8) complex and nsp12 are structurally constrained and required to promote polymerase activity.

Alanine substitutions of two aspartic acid residues in the conserved RdRp signature sequence SDD in motif C abolished both *de novo* and primer-dependent polymerase activity, demonstrating that nsp12 (rather than a potentially contaminating *E. coli* protein) was responsible for the activity characterized in this study.

In another set of experiments, the formation of nsp(7+8) complexes was investigated. Using a combination of biochemical methods and native mass spectrometry, FIPV and TGEV nsp(7+8) complexes were found to form predominantly heterotrimeric complexes (2:1), while the nsp(7+8) complexes of HCoV-229E, PEDV, SARS-CoV and MERS-CoV had a predominantly heterotetrameric architecture (2:2). The data also provided evidence that, in all coronaviruses included in this study, heterodimeric nsp(7+8) subcomplexes are formed that likely represent the building blocks for the assembly of larger complexes. In addition, a limited proportion of the nsp(7+8) complexes of HCoV-229E, PEDV, and MERS-CoV were confirmed to be heterotrimers (2:1).

A second major part of the study dealt with the replicase complex of another nidovirus family, focusing on equine arteritis virus (EAV), a member of the *Arteriviridae* family. In previous studies on the EAV nsp9 polymerase protein, no conclusive evidence was obtained to suggest that nsp9 (alone) has RdRp activity. Therefore, the present study investigated the polymerase activity of EAV nsp9 more extensively using a range of RNA substrates, a set of nsp9 NiRAN and RdRp mutants and other arterivirus nsps that, by analogy to the situation in coronaviruses, were speculated to have role in activating the arterivirus nsp9 RdRp. We were able to show that EAV nsp9 has RNA 3'-terminal adenylyltransferase activity while the protein failed to transfer guanylate. Attempts to obtain evidence for nsp9-mediated RNA synthesis in the presence of other EAV nsps and Mn^{2+} failed to demonstrate RdRp activity but indicated that the NiRAN domain

potentially transfers UMP and GMP to nsp6-7-8. Subsequent studies suggested that the nsp9-associated NiRAN domain specifically transfers UMP and GMP but not AMP or CMP to the nsp7α portion of nsp7 in the presence of Mn^{2+}. The analysis of deconvulated intact protein mass spectra suggested that nsp9 transfers (only) one NMP molecule to nsp6-7-8 or nsp7, suggesting a specific modification site in these target proteins. The data, along with the observed stability of the protein-NMP adduct at low pH and under daturing conditions, also suggest the formation of a covalent bond between NMP and the protein. Alanine replacement of two putative active-site residues in nsp9 (K94 and D165) abolished the nsp7-directed uridylylation and guanylylation activities whereas several other mutations in the conserved NiRAN motifs A_N, B_N or C_N in nsp9 caused only partial defects in these *in vitro* assays. Alanine replacement of a non-conserved residue in the NiRAN domain and a conserved Lys in motif A of the RdRp domain (D445A) retained protein nucleotidylation activity, indicating that (only) the NiRAN domain mediates the NMP transfer. The role of the newly identified nsp7-targeted NiRAN-mediated nucleotidylation activity in the EAV replication cycle remains to be determined.

8 Zusammenfassung

In dieser Arbeit wurden biochemische Studien zu nidoviralen Replikase-Proteinen durchgeführt. Die Experimente stützten sich auf Marker-freies nsp7 und 8 sowie einen vorgeformten nsp(7+8)-Komplex, der durch eine nsp5-vermittelte proteolytische Prozessierung aus einem in *E. coli* produzierten Vorläuferprotein *in vitro* erzeugt wurde. Die Ergebnisse zeigen, dass die Aktivität der Primer-abhängigen Coronavirus-nsp12-Polymerase in Gegenwart eines vorgeformten nsp(7+8)-Komplexes am höchsten ist. Dieser Komplex ermöglicht es dem nsp12-Protein, eine effiziente Elongation eines RNA-Primers im Kontext eines partiell doppelsträngigen RNA-Substrats zu katalysieren. Sowohl Mg^{2+} als auch Mn^{2+} unterstützten diese Elongationsaktivität, wenngleich die Kopiergenauigkeit (*fidelity*) in Gegenwart erhöhter Mn^{2+}-Konzentrationen reduziert war. Auch nsp8 allein unterstützte die Primer-abhängige nsp12-Polymerase-Aktivität, wenngleich auf niedrigerem Niveau und nur in Gegenwart von Mn^{2+}-Ionen. Weiterhin konnte eine 3'-terminale RNA-Adenylyltransferaseaktivität für nsp8 und den vorgeformten nsp(7+8) Komplex bestätigt werden. Diese Aktivität wurde deutlich stimuliert, wenn das partiell doppelsträngige RNA-Substrat einen oligo(U_{10}) Überhang besaß.

Vergleichende Experimente zu nsp(7+8)-Komplexen unterschiedlicher Alpha- und Betacoronaviren in Kombination mit dem dazugehörigen nsp12 zeigten übereinstimmend, dass diese Komplexe eine *De-novo*-RNA-Synthese in Gegenwart von Mn^{2+} (nicht jedoch Mg^{2+}) ermöglichen, wobei eine 3'-biotinylierte einzelsträngige RNA als Matrize verwendet wurde. EMSA-Experimente ergaben Hinweise darauf, dass der nsp(7+8)-Komplex fest an RNA bindet und dabei hilft, die RNA-Bindung von nsp12 zu verstärken.

Experimente, in denen die Polymeraseaktivität des TGEV-nsp12 in Anwesenheit des nsp(7+8)-Komplexes eines anderen Alpha- oder Betacoronavirus getestet wurde, ergaben, dass nsp12 nur dann seine Aktivität behielt, wenn es mit dem nsp(7+8)-Komplex von FIPV, also dem (zu TGEV) nächstverwandten Coronavirus, eingesetzt wurde. Die Daten zeigen die Notwendigkeit fein abgestimmter Protein-Protein-Wechselwirkungen zwischen dem nsp(7+8)-Komplex und nsp12, die für die Aktivierung der nsp12-Polymeraseaktivität zwingend erforderlich sind.

Alanin-Substitutionen zweier Aspartat-Reste innerhalb der konservierten RdRp-Signatursequenz SDD im Motif C blockierten sowohl die Primer-abhängige also auch die *de-novo*-initiierte RNA-Synthese. Diese Daten unterstützen die Schlussfolgerung, dass die in dieser Arbeit untersuchten Polymerase-Aktivitäten tatsächlich von nsp12 (und nicht von eventuell vorhandenen Verunreinigungen aus *E.-coli*-Zellen) vermittelt wurden.

In weiteren Experimenten wurde die Bildung von nsp(7+8)-Komplexen untersucht. Unter Verwendung unterschiedlicher biochemischer Methoden und der nativen Massenspektrometrie konnte gezeigt werden, dass FIPV- und TGEV-nsp(7+8)-Komplexe überwiegend als Heterotrimer (2:1) vorliegen, während die nsp(7+8)-Komplexe von HCoV-229E, PEDV, SARS-CoV und MERS-CoV hauptsächlich Heterotetramere (2:2) bildeten. Die Daten zeigten außerdem, dass alle in dieser Arbeit untersuchten coronaviralen nsp(7+8)-Komplexe auch als Heterodimere vorlagen, die vermutlich Ausgangspunkt und Grundbausteine für die Bildung höhermolekularer Proteinkomplexe sind. Schließlich ergaben sich Hinweise darauf, dass HCoV-229E, PEDV und MERS-CoV auch Heterotrimere (2:1) bilden, wenn auch nur in begrenztem Umfang.

Ein weiterer Hauptteil der Arbeit bechäftigte sich mit dem Replikase-Komplex einer anderen Nidovirusfamilie. Diese Arbeiten konzentrierten sich auf EAV (e*quine arteritis virus*), ein Virus aus der Familie der *Arteriviridae*. Frühere Studien hatten keinen überzeugenden Nachweis einer RdRp-Aktivität für die arterivirale nsp9-Polymerase gebracht. Aus diesem Grund sollten weitere Studien durchgeführt werden zur Identifizierung und Charakterisierung des EAV-nsp9-Proteins unter Verwendung unterschiedlicher RNA-Substarte, eines Satzes von NIRAN- und RdRp-Proteinvarianten mit spezifischen Aminosäuresubstitutionen und einer Reihe weiterer arteriviraler Nichtstrukturproteine, die (ähnlich wie bei Coronaviren) eine aktivierende (Kofaktor-) Funktion für die nsp9-RdRp haben könnten. Dabei ist es gelungen, eine 3'-terminale Adenylyltransferaseaktivität für EAV nsp9 nachzuweisen, während das Protein keine erkennbare Fähigkeit besaß, GMP zu übertragen. Versuche, eine nsp9-vermittelte Polymeraseaktivität in Gegenwart anderer Nichtstrukturproteine nachzuweisen, waren nicht erfolgreich. Die durchgeführten Experimente ergaben jedoch erste Hinweise darauf, dass die NiRAN-Domäne an der Übertragung von UMP und GMP auf nsp6-7-8 beteiligt sein könnte. In nachfolgenden Studien zeigte sich, dass die nsp9-assoziierte NiRAN-Domäne UMP und GMP (nicht jedoch AMP und CMP) spezifisch auf den nsp7α-Anteil des nsp7-Proteins überträgt und dass diese Reaktion Mn^{2+} benötigt. Massenspektrometrische Analysen legten nahe, dass nsp9 (exakt) ein NMP-Molekül auf nsp6-7-8 oder nsp7 überträgt, was wiederum auf eine sehr spezifische Modifikation im Zielprotein hindeutet. Diese Daten, ergänzt durch die beobachtete Stabilität des Protein-NMP-Addukts bei niedrigem pH und unter denaturierenden Bedingungen, lassen darauf schließen, dass das NMP über eine kovalente Bindung an das Zielprotein gekoppelt ist. Substitutionen von zwei Aminosäure-Resten im putativen aktiven Zentrum der NiRAN-Domäne mit Alanin

(K94A und D165A) blockierten die nsp7-gerichtete Uridylylierung und Guanylylierung, während einige andere Mutationen in den konservierten NiRAN-Sequenzmotiven A_N, B_N oder C_N nur partielle funktionelle Defekte in diesen *In-vitro*-Assays zeigten. Der Austausch eines nichtkonservierten Restes in der NiRAN-Domäne und eines konservierten Lysin-Restes im Motiv A der RdRp-Domäne (D445A) hatte hingegen keinen Einfluss auf die Nukleotidylierungsaktivität. Letzteres lässt den Schluss zu, dass (allein) die NiRAN-Domäne für die Nukleotidyltransferaseaktivität verantwortlich ist. Die Aufklärung der biologischen Relevanz dieser neu identifizierten, auf das nsp7 gerichteten Nukleotidylierungsaktivität im arteriviralen Replikationszyklus bleibt weiteren Studien vorbehalten.

9 References

Achuthan, V., Keith, B.J., Connolly, B.A., DeStefano, J.J., 2014. Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions in vitro. J Virol 88, 8514-8527.

Ackermann, M., Padmanabhan, R., 2001. De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. J Biol Chem 276, 39926-39937.

Adedeji, A.O., Marchand, B., Te Velthuis, A.J., Snijder, E.J., Weiss, S., Eoff, R.L., Singh, K., Sarafianos, S.G., 2012. Mechanism of nucleic acid unwinding by SARS-CoV helicase. PLoS One 7, e36521.

Ahn, D.G., Choi, J.K., Taylor, D.R., Oh, J.W., 2012. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. Arch Virol 157, 2095-2104.

An, S., Chen, C.J., Yu, X., Leibowitz, J.L., Makino, S., 1999. Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer. J Virol 73, 7853-7859.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., Hilgenfeld, R., 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J 21, 3213-3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 300, 1763-1767.

Angelini, M.M., Akhlaghpour, M., Neuman, B.W., Buchmeier, M.J., 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. mBio 4.

Arnold, J.J., Ghosh, S.K., Cameron, C.E., 1999. Poliovirus RNA-dependent RNA polymerase (3D(pol)). Divalent cation modulation of primer, template, and nucleotide selection. J Biol Chem 274, 37060-37069.

Arnold, J.J., Gohara, D.W., Cameron, C.E., 2004. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mn2+. Biochemistry 43, 5138-5148.

Ash, D.E., Schramm, V.L., 1982. Determination of free and bound manganese(II) in hepatocytes from fed and fasted rats. J Biol Chem 257, 9261-9264.

Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V., Lai, M.M., 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. J Virol 67, 6056-6063.

Baliji, S., Cammer, S.A., Sobral, B., Baker, S.C., 2009. Detection of nonstructural protein 6 in murine coronavirus-infected cells and analysis of the transmembrane topology by using bioinformatics and molecular approaches. J Virol 83, 6957-6962.

Barrette-Ng, I.H., Ng, K.K., Mark, B.L., Van Aken, D., Cherney, M.M., Garen, C., Kolodenko, Y., Gorbalenya, A.E., Snijder, E.J., James, M.N., 2002. Structure of arterivirus nsp4. The smallest

chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. J Biol Chem 277, 39960-39966.

Beerens, N., Selisko, B., Ricagno, S., Imbert, I., van der Zanden, L., Snijder, E.J., Canard, B., 2007. De novo initiation of RNA synthesis by the arterivirus RNA-dependent RNA polymerase. J Virol 81, 8384-8395.

Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4, 1011-1033.

Bonilla, P.J., Hughes, S.A., Weiss, S.R., 1997. Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. J Virol 71, 900-909.

Boscarino, J.A., Logan, H.L., Lacny, J.J., Gallagher, T.M., 2008. Envelope protein palmitoylations are crucial for murine coronavirus assembly. J Virol 82, 2989-2999.

Bougie, I., Charpentier, S., Bisaillon, M., 2003. Characterization of the metal ion binding properties of the hepatitis C virus RNA polymerase. J Biol Chem 278, 3868-3875.

Boursnell, M.E., Brown, T.D., Foulds, I.J., Green, P.F., Tomley, F.M., Binns, M.M., 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J Gen Virol 68 (Pt 1), 57-77.

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E.J., Canard, B., Decroly, E., 2010. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog 6, e1000863.

Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., Decroly, E., 2012. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proc Natl Acad Sci U S A 109, 9372-9377.

Bredenbeek, P.J., Pachuk, C.J., Noten, A.F., Charite, J., Luytjes, W., Weiss, S.R., Spaan, W.J., 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res 18, 1825-1832.

Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication. Curr Top Microbiol Immunol 287, 1-30.

Brierley, I., Dos Ramos, F.J., 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. Virus Res 119, 29-42.

Brooks, R.R., Andersen, J.A., 1978. Substrate, metal and template effects on inhibition of bacteriophage-qbeta ribonucleic acid polymerase by ortho- and pyro-phosphate. Biochem J 171, 725-732.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., Stuart, D.I., 2001. A mechanism for initiating RNA-dependent RNA polymerization. Nature 410, 235-240.

Castro, C., Arnold, J.J., Cameron, C.E., 2005. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Res 107, 141-149.

Chen, D., Patton, J.T., 2000. De novo synthesis of minus strand RNA by the rotavirus RNA polymerase in a cell-free system involves a novel mechanism of initiation. RNA 6, 1455-1467.

Chen, J., Xu, X., Tao, H., Li, Y., Nan, H., Wang, Y., Tian, M., Chen, H., 2017. Structural Analysis of Porcine Reproductive and Respiratory Syndrome Virus Non-structural Protein 7alpha (NSP7alpha) and Identification of Its Interaction with NSP9. Front Microbiol 8, 853.

Chen, S., Jonas, F., Shen, C., Hilgenfeld, R., 2010. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. Protein Cell 1, 59-74.

Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., Guo, D., 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. Proc Natl Acad Sci U S A 106, 3484-3489.

Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P., Ahola, T., Liang, Y., Liu, X., Guo, D., 2011. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. PLoS Pathog 7, e1002294.

Chen, Y., Tao, J., Sun, Y., Wu, A., Su, C., Gao, G., Cai, H., Qiu, S., Wu, Y., Ahola, T., Guo, D., 2013. Structure-function analysis of severe acute respiratory syndrome coronavirus RNA cap guanine-N7-methyltransferase. J Virol 87, 6296-6305.

Cheng, A., Zhang, W., Xie, Y., Jiang, W., Arnold, E., Sarafianos, S.G., Ding, J., 2005. Expression, purification, and characterization of SARS coronavirus RNA polymerase. Virology 335, 165-176.

Cirino, N.M., Cameron, C.E., Smith, J.S., Rausch, J.W., Roth, M.J., Benkovic, S.J., Le Grice, S.F., 1995. Divalent cation modulation of the ribonuclease functions of human immunodeficiency virus reverse transcriptase. Biochemistry 34, 9936-9943.

Coronaviridae Study Group of the International Committee on Taxonomy of, V., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5, 536-544.

Cowley, J.A., Dimmock, C.M., Spann, K.M., Walker, P.J., 2000. Gill-associated virus of Penaeus monodon prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. J Gen Virol 81, 1473-1484.

de Groot, R.J., Cowley, J. A., Enjuanes, L., Faaberg, K. S., Perlman, S., Rottier, P. J. M., Snijder, E. J., Ziebuhr, J. & Gorbalenya, A. E. (2012b). Order Nidovirales. In Virus Taxonomy, pp. 785-795. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. Amsterdam: Elsevier., 2012. Order Nidovirales. In Virus Taxonomy. Elsevier, Amsterdam.

de Haan, C.A., Li, Z., te Lintelo, E., Bosch, B.J., Haijema, B.J., Rottier, P.J., 2005. Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. J Virol 79, 14451-14456.

de Haan, C.A., Reggiori, F., 2008. Are nidoviruses hijacking the autophagy machinery? Autophagy 4, 276-279.

de Vries, A.A., Chirnside, E.D., Horzinek, M.C., Rottier, P.J., 1992. Structural proteins of equine arteritis virus. J Virol 66, 6294-6303.

Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., Canard, B., 2011a. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog 7, e1002059.

Decroly, E., Ferron, F., Lescar, J., Canard, B., 2011b. Conventional and unconventional mechanisms for capping viral mRNA. Nat Rev Microbiol 10, 51-65.

Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder, E.J., Canard, B., 2008. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. J Virol 82, 8071-8084.

den Boon, J.A., Snijder, E.J., Chirnside, E.D., de Vries, A.A., Horzinek, M.C., Spaan, W.J., 1991. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J Virol 65, 2910-2920.

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

Donaldson, E.F., Graham, R.L., Sims, A.C., Denison, M.R., Baric, R.S., 2007a. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J Virol 81, 7086-7098.

Donaldson, E.F., Sims, A.C., Graham, R.L., Denison, M.R., Baric, R.S., 2007b. Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. J Virol 81, 6356-6368.

Drake, J.W., 1999. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci 870, 100-107.

Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L., Denison, M.R., 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J Virol 81, 12135-12144.

Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., Canard, B., 2004. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc Natl Acad Sci U S A 101, 3792-3796.

Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., Imbert, I., 2018. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. Proc Natl Acad Sci U S A 115, E162-E171.

Frank, E.G., Woodgate, R., 2007. Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. J Biol Chem 282, 24689-24696.

Fullerton, S.W., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P.A., Rohayem, J., 2007. Structural and functional characterization of sapovirus RNA-dependent RNA polymerase. J Virol 81, 1858-1871.

Gadlage, M.J., Denison, M.R., 2010. Exchange of the coronavirus replicase polyprotein cleavage sites alters protease specificity and processing. J Virol 84, 6894-6898.

Gallagher, T.M., Buchmeier, M.J., 2001. Coronavirus spike proteins in viral entry and pathogenesis. Virology 279, 371-374.

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

Gohara, D.W., Ha, C.S., Kumar, S., Ghosh, B., Arnold, J.J., Wisniewski, T.J., Cameron, C.E., 1999. Production of "authentic" poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in Escherichia coli. Protein Expr Purif 17, 128-138.

Goldschmidt, V., Didierjean, J., Ehresmann, B., Ehresmann, C., Isel, C., Marquet, R., 2006. Mg2+ dependency of HIV-1 reverse transcription, inhibition by nucleoside analogues and resistance. Nucleic Acids Res 34, 42-52.

Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P., Koonin, E.V., 1989a. An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. J Mol Evol 28, 256-268.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res 117, 17-37.

Gorbalenya, A.E., Koonin, E.V., 1989. Viral proteins containing the purine NTP-binding sequence pattern. Nucleic Acids Res 17, 8413-8440.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1988. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Lett 235, 16-24.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1989b. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res 17, 4847-4861.

Gorbalenya, A.E., Koonin, E.V., Lai, M.M., 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. FEBS Lett 288, 201-205.

Gorbalenya, A.E., Pringle, F.M., Zeddam, J.L., Luke, B.T., Cameron, C.E., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H., Ward, V.K., 2002. The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J Mol Biol 324, 47-62.

Gordon, C.J., Tchesnokov, E.P., Feng, J.Y., Porter, D.P., Gotte, M., 2020. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. J Biol Chem 295, 4773-4779.

Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K., Baker, S.C., 2002. RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. J Virol 76, 3697-3708.

Graham, R.L., Denison, M.R., 2006. Replication of murine hepatitis virus is regulated by papain-like proteinase 1 processing of nonstructural proteins 1, 2, and 3. J Virol 80, 11610-11620.

Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S.C., Mesecar, A.D., 2008. Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. Virus Res 133, 63-73.

Guarino, L.A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., Kao, C., 2005. Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. J Mol Biol 353, 1106-1117.

Haider, S.R., Reid, H.J., Sharp, B.L., 2012. Tricine-SDS-PAGE. Methods Mol Biol 869, 81-91.

Hegyi, A., Friebe, A., Gorbalenya, A.E., Ziebuhr, J., 2002. Mutational analysis of the active centre of coronavirus 3C-like proteases. J Gen Virol 83, 581-593.

Herold, J., Raabe, T., Schelle-Prinz, B., Siddell, S.G., 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. Virology 195, 680-691.

Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. Nature.

Hofmann, M.A., Brian, D.A., 1991. The 5' end of coronavirus minus-strand RNAs contains a short poly(U) tract. J Virol 65, 6331-6333.

Hsu, M.F., Kuo, C.J., Chang, K.T., Chang, H.C., Chou, C.C., Ko, T.P., Shr, H.L., Chang, G.G., Wang, A.H., Liang, P.H., 2005. Mechanism of the maturation process of SARS-CoV 3CL protease. J Biol Chem 280, 31257-31266.

Hsue, B., Masters, P.S., 1997. A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. J Virol 71, 7567-7578.

Huang, Y., Beaudry, A., McSwiggen, J., Sousa, R., 1997. Determinants of ribose specificity in RNA polymerization: effects of Mn2+ and deoxynucleoside monophosphate incorporation into transcripts. Biochemistry 36, 13718-13728.

Hudson, C.B., Beaudette, F.R., 1932. Infection of the Cloaca with the Virus of Infectious Bronchitis. Science 76, 34.

Hyllseth, B., 1973. Structural proteins of equine arteritis virus. Arch Gesamte Virusforsch 40, 177-188.

Imbert, I., Guillemot, J.C., Bourhis, J.M., Bussetta, C., Coutard, B., Egloff, M.P., Ferron, F., Gorbalenya, A.E., Canard, B., 2006. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J 25, 4933-4942.

Imbert, I., Snijder, E.J., Dimitrova, M., Guillemot, J.C., Lecine, P., Canard, B., 2008. The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Res 133, 136-148.

Ivanov, K.A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A.E., Ziebuhr, J., 2004a. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. Proc Natl Acad Sci U S A 101, 12694-12699.

Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J., Ziebuhr, J., 2004b. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol 78, 5619-5632.

Ivanov, K.A., Ziebuhr, J., 2004. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. J Virol 78, 7833-7838.

Jones, D.H., Howard, B.H., 1991. A rapid method for recombination and site-specific mutagenesis by placing homologous ends on DNA using polymerase chain reaction. Biotechniques 10, 62-66.

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M., Moy, K., Yadav, M.K., Velasquez, J., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2006. Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. J Virol 80, 7894-7901.

Kanjanahaluethai, A., Chen, Z., Jukneliene, D., Baker, S.C., 2007. Membrane topology of murine coronavirus replicase nonstructural protein 3. Virology 361, 391-401.

Kao, C.C., Del Vecchio, A.M., Zhong, W., 1999. De novo initiation of RNA synthesis by a recombinant flaviviridae RNA-dependent RNA polymerase. Virology 253, 1-7.

Kao, C.C., Singh, P., Ecker, D.J., 2001. De novo initiation of viral RNA-dependent RNA synthesis. Virology 287, 251-260.

Kirchdoerfer, R.N., Ward, A.B., 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun 10, 2342.

Knoops, K., Kikkert, M., Worm, S.H., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M., Snijder, E.J., 2008. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. PLoS Biol 6, e226.

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

Krichel, B., Bylapudi, G., Schmidt, C., Blanchet, C., Schubert, R., Brings, L., Koehler, M., Zenobi, R., Svergun, D., Lorenzen, K., Madhugiri, R., Ziebuhr, J., Uetrecht, C., 2021. Hallmarks of Alpha- and Betacoronavirus non-structural protein 7+8 complexes. Sci Adv 7.

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

Kuo, L., Hurst, K.R., Masters, P.S., 2007. Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function. J Virol 81, 2249-2262.

Kuo, L., Masters, P.S., 2003. The small envelope protein E is not essential for murine coronavirus replication. J Virol 77, 4597-4608.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv Virus Res 48, 1-100.

Lauber, C., Goeman, J.J., Parquet Mdel, C., Nga, P.T., Snijder, E.J., Morita, K., Gorbalenya, A.E., 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog 9, e1003500.

Laurila, M.R., Makeyev, E.V., Bamford, D.H., 2002. Bacteriophage phi 6 RNA-dependent RNA polymerase: molecular details of initiating nucleic acid synthesis without primer. J Biol Chem 277, 17117-17124.

Laurila, M.R.L., Salgado, P.S., Stuart, D.I., Grimes, J.M., Bamford, D.H., 2005. Back-priming mode of phi6 RNA-dependent RNA polymerase. J Gen Virol 86, 521-526.

Lee, H.J., Shieh, C.K., Gorbalenya, A.E., Koonin, E.V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., Lai, M.M., 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology 180, 567-582.

Lehmann, K.C., Gorbalenya, A.E., Snijder, E.J., Posthuma, C.C., 2016. Arterivirus RNA-dependent RNA polymerase: Vital enzymatic activity remains elusive. Virology 487, 68-74.

Lehmann, K.C., Gulyaeva, A., Zevenhoven-Dobbe, J.C., Janssen, G.M., Ruben, M., Overkleeft, H.S., van Veelen, P.A., Samborskiy, D.V., Kravchenko, A.A., Leontovich, A.M., Sidorov, I.A., Snijder, E.J., Posthuma, C.C., Gorbalenya, A.E., 2015. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. Nucleic Acids Res 43, 8416-8434.

Lei, J., Kusov, Y., Hilgenfeld, R., 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res 149, 58-74.

Li, S., Zhao, Q., Zhang, Y., Zhang, Y., Bartlam, M., Li, X., Rao, Z., 2010. New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. Protein Cell 1, 198-204.

Li, Y., Breaker, R.R., 2001. In vitro selection of kinase and ligase deoxyribozymes. Methods 23, 179-190.

Li, Y., Tas, A., Snijder, E.J., Fang, Y., 2012. Identification of porcine reproductive and respiratory syndrome virus ORF1a-encoded non-structural proteins in virus-infected cells. J Gen Virol 93, 829-839.

Li, Y., Tas, A., Sun, Z., Snijder, E.J., Fang, Y., 2015. Proteolytic processing of the porcine reproductive and respiratory syndrome virus replicase. Virus Res 202, 48-59.

Littler, D.R., Gully, B.S., Colson, R.N., Rossjohn, J., 2020. Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. iScience 23, 101258.

Liu, D.X., Brown, T.D., 1995. Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. Virology 209, 420-427.

Lorsch, J., 2013. Methods in Enzymology. Laboratory methods in enzymology: DNA. Preface. Methods Enzymol 529, xix.

Luo, G., Hamatake, R.K., Mathis, D.M., Racela, J., Rigat, K.L., Lemm, J., Colonno, R.J., 2000. De novo initiation of RNA synthesis by the RNA-dependent RNA polymerase (NS5B) of hepatitis C virus. J Virol 74, 851-863.

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., Rao, Z., 2015. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci U S A 112, 9436-9441.

Machin, A., Martin Alonso, J.M., Parra, F., 2001. Identification of the amino acid residue involved in rabbit hemorrhagic disease virus VPg uridylylation. J Biol Chem 276, 27787-27792.

Madhugiri, R., Fricke, M., Marz, M., Ziebuhr, J., 2014. RNA structure analysis of alphacoronavirus terminal genome regions. Virus Res 194, 76-89.

Maier, H.J., Hawes, P.C., Cottam, E.M., Mantell, J., Verkade, P., Monaghan, P., Wileman, T., Britton, P., 2013. Infectious bronchitis virus generates spherules from zippered endoplasmic reticulum membranes. mBio 4, e00801-00813.

Manolaridis, I., Gaudin, C., Posthuma, C.C., Zevenhoven-Dobbe, J.C., Imbert, I., Canard, B., Kelly, G., Tucker, P.A., Conte, M.R., Snijder, E.J., 2011. Structure and genetic analysis of the arterivirus nonstructural protein 7alpha. J Virol 85, 7449-7453.

Markesbery, W.R., Ehmann, W.D., Alauddin, M., Hossain, T.I., 1984. Brain trace element concentrations in aging. Neurobiol Aging 5, 19-28.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., Ziebuhr, J., 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci U S A 103, 5108-5113.

Morin, B., Rahmeh, A.A., Whelan, S.P., 2012. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. EMBO J 31, 1320-1329.

Muramatsu, T., Takemoto, C., Kim, Y.T., Wang, H., Nishii, W., Terada, T., Shirouzu, M., Yokoyama, S., 2016. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. Proc Natl Acad Sci U S A 113, 12997-13002.

Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J., Brierley, I., 2006. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature 441, 244-247.

Nedialkova, D.D., Ulferts, R., van den Born, E., Lauber, C., Gorbalenya, A.E., Ziebuhr, J., Snijder, E.J., 2009. Biochemical characterization of arterivirus nonstructural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease. J Virol 83, 5671-5682.

Neufeld, K.L., Galarza, J.M., Richards, O.C., Summers, D.F., Ehrenfeld, E., 1994. Identification of terminal adenylyl transferase activity of the poliovirus polymerase 3Dpol. J Virol 68, 5811-5818.

Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Makino, S., Sawicki, S.G., Siddell, S.G., Stamou, D.G., Wilson, I.A., Kuhn, P., Buchmeier, M.J., 2011. A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol 174, 11-22.

Ng, K.K., Pendas-Franco, N., Rojo, J., Boga, J.A., Machin, A., Alonso, J.M., Parra, F., 2004. Crystal structure of norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. J Biol Chem 279, 16638-16645.

Nomaguchi, M., Ackermann, M., Yon, C., You, S., Padmanabhan, R., 2003. De novo synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. J Virol 77, 8831-8842.

Nomaguchi, M., Teramoto, T., Yu, L., Markoff, L., Padmanabhan, R., 2004. Requirements for West Nile virus (-)- and (+)-strand subgenomic RNA synthesis in vitro by the viral RNA-dependent RNA polymerase expressed in Escherichia coli. J Biol Chem 279, 12141-12151.

Noton, S.L., Aljabr, W., Hiscox, J.A., Matthews, D.A., Fearns, R., 2014. Factors affecting de novo RNA synthesis and back-priming by the respiratory syncytial virus polymerase. Virology 462-463, 318-327.

Ogando, N.S., Ferron, F., Decroly, E., Canard, B., Posthuma, C.C., Snijder, E.J., 2019. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. Front Microbiol 10, 1813.

Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P., te Lintelo, E.G., Rottier, P.J., de Haan, C.A., 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. J Virol 82, 12392-12405.

Opstelten, D.J., Raamsman, M.J., Wolfs, K., Horzinek, M.C., Rottier, P.J., 1995. Envelope glycoprotein interactions in coronavirus assembly. J Cell Biol 131, 339-349.

Oudshoorn, D., Rijs, K., Limpens, R., Groen, K., Koster, A.J., Snijder, E.J., Kikkert, M., Barcena, M., 2017. Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. mBio 8.

Pasternak, A.O., Spaan, W.J., Snijder, E.J., 2006. Nidovirus transcription: how to make sense...? J Gen Virol 87, 1403-1421.

Paul, A.V., Rieder, E., Kim, D.W., van Boom, J.H., Wimmer, E., 2000. Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. J Virol 74, 10359-10370.

Paul, A.V., van Boom, J.H., Filippov, D., Wimmer, E., 1998. Protein-primed RNA synthesis by purified poliovirus RNA polymerase. Nature 393, 280-284.

Pinto, D., Sarocchi-Landousy, M.T., Guschlbauer, W., 1979. 2'-Deoxy-2'-fluorouridine-5'-triphosphates: a possible substrate for E. coli RNA polymerase. Nucleic Acids Res 6, 1041-1048.

Plotch, S.J., Bouloy, M., Ulmanen, I., Krug, R.M., 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23, 847-858.

Polyanichko, A.M., Andrushchenko, V.V., Chikhirzhina, E.V., Vorob'ev, V.I., Wieser, H., 2004. The effect of manganese(II) on DNA structure: electronic and vibrational circular dichroism studies. Nucleic Acids Res 32, 989-996.

Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R., Hilgenfeld, R., 2008. Variable oligomerization modes in coronavirus non-structural protein 9. J Mol Biol 383, 1081-1096.

Poranen, M.M., Koivunen, M.R., Bamford, D.H., 2008a. Nontemplated terminal nucleotidyltransferase activity of double-stranded RNA bacteriophage phi6 RNA-dependent RNA polymerase. J Virol 82, 9254-9264.

Poranen, M.M., Salgado, P.S., Koivunen, M.R., Wright, S., Bamford, D.H., Stuart, D.I., Grimes, J.M., 2008b. Structural explanation for the role of Mn2+ in the activity of phi6 RNA-dependent RNA polymerase. Nucleic Acids Res 36, 6633-6644.

Posthuma, C.C., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., Blokhuis, J.H., Gorbalenya, A.E., Snijder, E.J., 2006. Site-directed mutagenesis of the Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle. J Virol 80, 1653-1661.

Puustinen, P., Makinen, K., 2004. Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. J Biol Chem 279, 38103-38110.

Ranjith-Kumar, C.T., Gajewski, J., Gutshall, L., Maley, D., Sarisky, R.T., Kao, C.C., 2001. Terminal nucleotidyl transferase activity of recombinant Flaviviridae RNA-dependent RNA polymerases: implication for viral RNA synthesis. J Virol 75, 8615-8623.

Ranjith-Kumar, C.T., Gutshall, L., Kim, M.J., Sarisky, R.T., Kao, C.C., 2002a. Requirements for de novo initiation of RNA synthesis by recombinant flaviviral RNA-dependent RNA polymerases. J Virol 76, 12526-12536.

Ranjith-Kumar, C.T., Kim, Y.C., Gutshall, L., Silverman, C., Khandekar, S., Sarisky, R.T., Kao, C.C., 2002b. Mechanism of de novo initiation by the hepatitis C virus RNA-dependent RNA polymerase: role of divalent metals. J Virol 76, 12513-12525.

Ratia, K., Saikatendu, K.S., Santarsiero, B.D., Barretto, N., Baker, S.C., Stevens, R.C., Mesecar, A.D., 2006. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. Proc Natl Acad Sci U S A 103, 5717-5722.

Ricagno, S., Egloff, M.P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J., Canard, B., 2006. Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. Proc Natl Acad Sci U S A 103, 11892-11897.

Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A., Gorbalenya, A.E., 2018. A planarian nidovirus expands the limits of RNA genome size. PLoS Pathog 14, e1007314.

Salanueva, I.J., Carrascosa, J.L., Risco, C., 1999. Structural maturation of the transmissible gastroenteritis coronavirus. J Virol 73, 7952-7964.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. J Virol 81, 20-29.

Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H., Siddell, S.G., 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog 1, e39.

Schagger, H., 2006. Tricine-SDS-PAGE. Nat Protoc 1, 16-22.

Sexton, N.R., Smith, E.C., Blanc, H., Vignuzzi, M., Peersen, O.B., Denison, M.R., 2016. Homology-Based Identification of a Mutation in the Coronavirus RNA-Dependent RNA Polymerase That Confers Resistance to Multiple Mutagens. Journal of virology 90, 7415-7428.

Seybert, A., Posthuma, C.C., van Dinten, L.C., Snijder, E.J., Gorbalenya, A.E., Ziebuhr, J., 2005. A complex zinc finger controls the enzymatic activities of nidovirus helicases. J Virol 79, 696-704.

Seybert, A., van Dinten, L.C., Snijder, E.J., Ziebuhr, J., 2000. Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. J Virol 74, 9586-9593.

Shim, J.H., Larson, G., Wu, J.Z., Hong, Z., 2002. Selection of 3'-template bases and initiating nucleotides by hepatitis C virus NS5B RNA-dependent RNA polymerase. J Virol 76, 7030-7039.

Siddell, S., Wege, H., ter Meulen, V., 1982. The structure and replication of coronaviruses. Curr Top Microbiol Immunol 99, 131-163.

Siddell, S., Wege, H., Ter Meulen, V., 1983. The biology of coronaviruses. J Gen Virol 64 (Pt 4), 761-776.

Siddell, S.G., Walker, P.J., Lefkowitz, E.J., Mushegian, A.R., Adams, M.J., Dutilh, B.E., Gorbalenya, A.E., Harrach, B., Harrison, R.L., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M., Rubino, L., Sabanadzovic, S., Sanfacon, H., Simmonds, P., Varsani, A., Zerbini, F.M., Davison, A.J., 2019. Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). Arch Virol 164, 943-946.

Simpson, R.J., 2006. Electrophoresis of Peptides (Tricine-SDS-PAGE). CSH Protoc 2006.

Slanina, H., Madhugiri, R., Bylapudi, G., Schultheiss, K., Karl, N., Gulyaeva, A., Gorbalenya, A.E., Linne, U., Ziebuhr, J., 2021. Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. Proc Natl Acad Sci U S A 118.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 331, 991-1004.

Snijder, E.J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J.J., van der Meulen, J., Koerten, H.K., Mommaas, A.M., 2006. Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. J Virol 80, 5927-5940.

Snijder, E.J., van Tol, H., Roos, N., Pedersen, K.W., 2001. Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. J Gen Virol 82, 985-994.

Snijder, E.J., Wassenaar, A.L., Spaan, W.J., 1992. The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. J Virol 66, 7040-7048.

Snijder, E.J., Wassenaar, A.L., Spaan, W.J., 1994. Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. J Virol 68, 5755-5764.

Snijder, E.J., Wassenaar, A.L., Spaan, W.J., Gorbalenya, A.E., 1995. The arterivirus Nsp2 protease. An unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. J Biol Chem 270, 16671-16676.

Snijder, E.J., Wassenaar, A.L., van Dinten, L.C., Spaan, W.J., Gorbalenya, A.E., 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. J Biol Chem 271, 4864-4871.

Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., Qin, C., 2019. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses 11.

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D., Rappuoli, R., 2003. SARS--beginning to understand a new virus. Nat Rev Microbiol 1, 209-218.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M., Rao, Z., 2006. Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. J Virol 80, 7902-7908.

Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B., Imbert, I., 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci U S A 111, E3900-3909.

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L.L., Diprose, J., Alderton, D., Walsh, M., Grimes, J.M., Stuart, D.I., 2004. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure 12, 341-353.

Svitkin, Y.V., Costa-Mattioli, M., Herdy, B., Perreault, S., Sonenberg, N., 2007. Stimulation of picornavirus replication by the poly(A) tail in a cell-free extract is largely independent of the poly(A) binding protein (PABP). RNA 13, 2330-2340.

Tabor, S., Richardson, C.C., 1989. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. Proc Natl Acad Sci U S A 86, 4076-4080.

Tan, Y.W., Fung, T.S., Shen, H., Huang, M., Liu, D.X., 2018. Coronavirus infectious bronchitis virus non-structural proteins 8 and 12 form stable complex independent of the non-translated regions of viral RNA and other viral proteins. Virology 513, 75-84.

te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H., Snijder, E.J., 2010a. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. Nucleic Acids Res 38, 203-214.

te Velthuis, A.J., van den Worm, S.H., Sims, A.C., Baric, R.S., Snijder, E.J., van Hemert, M.J., 2010b. Zn(2+) inhibits coronavirus and arterivirus RNA polymerase activity in vitro and zinc ionophores block the replication of these viruses in cell culture. PLoS Pathog 6, e1001176.

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

Tibbles, K.W., Brierley, I., Cavanagh, D., Brown, T.D., 1996. Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. J Virol 70, 1923-1930.

Tomar, S., Hardy, R.W., Smith, J.L., Kuhn, R.J., 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. J Virol 80, 9962-9969.

Tong, S., Conrardy, C., Ruone, S., Kuzmin, I.V., Guo, X., Tao, Y., Niezgoda, M., Haynes, L., Agwanda, B., Breiman, R.F., Anderson, L.J., Rupprecht, C.E., 2009. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerg Infect Dis 15, 482-485.

Tvarogova, J., Madhugiri, R., Bylapudi, G., Ferguson, L.J., Karl, N., Ziebuhr, J., 2019. Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3'-Terminal Adenylyltransferase Activity. J Virol 93.

van Aken, D., Zevenhoven-Dobbe, J., Gorbalenya, A.E., Snijder, E.J., 2006. Proteolytic maturation of replicase polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7. J Gen Virol 87, 3473-3482.

Van Breedam, W., Delputte, P.L., Van Gorp, H., Misinzo, G., Vanderheijden, N., Duan, X., Nauwynck, H.J., 2010a. Porcine reproductive and respiratory syndrome virus entry into the porcine macrophage. J Gen Virol 91, 1659-1667.

Van Breedam, W., Van Gorp, H., Zhang, J.Q., Crocker, P.R., Delputte, P.L., Nauwynck, H.J., 2010b. The M/GP(5) glycoprotein complex of porcine reproductive and respiratory syndrome virus binds the sialoadhesin receptor in a sialic acid-dependent manner. PLoS Pathog 6, e1000730.

van den Ent, F., Lowe, J., 2006. RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 67, 67-74.

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

van der Meer, Y., van Tol, H., Locker, J.K., Snijder, E.J., 1998. ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. J Virol 72, 6689-6698.

van Dinten, L.C., Rensen, S., Gorbalenya, A.E., Snijder, E.J., 1999. Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and Is essential for virus replication. J Virol 73, 2027-2037.

van Dinten, L.C., van Tol, H., Gorbalenya, A.E., Snijder, E.J., 2000. The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis. J Virol 74, 5213-5223.

van Dinten, L.C., Wassenaar, A.L., Gorbalenya, A.E., Spaan, W.J., Snijder, E.J., 1996. Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. J Virol 70, 6625-6633.

van Hemert, M.J., de Wilde, A.H., Gorbalenya, A.E., Snijder, E.J., 2008a. The in vitro RNA synthesizing activity of the isolated arterivirus replication/transcription complex is dependent on a host factor. J Biol Chem 283, 16525-16536.

van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E., Snijder, E.J., 2008b. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. PLoS Pathog 4, e1000054.

van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J., Snijder, E.J., 1999. Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. Proc Natl Acad Sci U S A 96, 12056-12061.

von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R., Haas, J., 2007. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. PLoS One 2, e459.

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Dempsey, D.M., Dutilh, B.E., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M., Orton, R.J., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D.B., Varsani, A., Zerbini, F.M., Davison, A.J., 2020. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). Arch Virol 165, 2737-2748.

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

Wassenaar, A.L., Spaan, W.J., Gorbalenya, A.E., Snijder, E.J., 1997. Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. J Virol 71, 9313-9322.

Wedekind, J.E., Frey, P.A., Rayment, I., 1996. The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyltransferase provides insight into phosphoryl group transfer. Biochemistry 35, 11560-11569.

Weiss, S.R., Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol Mol Biol Rev 69, 635-664.

Westerbeck, J.W., Machamer, C.E., 2015. A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway. J Virol 89, 9313-9323.

Wilson, L., McKinlay, C., Gage, P., Ewart, G., 2004. SARS coronavirus E protein forms cation-selective ion channels. Virology 330, 322-331.

Wright, S., Poranen, M.M., Bamford, D.H., Stuart, D.I., Grimes, J.M., 2012. Noncatalytic ions direct the RNA-dependent RNA polymerase of bacterial double-stranded RNA virus varphi6 from de novo initiation to elongation. J Virol 86, 2837-2849.

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

Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S.G., Ding, J., 2003. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. Nucleic Acids Res 31, 7117-7130.

Yao, Z., Jones, D.H., Grose, C., 1992. Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. PCR Methods Appl 1, 205-207.

Yi, G.H., Zhang, C.Y., Cao, S., Wu, H.X., Wang, Y., 2003. De novo RNA synthesis by a recombinant classical swine fever virus RNA-dependent RNA polymerase. Eur J Biochem 270, 4952-4961.

Yu, F., Hasebe, F., Inoue, S., Mathenge, E.G., Morita, K., 2007. Identification and characterization of RNA-dependent RNA polymerase activity in recombinant Japanese encephalitis virus NS5 protein. Arch Virol 152, 1859-1869.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367, 1814-1820.

Zeegers, J.J., Van der Zeijst, B.A., Horzinek, M.C., 1976. The structural proteins of equine arteritis virus. Virology 73, 200-205.

Zeng, Z., Deng, F., Shi, K., Ye, G., Wang, G., Fang, L., Xiao, S., Fu, Z., Peng, G., 2018. Dimerization of Coronavirus nsp9 with Diverse Modes Enhances Its Nucleic Acid Binding Affinity. J Virol 92.

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

Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z., Rao, Z., 2018. Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. J Virol 92.

Zhang, M., Cao, Z., Xie, J., Zhu, W., Zhou, P., Gu, H., Sun, L., Su, S., Zhang, G., 2013. Mutagenesis analysis of porcine reproductive and respiratory syndrome virus nonstructural protein 7. Virus Genes 47, 467-477.

Zhang, R.Q., Ellis, K.J., 1989. In vivo measurement of total body magnesium and manganese in rats. Am J Physiol 257, R1136-1140.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., China Novel Coronavirus, I., Research, T., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382, 727-733.

Ziebuhr, J., 2004. Molecular biology of severe acute respiratory syndrome coronavirus. Curr Opin Microbiol 7, 412-419.

Ziebuhr, J., 2005. The coronavirus replicase. Curr Top Microbiol Immunol 287, 57-94.

Ziebuhr, J., Herold, J., Siddell, S.G., 1995. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. J Virol 69, 4331-4338.

Ziebuhr, J., Heusipp, G., Siddell, S.G., 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J Virol 71, 3992-3997.

Ziebuhr, J., Schelle, B., Karl, N., Minskaia, E., Bayer, S., Siddell, S.G., Gorbalenya, A.E., Thiel, V., 2007. Human coronavirus 229E papain-like proteases have overlapping specificities but distinct functions in viral replication. J Virol 81, 3922-3932.

Ziebuhr, J., Siddell, S.G., 1999. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. J Virol 73, 177-185.

Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol 81, 853-879.

Ziebuhr, J., Thiel, V., Gorbalenya, A.E., 2001. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. J Biol Chem 276, 33220-33232.

Züst, R., Miller, T.B., Goebel, S.J., Thiel, V., Masters, P.S., 2008. Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. J Virol 82, 1214-1228.

10 Supplementary data

10.1 Comparative analysis of alpha- and betacoronavirus nsp(7+8) complexes regarding their quarternary structures and roles in supporting nsp12-mediated initiation and elongation of RNA synthesis

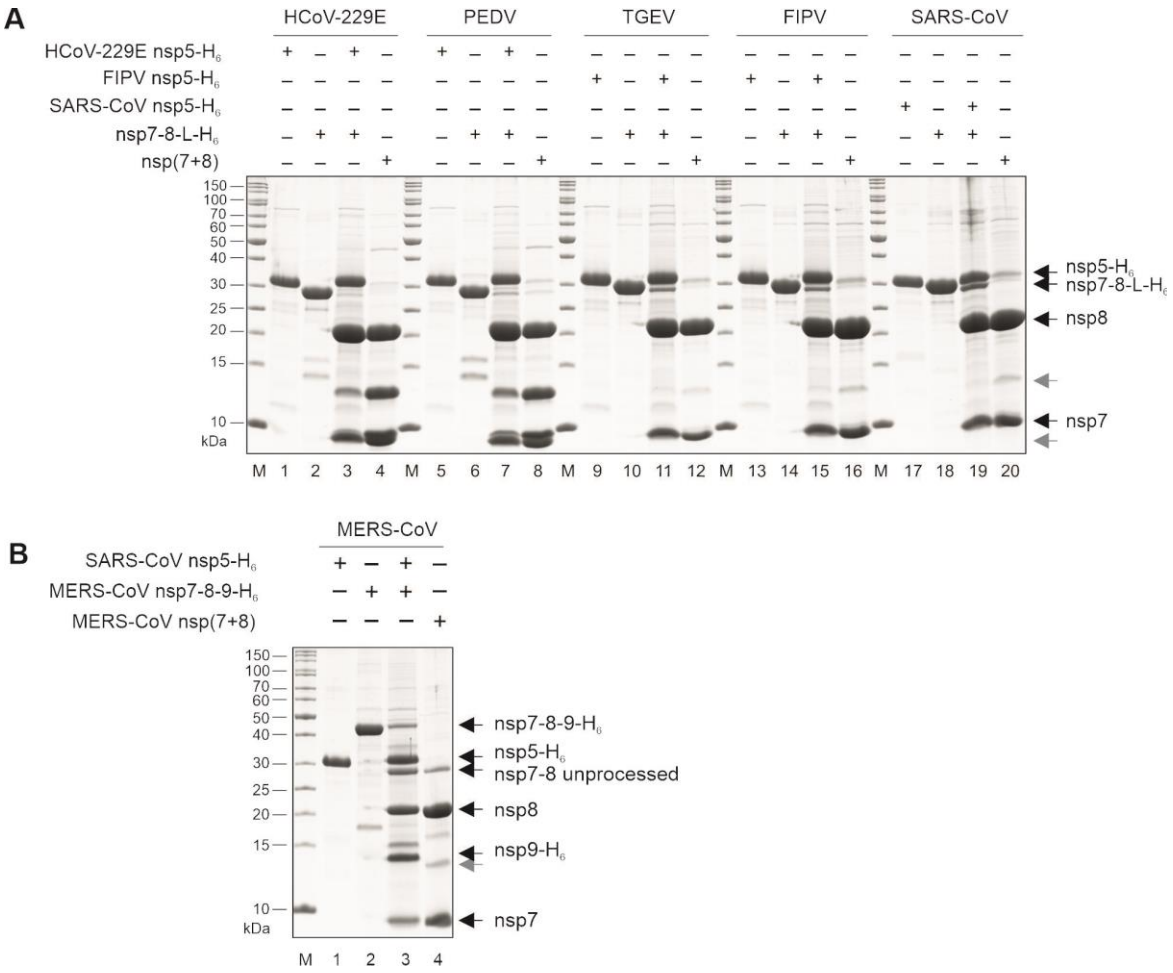
Results for the initiation of primer-dependent and *de novo* RNA synthesis indicated that the both nsp12 and the nsp(7+8) complex are essential for both activities. Interestingly, Mn^{2+} played an important role in facilitating *de novo* initiation of RNA synthesis by the nsp(7+8)+nsp12 complex. This raised several questions: i) Do all coronavirus nsp(7+8) complexes adopt a similar architecture, although the available structures of nsp7+8 complexes obtained for SARS-CoV, SARS-CoV-2 and FCoV seem to contradict this idea (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; te Velthuis et al., 2012; Xiao et al., 2012; Zhai et al., 2005)). ii) Are coronavirus replicase complexes equally capable of synthesizing RNA *in vitro* using primer-dependent and *de novo* initiated RNA synthesis despite (potentially different) stoichiometries of the nsp(7+8) complexes involved. iii) Do coronavirus replicase complexes respond differently to different metal ions for the initiation of RNA synthesis. To answer these questions, we decided to produce nsp(7+8) complexes from four alphacoronaviruses (HCoV-229E, PEDV, TGEV and FIPV) and two betacoronaviruses (SARS-CoV and MERS-CoV) along with their respective nsp12 enzymes.

10.1.1 Generation of nsp(7+8) complexes from alpha- and betacoronaviruses

Initially, we produced pre-assembled nsp(7+8) complexes from nsp7-8-9-H_6 polyprotein precursors by nsp5-H_6-mediated proteolytic cleavage. To simplify this procedure, nsp(7+8) complexes were generated from nsp7-8-L-H_6 rather than nsp7-8-

9-H_6 precursors. To this end, 12 nucleotides coding for the four amino acids SGSG were inserted between the C-terminal residue of nsp8 (corresponding to the P1 position of the nsp8|9 M^{pro} cleavage site) and the H_6-tag (except MERS-CoV). The nsp7-8 coding sequences of HCoV-229E, PEDV, TGEV, FIPV, SARS-CoV, and MERS-CoV nsp7-8-9 were cloned into the pASK-Ub-nsp-H_6 plasmid and protein expression, and purification was carried out as described in Material and Methods. The Ni-NTA elution fractions of HCoV-229E, PEDV, TGEV, FIPV, SARS-CoV nsp7-8-L-H_6 proteins and MERS-CoV nsp7-8-9-H_6 fractions were subjected to anion exchange chromatography. Peak fractions of the eluted proteins were combined and dialyzed at 4°C against cleavage buffer.

As described in Material and Methods, the proteases of HCoV-229E, FIPV and SARS-CoV were purified with a C-terminal H_6-tag. HCoV-229E nsp5-H_6 was used to cleave the nsp7-8-L-H_6 precursors of HCoV-229E and PEDV, FIPV nsp5-H_6 was used to process the nsp7-8-L-H_6 precursors of FIPV and TGEV, and SARS-CoV nsp5-H_6 was used to cleave the nsp7-8-L-H_6 precursor of SARS-CoV and MERS-CoV to produce nsp7, nsp8, and L-H_6. After completion of the cleavage, the reaction mixture was loaded onto a Ni-NTA column. The nsp(7+8) complexes in the flowthrough were further purified by anion exchange chromatography. Supplementary Figure S1 (A and B) shows an SDS-PAGE analysis of the processing of nsp7-8-L-H_6 and nsp7-8-9-H_6 by nsp5-H_6 (5 µg of proteases were mixed with 15 µg polyprotein and incubated at 4°C for 48 h). The data show efficient cleavage of the precursor proteins with recombinant nsp5 and subsequent enrichment of the proteins constituting the nsp(7+8) complexes.

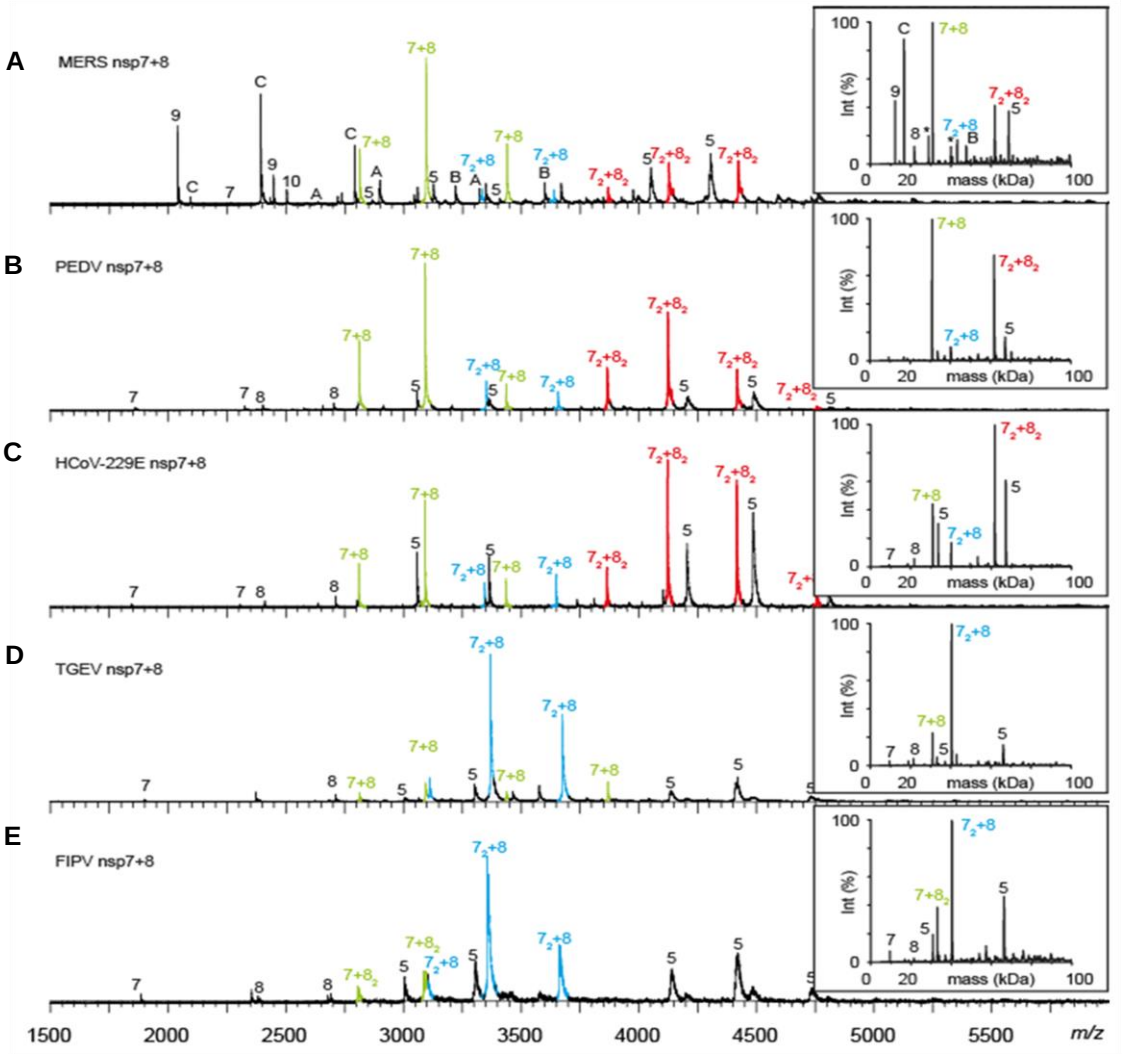


Supplementary Figure S1: Tris-tricine PAGE showing the proteolytic release and purification of pre-formed alpha- and betacoronavirus nsp(7+8) complexes with authentic N- and C-termini after nsp5-H$_6$ (3CL^{pro}) cleavage. Tris-tricine-PAGE analysis of 3CL^{pro}-mediated processing and generation of coronavirus nsp(7+8) complexes from the polyprotein precursors nsp7-8-L-H$_6$ and nsp7-8-9-H$_6$, respectively. To generate nsp(7+8) complexes, nsp7-8-L-H$_6$ or nsp7-8-9-H$_6$ and nsp5-H$_6$ proteins were purified using Ni-IMAC and anion-exchange chromatography. Nsp7-8-L-H$_6$ or nsp7-8-9-H$_6$ (15 µg) were cleaved with 5 µg 3CL^{pro} (nsp5-H$_6$) for 48 h at 4°C, followed by the removal of L-H$_6$, nsp9-H$_6$ and nsp5-H$_6$ using Ni-IMAC. Subsequently, nsp(7+8) complexes were enriched using anion-exchange chromatography. **(A)** Nsp(7+8) complexes of HCoV-229E, PEDV, FIPV, TGEV, and SARS-CoV were generated using nsp7-8-L-H$_6$. Coomassie brilliant blue-stained Tris-tricine polyacrylamide gel. Lanes 1, 5, 9, 13, 17: purified 3CL^{pro} (nsp5-H$_6$); lanes 2, 6, 10, 14, 18: nsp7-8-L-H$_6$; lanes 3, 7, 11, 15, 19: 3CL^{pro}-mediated cleavage reaction; lanes 4, 8, 12, 16, 20: enriched nsp(7+8) complexes. **(B)** The MERS-CoV nsp(7+8) complex was generated using a nsp7-8-9-H$_6$ precursor substrate. Lane 1: 3CL^{pro} (nsp5-H$_6$); lane 2: nsp7-8-9-H$_6$; lane 3: 3CL^{pro}-mediated cleavage reaction; lane 4: enriched nsp(7+8) complex. Lane M: marker proteins with molecular masses (in kilodaltons) indicated to the left. Black arrows on the right indicate the identities of individual proteins after 3CL^{pro}-mediated cleavage. Gray arrowheads indicate aberrant *in vitro* cleavage products of nsp8. nsp – nonstructural protein. + or – on top of the gels indicate the presence or absence of the respective protein in the reaction.

10.1.2 Native mass spectrometry of alpha- and betacoronavirus nsp(7+8) complexes

To analyze the multimeric state of nsp(7+8) complexes of alpha- and betacoronaviruses, purified HCoV-229E, PEDV, TGEV, FIPV and SARS-CoV nsp7-8-L-H_6 and MERS-CoV nsp7-8-9-10-11-H_6 were subjected to proteolytic processing. HCoV-229E nsp5-H_6 was used to cleave HCoV-229E and PEDV nsp7-8-L-H_6, FIPV nsp5-H_6 was used to cleave FIPV and TGEV nsp7-8-L-H_6, and SARS-CoV nsp5-H_6 was used to cleave SARS-CoV nsp7-8-L-H_6 and MERS-CoV nsp7-8-9-10-11-H_6 into individual products. The reactions were analyzed by native MS. Native mass spectrometry suggested that the nsp5 homodimer did not participate in stable interactions with other proteins. As shown in Supplementary Figure S2 A, proteolytic processing of the MERS-CoV nsp7-8-9-10-11-H_6 polyprotein gave rise to heterodimeric (1:1), heterotrimeric (2:1) and heterotetrameric complexes (2:2). Also detected were incompletely processed intermediates, such as nsp7-8-9, nsp9-10 and nsp10-11-H_6, which appeared as intact masses and did not participate in interactions (Supplementary Figure S2 A), and unbound monomeric nsp7 and nsp8. The HCoV-229E and PEDV nsp(7+8) complexes were identified as heterodimers (1:1), heterotrimers (2:1) and, predominantely, as heterotetramers (2:2) (Supplementary Figure S2 B and C). TGEV and FIPV complexes were heterodimers (1:1) and predominantly heterotrimers (2:1) (Supplementary Figure S2 D and E). These stoichiometries are predicted based on the molecular weights of the subunits. However, to provide further evidence, molecular ions of three different charge states of the nsp7+8 heterotetramer of HCoV-229E, PEDV, MERS-CoV and heterotrimers of TGEV and FIPV nsp(7+8) complexes were selected and subjected to

CID-MS/MS to successively induce subunit dissociation, by stepwise increasing the collision voltage (Krichel et al., 2021; Krichel et al., 2020).

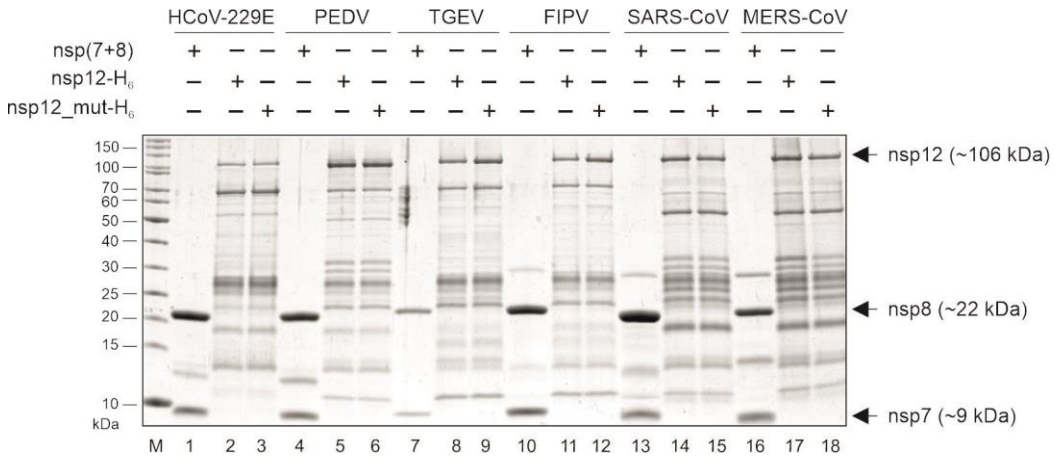


Supplementary Figure S2: Native MS of alpha- and betacoronavirus nsp(7+8) complex formation. Shown are representative native mass spectra and their deconvoluted mass spectra (insets) depicting nsp(7+8) complex formation from **(A)** MERS-CoV, **(B)** PEDV, **(C)** HCoV-229E, **(D)** TGEV, **(E)** FIPV. Complex formation was measured following nsp5-mediated cleavage of precursor (at 15 µM concentration) in 300 mM NH_4COOH (pH 8) and 1 mM DTT. Peak series assigned to specific proteins are indicated and complexes of nsp7 and nsp8 are additionally colored.

10.1.3 Expression and purification of alpha- and betacoronavirus nsp12

To compare the role of pre-assembled alpha- and betacoronavirus nsp(7+8) complexes in RNA synthesis, the 106-kDa protein nsp12 and the mutant derivative, nsp12_SAA,

of HCoV-229E, PEDV, TGEV, MERS-CoV and SARS-CoV nsp12_D618A were purified using the ubiquitin-nsp expression system (Supplementary Figure S3: lanes 3, 6, 9, 12, 15, 18). As shown in Figure S 3, there was no obvious difference in the expression and enrichment of the respective nsp12 proteins used in subsequent polymerase activity assays.



Supplementary Figure S3: Purified nsp(7+8) and nsp12-H_6 of alpha- and betacoronaviruses. Coomassie brilliant blue-stained 12 % Tris-tricine-polyacrylamide gel showing nsp(7+8) complexes purified by anion exchange and size-exlusion chromatography and the nsp12 and nsp12_mut proteins purified by anion exchange chromatography: 5 µg of nsp(7+8) complexes of HCoV-229E, PEDV, FIPV, SARS-CoV, MERS-CoV and 2 µg of TGEV nsp(7+8) and 2 µg of all nsp12 and nsp12_mut proteins were analyzed. Lanes 1, 2, 3: HCoV-229E; lanes 4, 5, 6: PEDV; lanes 7, 8, 9: TGEV; lanes 10, 11, 12: FIPV; lanes 13, 14, 15: SARS-CoV and lanes 16, 17, 18: MERS-CoV. nsp12_mut indicates nsp12 variants containing alanine substitutions of conserved active-site aspartic acid residues. HCoV-229E: D4828 and D4829, PEDV: D4855 and D4856, TGEV: D4755 and D4756, FIPV: D4780 and D4781, SARS-CoV: D4987, MERS: D5138 and D5139. Numbers indicate the positions of residues in the polyprotein 1ab of the respective virus. nsp: nonstructural protein. Lane M: marker proteins with molecular masses (in kilodaltons) indicated to the left.

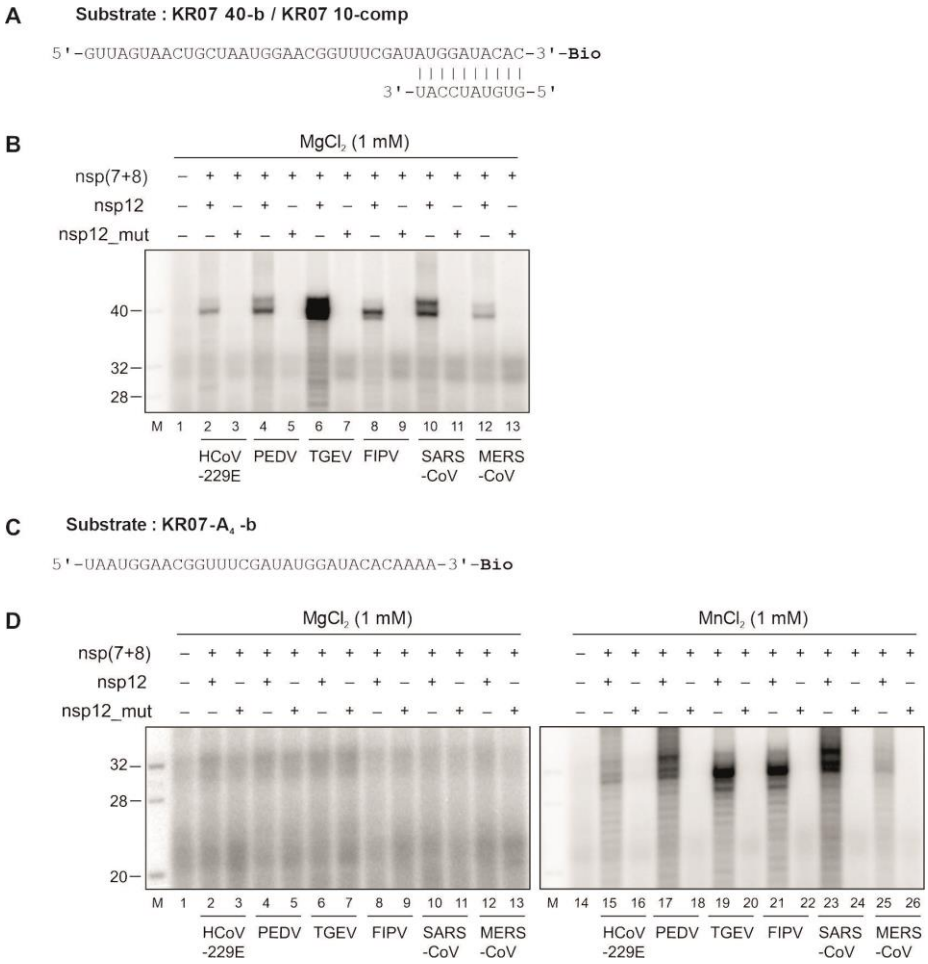
10.1.4 Comparative analysis of polymerase activities of coronavirus nsp(7+8)/nsp12 replicase complexes

Elongation and *de novo* RNA synthesis activities of nsp(7+8)+nsp12 complexes of alpha- and betacoronaviruses were compared. To determine elongation activity, an *in vitro* primer-dependent polymerase activity assay was performed using the KR07 40 -b/ KR07 10-comp RNA substrate (Supplementary Figure S4 A). The heteropolymeric 5'-overhang of KR07 40 -b served as template during the elongation reaction. For *de*

novo RNA synthesis, a 32-oligonucleotide RNA containing four adenine nucleotides at the 3'-end served as template. The RNAs (except the KR07 10 primer in the KR07 40 - b/ KR07 10-comp RNA substrate) were 3'-biotinylated to exclude terminal nucleotide transferase activity (Supplementary Figure 4 A, C) and reactions were performed in the presence of all four nucleotides.

Reaction products were detected by the incorporation of radiolabeled [α-^{32}P]-UTP during RNA synthesis (Supplementary Figure S4 B and D). Reactions containing no enzyme were used as negative controls. All replicase complexes generated elongated products of the expected length in the presence of Mg^{2+} (Supplementary Figure 4 B; lanes 2, 4, 6, 8, 10, 12). The TGEV nsp(7+8)+nsp12 complex had the highest activity among the enzymes analyzed (Supplementary Figure S4 B, lane 6). The SARS-CoV, PEDV and FIPV complexes had somewhat lower activity (Supplementary Figure S4 B; lanes 4, 8, 10), while the HCoV-229E and MERS-CoV complexes had the lowest activity (Supplementary Figure 4 B, lanes 2 and 12). As expected, there was no evidence for RNA polymerase activity in mixtures of the respective nsp(7+8) complexes with mutant forms of nsp12 containing single or double substitutions in the RdRp active site. Next, *de novo* RNA synthesis activity was tested in the presence of both Mg^{2+} and Mn^{2+}. In line with the data shown above (Fig. 24), Mg^{2+} did not support the initiation of RNA synthesis (Supplementary Figure S4 D; lanes 2-13). However, in the presence of Mn^{2+}, all nsp(7+8)+nsp12 replicase complexes generated products of (or close to) the expected length (around 32 nucleotides) with some variations by *de novo* initiation of RNA synthesis (Supplementary Figure S4 D; lanes 15, 17, 19, 21, 23, 25).

No activity was detected for the nsp12 mutant forms (Supplementary Figure S4 D; lanes 16, 18, 20, 22, 24, 26). Collectively, this data indicates that all replicase complexes can use *de novo* initiation to initiate RNA synthesis and Mn^{2+} is an essential catalytic ion for the initiation of RNA synthesis.



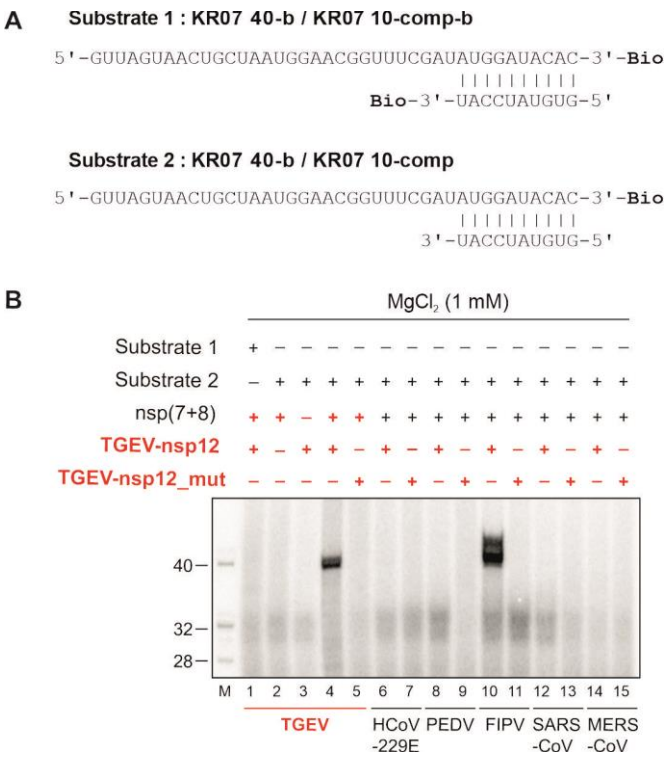
Supplementary Figure S4. Replicase complexes nsp(7+8)+nsp12 from different alpha- and betacoronaviruses display efficient *de novo* initiation and primer-dependent elongation activities *in vitro*. (A) Partially double-stranded RNA used in the primer-dependent polymerase assay shown in **(B)**, 'Bio' indicates RNA 3'-biotinylation **(B)** Primer-dependent elongation assays were performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of protein, 1 µM of RNA substrate, NTP mix (100 µM of ATP, CTP, GTP and 75 µM UTP) and 0.17 µM [α-^{32}P]-UTP in the presence of either $MgCl_2$ or $MnCl_2$ as indicated above the autoradiograms. **(C)** single-stranded RNA used in *de novo* initiation assays. 'Bio' indicates RNA 3'-biotinylation. **(D)** *De novo* initiation assays were performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of protein, 1 µM of RNA substrate, NTP mix (100 µM of ATP, CTP, GTP and 75 µM UTP) and 0.17 µM [α-^{32}P]-UTP in the presence of either $MgCl_2$ or $MnCl_2$ as indicated above the autoradiograms. *De novo* and primer-dependent elongation assays contained a 1:1 molar ratio of nsp(7+8) and nsp12. nsp12_mut: nsp12 variant containing alanine substitutions of conserved aspartic acid residues (HCoV-229E: D4828 and D4829; PEDV: D4855 and 4856; TGEV: D4755 and D4756; FIPV: D4780 and D4781; SARS-CoV: D4987; MERS-CoV: D5138 and D5139. Numbers indicate the positions of residues in the polyprotein 1ab of the indicated virus). + or -: reactions performed in the presence or absence of the indicated protein. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea

gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of the autoradiograms.

10.1.5 Elongation activities in reactions containing different alpha- and betacoronavirus nsp(7+8) complexes in combination with TGEV nsp12

All nidoviral RNA-dependent RNA polymerases have a similar fold consisting of the classical “cupped right hand” structure with palm, fingers, and thumb subdomains (Gao et al., 2020; Gorbalenya et al., 2002; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Xu et al., 2003). Co-structures for nsp7+8+nsp12 revealed that a nsp7+8 heterodimer associates with the thumb domain of the polymerase (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019) and this interaction is thought to be conserved among other coronaviruses. To test the level of conservation of interactions contributing to nsp(7+8)/nsp12 complex formation and functionality, we used TGEV nsp12 in combination with different alpha- and betacoronavirus nsp(7+8) complexes in a primer-dependent polymerase assay using KR07 40-b/ KR07 10 comp partial duplex RNA as substrate in the presence of Mg^{2+} and all four nucleotides (Supplementary Figure S5). The [α-^{32}P]-UTP-labeled reaction products were resolved by denaturing PAGE and detected by phsophorimaging. In a control reaction, a substrate was used in which both strands were 3’-biotinylated (KR07 40-b/ KR07 10-comp-b). As expected, no radiolabeled product was detected if this substrate was used in a reaction containing the TGEV nsp(7+8)+nsp12 complex (Supplemetary Figure S5 B; lane 1). Also, no activity was detected in reactions containing only nsp(7+8) or only nsp12 along with the KR07 40-b / KR07 10 substrate-2 (Supplementary Figure 5S B; lanes 2 and 3). As shown in Supplementary Figure S5 B (lane 4), the combination of TGEV nsp(7+8) and TGEV nsp12 generated products of the expected length of approximately 40 nucleotides in the

presence of the KR07 40-b / KR07 10-comp substrate-2. This different result obtained with the two versions of the same substrate (3'-modified at both ends or 3'-modified only at the bottom strand) confirmed that the observed polymerase activity resulted from the use of the synthetic RNA template provided in the reaction (KR07 40-b / KR07 10-comp substrate-2) rather than potentially co-purified *E. coli*-derived nucleic acids (RNA or DNA). As expected, the mutant TGEV nsp12 protein in which two conserved aspartate residues in the RdRp motif C were replaced with alanine (nsp12_RdRp_SDD to SAA) did not give rise to a radiolabeled product (Supplementary Figure S5 B; lane 5). In reactions in which the nsp(7+8) complexes of HCoV-229E, PEDV, SARS-CoV and MERS-CoV, respectively, were mixed with (non-cognate) TGEV nsp12, no primer-dependent elongation activity was detected (Supplementary Figure S5 B; lanes 6, 8, 12, 14). In striking contrast, if the FIPV nsp(7+8) complex was used in combination with TGEV nsp12, the production of a radiolabeled RNA product was readily detectable (Supplementary Figure S5 B; lane 10). Collectively, these results suggest specific interactions between nsp12 and its cognate nsp(7+8) complex or an nsp(7+8) complex from a very closely related virus. It is important to note that FIPV and TGEV are members of the same species called *Alphacoronavirus 1*, while all other viruses belong to other virus species.



Supplementary Figure S5: Activity of the TGEV nsp12 in the presence of the preformed TGEV nsp(7+8) complex or (noncognate) nsp(7+8) complexes of other coronaviruses. (A) Partially double-stranded RNA substrates used in this experiment. 'Bio' indicates RNA 3'-biotinylation. **(B)** The primer-dependent elongation assay was performed at 30°C for 60 min in standard reaction buffer supplemented with 1 µM of protein, 1 µM of RNA substrate, NTP mix (100 µM of ATP, CTP, GTP and 75 µM UTP) and 0.17 µM [α-^{32}P] GTP in the presence of $MgCl_2$ as indicated above the autoradiograms. Primer-dependent elongation assays contained a 1:1 molar ratio of nsp(7+8) and nsp12. nsp12_mut: TGEV nsp12 variant containing alanine substitutions of conserved aspartic acid residues D4755 and D4756 (numbers indicate the positions of residues in the TGEV polyprotein 1ab). + or -: reactions performed in the presence or absence of a given protein as indicated above the autoradiogram. „+ or "- in„ red"indicate TGEV nsp(7+8) and nsp12 wild-type or mutant proteins. Products were resolved in a TBE-buffered 12% polyacrylamide-7M urea gel and visualized by phosphorimaging. In lane M, 5'-[^{32}P]-labeled oligoribonucleotides were loaded as markers. Sizes (in nucleotides) are indicated on the left side of autoradiograms.

Publications

1. Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit.
   Slanina H[#], Madhugiri R[#], **Bylapudi G**, Schultheiß K, Karl N, Gulyaeva A, Gorbalenya AE, Linne U, Ziebuhr J.
   Proc Natl Acad Sci USA**.** 2021 Feb 9;118(6): e2022310118. doi: 10.1073/pnas.2022310118.

2. Hallmarks of Alpha- and Betacoronavirus nonstructural protein 7+8 complexes.
   Krichel B, **Bylapudi G**, Schmidt C, Blanchet C, Schubert R, Brings L, Koehler M, Zenobi R, Svergun D, Lorenzen K, Madhugiri R, Ziebuhr J and Uetrecht C*
   Science Advances. 03 Mar 2021: Vol. 7, no. 10, eabf1004, DOI: 10.1126/sciadv.abf1004.

3. Identification and characterization of a human coronavirus 229E nonstructural protein 8-Associated RNA 3'-Terminal Adenylyl transferase Activity.
   Tvarogová J[#], Madhugiri R[#], **Bylapudi G**, Ferguson LJ, Karl N, Ziebuhr J.
   J Virol. 2019 May 29;93(12). pii: e00291-19. doi: 10.1128/JVI.00291-19. Print 2019 Jun 15.

Manuscript in preparation

4. Characterization of primer-dependent and *de novo*-initiated RNA synthesis catalyzed by nsp(7+8)/nsp12 complexes of representative alpha- and betacoronaviruses.
   Bylapudi G[#], Madhugiri R[#], Slanina H, Krichel B, Tvarogová J, Karl N, Uetrecht C, and Ziebuhr J.
   [#] Indicates equal contribution

11. Declaration

I declare that I have prepared the submitted dissertation independently and without unauthorized outside help and only with the help that I have indicated in the dissertation. All text passages taken verbatim or in spirit from published writings and all information based on oral information are marked as such. I agree to a possible check of my dissertation by an antiplagiarism software. In the research I have carried out and mentioned in the dissertation, I have complied with the principles of good scientific practice as laid down in the "Statutes of Justus Liebig University Giessen for Ensuring Good Scientific Practice".

_____ _____

Date Signature

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.